ABSTRACT

Title of Dissertation: THE ACQUISITION OF *WH*-QUESTIONS IN TASHLHIYT BERBER: NOVEL BEHAVIORAL AND CORPUS STUDIES

Imane Bou-Saboun
Doctor of Philosophy, 2025

Dissertation Directed by: Professor Jeffrey Lidz
Department of Linguistics

Professor Maria Polinsky
Department of Linguistics

The central quest in this dissertation is how a child would acquire abstract structure in a language that masks the overt expressions of said structure. More specifically, the interest here is in studying the acquisition of *wh*-questions in a language with a grammar that shows the following properties: (1) the *wh*-word is always followed by an obligatory complementizer that is rendered inaudible by the phonology in most instances and (2) it does not differentiate between subject and object Ā-extraction in word order but through verbal and nominal morphology only. The case study explored here is that of the Afro-Asiatic language Tashlhiyt Berber (TB). In this language, the morphology is the predominant source of information about the position of a gap in Ā-dependencies.

We test the children's knowledge with a behavioral study and make hypotheses about the learning problem through a corpus of child-directed Tashlhiyt Berber. The comprehen-

sion study involved children between ages 2 and 4 and it was designed to test knowledge of reversible action verbs in subject and object wh-questions and relative clauses. We find that children's knowledge of wh-questions and relative clauses in the language is equivalent, meaning that these construction types are not predictive of performance. By ages 3 and 4, TB-acquiring kids show high comprehension of subject extraction which featured the anti-agreement morphology.

The behavioral study result that construction type (WH vs. RC) is non-predictive of performance comes despite some of the surface asymmetries between the two constructions and an asymmetry in terms of their input distribution as measured in the corpus. Although the input appears to be informative in terms of the distribution and quality of the data on anti-agreement, the input turns out to be uninformative about the obligatory complementizer and the nominal morphology necessary to infer the grammar of TB $\bar{A}$-construction. This would be problematic unless the child recruits informative data for well-represented constructions to learn about the less represented constructions in the input, which would require innate knowledge of the abstract identity of $\bar{A}$-constructions.

This is the first work in the Berber developmental literature that combines facts about the morphosyntax with a phonological conspiracy present in the language to formulate a developmental puzzle. Beyond this, the contribution of this dissertation is twofold. Firstly, a significant portion of the findings in it builds on the construction of the first-of-its-kind corpus of caregiver-child interactions in Tashlhiyt Berber. This corpus was collected for the purposes of this dissertation and is meant to be contributed, fully transcribed and media-linked, to the Child Language Data Exchange System (CHILDES) (MacWhinney 2000) repository. Secondly, this contribution diversifies the growing work on the study

of V1 languages with anti-agreement phenomena and the idiosyncratic language-specific Construct State from a developmental standpoint.

THE ACQUISITION OF *WH*-QUESTIONS IN TASHLHIYT BERBER: NOVEL BEHAVIORAL AND CORPUS STUDIES

by

Imane Bou-Saboun

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:
Professor Jeffrey Lidz, Chair/Advisor
Professor Maria Polinsky, Chair/Advisor
Associate Professor Alexander Williams
Assistant Professor Aron Hirsch
Professor Yi Ting Huang, Dean's Representative

Dedication

To Malika El Aliani, my *mamà*, who keeps the dreams alive.

To my *nonno*, M'barek El Aliani, who unchained the butterfly effect of opportunity with a little trip to Trapani back in 1989.

Iggut madnnan, idrus maissnn. Much gets said, very few know.

Amazigh Proverb

Acknowledgments

As Marie Kondo (or anyone with commonsense really) would say, it's inadvisable to peek into a photo album in the middle of a tidying or packing spree. You catch a glimpse of a memory and that's it, you're not here anymore, and you can kiss your goals to be all packed up by the time movers come goodbye. Yet, despite the pinch I'm in, I can't resist the need to open the metaphorical photo album of my five years in this program, as I am keen on properly paying homage to the people that have made it one of the most transformative times of my existence here on Earth.

I would like to start by thanking my advisors who created this wonderful opportunity in the form of a PhD program admission for me, Jeffrey Lidz and Masha Polinsky, who both pushed me intellectually and believed in the corpus work featured here since day zero - when we all had a virtual meeting about it back in November of 2020 - and they encouraged me to dedicate my time and efforts towards bringing it to the light.

To Jeff, thank you for being there every step of the way, including the last moments of intense editing. I am deeply grateful for the rigorous training one gets by being your student and your eye for good writing. And thank you for your brand of interfaith jokes, especially during Ramadan.

To Masha, thank you for your no-nonsense approach, for your signature lightning-fast response times from all time zones and for your travel stories. Teaching the course on heritage linguistics with you taught me so much and I'm honored for that precious opportunity and for your trust in me.

I want to thank my committee, Alexander Williams, Aron Hirsch and Yi Ting Huang. I am very grateful to have had the opportunity to make you read my thesis and receive deeply careful and thorough feedback colored with your own individual interests and expertise.

I want to take the time to thank all the people at the UMD department of Linguistics during my time here. I have limited time to do so but I will attempt it. Many thanks to Peggy Antonisse, for her guidance throughout the teaching process and for making my first time ever experience as a TA feel genuinely exciting. Thank you so much to Tonia Bleam for your guidance also and for entrusting me with some substitutions for you and for our various conversations (special mention to the winks). Many thanks to Howard Lasnik, for his passion and his availability to answer my syntactic questions. Thank you for your patience and for being happy to look through all those squib drafts and my 11-page compendium of our office hour discussions. Many thanks to Norbert Hornstein for answering my doubts about subjacency and for your Socratic method and for lending me some books. Thank you, Alex Chabot, for the peppermint green teas, the French conversations about Moroccan cuisine as well as, of course, Berber syllable structure. Speaking of French, I want to thank

Valentine Hacquard for her amazing seminar on the acquisition of semantics back in my second year. I am deeply grateful for your nudging me in the direction of Italian acquisition and for making me write a squib that I still would love to flesh out into a real paper. You made me feel like I had something to say in a domain that I was just getting acquainted to and emboldened me with your feedback. I want to thank Naomi Feldman for always having your office door open throughout that semester I was attending your ling620 course and for looking for and finding me RAs and possible collaborators since the moment you learned I was working on a corpus. Thank you, Ellen Lau, for your time, your precious contribution as a committee member in my 888, for your teaching and the careful, thoughtful way that you read, formulate questions and encourage others to do the same. I want to thank Bill Idsardi for approving my first expenses and trip to Morocco in the summer of 2022, as well as his support to travel to and from conferences to disseminate the knowledge I was lucky to gather from there. Thank you so much Kim, you're a wonderful human being that has made this degree not only bureaucratically smoother than an Italian could ever wish for in her lifetime, but also humanly pleasant. Thank you for our round-ups and for helping out each and every single time with the never-ending questions I had for you.

I am very grateful for the people at the UMD Acquisition Lab, ever since the beginning, for all your thoughtful questions that helped me refine presentations and ways of thinking. I want to thank Tara Mease for your time, helping me film the brand new stims for the Berber study with Sterling Mullenix and also for walking me through how to run studies the first time. I want to thank my mentee Josh Kwak for his assistance with the earliest transcriptions of the corpus of child-directed Berber. Yu'an Yang and Hisao Kurokami get a special mention for paving the way for me to run the study in Morocco with their respective recruitment experiences in China and Japan as well as their advice with travel.

I would like to thank the students who decided to sign up and attend my advanced course on the acquisition of understudied languages for all the discussions and for really sharing your minds and wonderfully daring ideas and opinions with me.

I also want to mention with gratitude all the people that I met over the years at talks, colloquia and poster sessions for their input and time in discussing and developing ideas and teaching me directly or indirectly how to become a clearer explainer. I'm very grateful to Athulya Aravind, Matt Wagers, Nozomi Tanaka, Kamil Deen, Nan Ratner in particular. Many thanks to Masaya Yoshida, who taught me a lot about how to answer any typology of question during QAs.

I want to thank some new colleagues and friends that I made in the setting of conferences which then simply became life. Thank you, Esra Eldem Tunç, a wonderful human I got to meet and spend time with at the UCLA Heritage Institute back in 2023 and with whom I've stayed in contact ever since. Seeing you again in person at the UMD HSP this year was such a treat. I want to thank Martine Gallardo as well for being an amazing talker, thinker and for your astute understanding of the dynamics in our field. Thank you very much to Giovanni Roversi, fellow Aemilian across the pond, and lovely and hospitable Keely New.

This corpus reports my name, but it is the product of the invisible contributions of myriads of people far beyond authorship and the direct crafting of the final result. This section is an attempt to show gratitude to those many hands that were lent to me, visibly

or invisibly, as the thick tapestry that forms this work was being woven. I would like to thank Rachida El Aliani, her sister Naima El Aliani and her mother *ammti* Mleid who first agreed to let me go visit them in Casablanca in September of 2019 to let me conduct my first ever survey and behavioral study with heritage speakers of Tashlhiyt Berber who speak Moroccan Darija in their day-to-day.

I want to take a moment to thank people without whom the research couldn't have taken place practically in Morocco. Ayad Alhyane, Abdelkarim Oulgoud, M'barek Aderdor, Ustadh Zaydoune. My dearest *nonno*, M'barek El Aliani for driving me around, making us tea, and taking me shopping for avocados and mini bananas every week at the nearby market. Thank you Keltoum Nait-Ben-El-Hajj for your hospitality and for your help in recruiting children where you live.

The Alamy family who are based in Italy and who agreed to help me construct the corpus by visiting and filming personal moments in the development of their daughters. Without your openness to experimenting, I wouldn't have had a topic, or at least it wouldn't have been this, which has been truly incredible and has made me fall in love multiple times with the job of being a linguist.

I want to give really special thanks to the entire Aderdor family who hosted me while I was collecting data in Biougra for part of the summer of 2022. Thank you so much, Fatima Aderdor, for bringing me into your world of teachers, kids and preschools. Thanks to your parents, Yamna and M'barek Aderdor for the car rides to the kindergartens in the mountains, the market visits and for all of the couscous *mashallah*. Thanks to your brothers for the rides and the food. Special thanks to Aicha Omar, and the Principal in Ait Izza, for putting me in contact with the first pre-K I visited in Chtouka-Ait-Baha and then led to the introduction of a whole network of schools. Thank you so much Malika in Lqsbt-n-Tallaght for introducing me and liaising between me and the local families in the village and for your friendship. I want to thank all of the villages for opening their doors to me, and letting me run my camera and studies.

I would like to thank the people who are part of my growing network of Amazighologist acquaintances and colleagues. Thank you very much Jenia Gutova, Abdellah Elouatiq and Dalila Dehbia Gaoua for your advice, ideas and judgments.

Originally this entire idea of working on Berber came from a brief encounter with Ianthi Tsimpli in her office sometime in Easter Term of 2019. We were brainstorming ideas in her office on Sidgwick site about a possible BA Hons thesis topic and I am so grateful to her for pointing out to me this little neat phenomenon called Anti-Agreement. And that led to Ouhalla's work and then came heritage linguistics which led to reading the Polinsky 2018 book and then looking up where Masha Polinsky herself worked and then there I was, doing the real thing. It's been an amazing journey.

I owe a lot to David Willis who was my academic advisor at Selwyn College, Dora Alexopoulou my BA Hons dissertation advisor and to my final year advisor Kirsty McDougall. Thank you all for sparking my curiosity, modelling academia and for the reference letters that led me here.

Speaking of Cambridge I want to thank the friends there who I kept through the years and who kept me grounded and steady through this journey. I wanted to thank my Murray-Edwards girls, Majida Begum, Haneen Zeglam and Amal Abdullah, the most precious girl squad I could ask to be a part of in my undergrad years. I am so proud

of you all for what you have accomplished thus far, and how you've all made a name for yourselves since leaving Cam. I marvel at how we have managed to keep contact across the continents we respectively reside in. I will never forget you all travelling from Manchester, Cambridge and London, to explore Agadir when I was there for fieldwork. It was the best way to celebrate completing my data collection in 2022. I want to thank Edson Owusu, undergrad housemate and travel companion extraordinaire. I will always associate Nashville,Tennessee and Cinque Terre with you now. Thank you so much Suchitra Sebastian, for pretty much existing. You're a wonderful friend and so insightful and inspiring. It was so great talking to you about academia and we've been talking about it pretty much since I ever even considered applying for a PhD. From the Edinburgh Fringe to Planet Word in Washington, D.C., seeing you is always a treat. Yolanda Lam, I owe you a trip to Crema and surroundings. Zainab Haider, Zareen Bhatti for the spiritual conversations and the relatability.

I want to thank some of the people that made my time here and full of camaraderie and warmth.

Thank you to my cohort of inspiring and wildly eclectic personalities, Rosa Eun-Kyoung Lee, Yuanfan Ying, Luisa Seguin, Clara Cuonzo, Xinchi Yu and Leslie Famularo. Thank you for the annual dinners, the fifth-year solidarity, which eased the admin burden for sure. It's hard to encapsulate in a paragraph how much we have experienced as a group, going from pixelated squares on a screen during the "p*ndemic year" to seeing each other everyday, some of us even sharing office spaces.

Rosa, thank to you for our hangouts, the concerts, our charged iced matcha latte roundups, which I hope were cathartic for both. Thank you, Jack, for being an admirable academic sibling. Your availability to discuss anything with rigor whilst being encouraging is something that really struck me. As well as your generosity with everything you have and openness to others. I want to thank Mina Hirzel. I always looked forward to catching up with you in Boston during BUCLD about how the year went and how to stay grounded. Thank you for your whimsy, for literally opening my eyes on so many patterns, sociological and linguistic. You're the best involuntary mentor a grad can ask for. Thank you, Sathvik Nair, for the iftars during Ramadan, the music, the deeply intercultural chats and our whole ongoing bit about Italy being the India of the West and vice-versa. Sarah Boukendour, mon amie Kabyle préférée. On est l'épreuve que les Marocains et les Algeriens peuvent être amis! Disons-le a la génération de nos parents, maintenant. But also for so much more than this, of course, and you know it. I want to thank Allison Dods for her friendship, the BUCLD-buddying, the themed parties, the long long conversations, the personality tests and just so much more that cannot all fit here, including baby Tortie. Thank you Lydia Quevedo for being a perfect office-mate throughout my last year spent in front of that huge screen sifting through spreadsheets of Berber glosses. I will genuinely miss the rhythm of your keyboard typing. Thank you Utku Turk for the all-nighters on R fitting GLMs and so much more beyond that. Thanks to Zulfiyya Aghakishiyeva. I was so grateful to have you there in person, and for the new years we spent here in 2024, and the roadtrips with Jack to Philly and Colonial Williamsburg and for having someone here who felt like a sister. Jéssica Mendes, my adventure-pal. I was never a sporty person, but getting to know you marked that new phase of my life where if I don't move I feel sick. Thank you Jingyi Cheng, Joselyn Rodriguez, London Dixon and Katherine Howitt for the extracurriculars,

the rides, the food, the technical, practical and research favors and overall dependability.

I want to thank my roommates on the Hill House, seasons 1 through 4. I can't be trusted to keep the casts in each season straight, but I just want to thank you all for the laughter, the kitchen banters and hot takes, the political jokes, finding a light on at 3 am and feeling less insane, as well as the collective flooded basement fixing and so much more. Teamwork truly makes the dreamwork and no one teaches you that notion better than the Hill House crew. Therefore, thank you Jéssica Mendes, Mina Hirzel, Rodrigo Ranero, Luisa Seguin, Clara Cuonzo, Fedya Golosov, Marta Ponciano Lázaro, Salma Elhaoudi, Utku Turk, Filippo Grassi, and Apoorva Verma. Thank you TJ, for being an awesome landlord.

I can't even begin to express my love, adoration and admiration for Jessie Dong and Lil Chines Marzo, what started off as independent European connections that ended up becoming my core US family. Words cannot begin to express how much I owe you (and your dogs). You characterized my last two years here and this is the kind of community that one could only wish for. I want to thank Marco Forza, my long-time friend who gave me the strength to endure and make sense of many things in life with his book recs. I cherish the time you came over to visit pretty much the whole Smithsonian Museal complex and more and gave me an excuse to finally play the tourist in D.C. And for being the reason why I saw Bernie Sanders in person.

Thank you S.T., for the karak chai, the music and the verses and the wisdom that kept me going.

Grazie mamma per avermi detto che il tashlhiyt è una lingua inutile quando da piccola ti chiedevo perché parlassimo arabo con te mentre tu parlavi tashlhiyt coi nonni e gli zii. Se non l'avessi fatto, non me la sarei studiata per ripicca. Grazie mille ai miei fratelli, Hanane e Khalid, a mio zio Rachid e mio zio Mohamed e i miei cari nonni. Grazie alla famiglia Patelli, Mario e Natalina. Grazie di tutto cuore a Nadia Nicolai, la mia primissima reference-letter writer! Grazie infinite alla famiglia Bussola: Cecilia, Giovanna Stanzani, Simone Negrello, Leonardo Mariggió, Lorena Bergamini, Stefano Accorsi (non l'attore), Giuliano Bonori, Betta Tonni. Grazie a Mauro Caselli dei Sempar in Baraca. È anche grazie a voi se sono potuta arrivata fino qui.

Maddalena Fumagalli, non so neanche da dove cominciare, sei stata preziosissima, un'amica storica, con cui spero di condividere altrettanta storia di quella trascorsa. Grazie per il Capodanno, per le scampagnate, il tuo folklore familiare, i viaggi e i mille tè e le mille tisane. Nonché i mille gatti. Grazie mille a Lucia Fiorani, ti ho conosciuta in questo periodo e la tua arte e creatività mi hanno molto ispirato. Mille grazie ad Alessandro Olmi, compagno di innumerevoli visioni di film, picnic vegani e letturine classiche. Grazie per le nostre gite estive ogni volta che torno ma soprattutto per la tua sensatezza ed empatia quelle volte che ci vedevamo a Bologna quando non era in zona rossa (mamma che ricordi) il primo anno del dottorato su Zoom. Andrea Lupo, mio mentore teatrale, ti meriti di essere annoverato tra i nomi qui perché ti voglio bene e me ne hai voluto tanto e questo mi ha sostenuto tantissimo negli anni.

This is in no way an exhaustive list and I am running out of time. If I have not mentioned you, it was a performance error, not a competence one! I will attempt to make

up for whenever we meet in person and reminisce about that time you said or did something that contributed to making this a more inspiring time. I want to thank from the bottom of my heart every single person placed on my path over the course of these past 5 years.

Last but far from being least: all Glory and Praise be to the One in Whose Hands is Authority over all Matters.

Table of Contents

List of Tables

List of Figures

Abbreviations

1	first person
2	second person
3	third person
AAE	anti agreement
ALL	allative
COMP	complementizer
CS	construct state
DEM	demonstrative
EMPH	emphasis marker
EP	epenthetic morpheme
F	feminine
FS	free state
FUT	future
INT	question particle/interrogative
M	masculine
NEG	negative

PL	plural
POSS	possessive
PROG	progressive
REL	relative
SG	singular
WH	wh-element

Chapter 1: Introduction

This dissertation is about children's acquisition of *wh*-questions in Tashlhiyt Berber (TB henceforth), an Afro-Asiatic language spoken by approximately 7 million people in southern Morocco. *Wh*-questions in this language are characterized by a clause-initial wh-phrase followed by a complementizer that introduces a construction that shares the structure of headless relative clauses (HRC) in the language. Moreover, when the gap in the relative portion is a subject gap, the verb surfaces with special morphology and default masculine singular agreement, which is known by the name of Anti-Agreement Effect (AAE henceforth, from Anti-Agreement Effect), optionally included in parentheses in item (1) below.

(1) [WH-phrase [Complementizer HRC-(AAE)]]

This is unlike how questions are realized in many of the languages that have been studied both in the syntactic and developmental literature such as English, French or Mandarin Chinese among others. This is where the necessity of investigating the acquisition of such a language lies. Diversifying the linguistic literature by studying the development of this typologically underexplored language could provide new evidence about the linguistic endowment of children and their capacity to acquire a first language. Here I will focus on two observations related to some properties of TB, (2a) and (2b), which will be needed to set up related learning challenges.

(2) a. Word order is not informative by itself about whether the extracted constituent is a subject or an object. Both types of extraction surface the same as WH(-NP) COMP V NP. This applies to *wh*-questions, relative clauses, and clefts alike.

b. The post-*wh*-word complementizer -a is mandatory in *wh*-questions as well as cleft constructions.

The fact in (2a) can be seen in the examples below, where (57a) is a subject *wh*-question and (57b) is an object *wh*-question.

(3) a. Manwa gitn ħifrxan a i-bus-**n** amuʃʃ?
WH in.them in.boys COMP 3SGM-kissed-**aae** FS.cat
'Which boy kissed the cat?'

b. Manwa gitn ħifrxan a i-bus **w**muʃʃ?
WH in.them in.boys COMP 3SGM-kissed CS.cat
'Which boy did the cat kiss?'

There are, however, two kinds of morphological evidence that can signal the structure of the dependency. The first piece of morphological evidence comes in the form of the AAE morphology. This is a process whereby subject-verb agreement is suppressed and the verbal agreement is marked by the default masculine singular prefix and the participial suffix *-n.* This occurs only when the subject has been extracted. Consequently both its presence and absence give unequivocal evidence for whether the WH does or does not correspond to the subject. The second piece of evidence comes from construct state, a form of nominal morphology. The Construct State (CS) is a marked form taken only by post-verbal subjects, whether in main clauses or dependent clauses. In the absence of AAE, this marking on post-verbal subjects can indirectly signal that the NP that was pronounced higher up in the question was the object, not the subject.

Here are the learning challenges in the language, related to points (2a) and (2b):

(4) There is a rule in Tashlhiyt Berber that bans sequences of identical vowels. Thus, the complementizer *-a* is potentially phonologically obscured in many contexts.

This is the case of the *wh*-word *ma* which means both 'who' and 'what'. It ends in *-a*, which means that in those questions, the obligatory COMP is inaudible, as I illustrate in (5) below.

(5) Ma ~~a~~ i-bus?
WH ~~COMP~~ 3SGM-kissed
'Who did he kiss?'

Compare this example with the examples in (57a)and (57b), which involve the *wh*-word *manwa gitn* 'which one' which renders the *-a* visible.

If a significantly large proportion of questions exhibit phonological deletion of the complementizer, then children might fail to identify the correct structure, or fail to learn the obligatory status of the complementizer.

(6) The distribution of AAE could be informative in the input only if specific conditions are met, because it shares some of its morphology with the masculine singular agreement prefix and some of it, the -n suffix, with masculine plural agreement.. AAE is merely an added -n suffix to the verb stem, but its prefix half takes the form of the masculine singular, as can be seen in (57a) replicated below as (7a) , compared with the masculine singular of the verb 'kissed' in (5) replicated below as (7b) .

(7) a. Manwa gitn ħifrxan a i-bus-**n** amuʃʃ?
WH in.them in.boys COMP 3SGM-kissed-**aae** FS.cat
'Which boy kissed the cat?'

b. Ma ~~a~~ i-bus?
WH ~~COMP~~ 3SGM-kissed
'Who did he kiss?'

Because AAE is so similar to other morphological forms such as the masculine singular prefix and the masculine plural suffix, there may be a possibility that children would fail to learn its role as an agreement feature impoverishment process that signals structure, which would force them to rely on less formal cues to the structure.

The question that now arises is, could (4) and (6) pose challenges to the acquisition of *wh*-questions in TB for first language learners?

In the case of (4), could there be a distribution of complementizers and *wh*-words in the input that leads the child to the wrong grammar where there is no complementizer? Towards this goal, it will be necessary to test children's knowledge of relative clauses as well to see whether there is a difference in the developmental timeline or in the children's performance when it comes to these constructions that underlyingly are based on the same mechanism but superficially look different due to the complementizer. I will be approaching this question by conducting a behavioral study that directly examines children's understanding of *wh*-questions and relative clauses. This behavioral study situates itself in the developmental literature by exploring previously untested ways in which a filler gap dependency is signaled, i.e. via anti-agreement and the construct state. In the case of (6) is the morphological signal present and detectable in a way that could support learning? If not, in what measure is it present? This is where the necessity of a corpus of ambient speech or caregiver-child interaction comes into play. Towards answering both (4) and (6), I built a corpus over 4 years that allowed us to gather as much data as possible about the

occurrence of Ā-constructions in the input (in child-directed speech as well as adult-directed or overheard speech), since they all have AAE in common. As we will see more in detail though an analysis of the corpus in Chapter 4, there is identity between *wh*-questions and focus clefts (they share the same obligatory complementizer) and, as I will argue, relative clauses are the basic Ā-construction that the focus clefts and *wh*-questions are derived from. Thus, any kind of evidence that shows that children know these constructions will be useful in the quest to understanding how and when children acquire *wh*-questions in TB. More generally , the corpus, which is the primary contribution of this dissertation, is about identifying the signal that children use in acquiring the grammar as well as the nature of the input more in general. We will be using the corpus to ask how clear the morphosyntactic properties of the language are and if they can be used to infer structure. Here is a three-way paradigm that exemplifies these parallels and identity, also instantiated in the fact that the AAE morphology is triggered in all three instances. In the threeway example below the instances are, in order from (8a) to (8b) , a subject *wh*-question, a focus cleft and a subject relative clause.

(8) a. Ma a i-ʃʃa-**n** ibrejn?
WH COMP 3SGM-eaten-AAE couscous.
'Who has eaten the couscous?'

b. Hanane a i-ʃʃa-**n** ibrejn.
Hanane COMP 3SGM-eaten-AAE couscous.
'It's Hanane who has eaten the couscous'

c. tafruxt lli i-ʃʃa-**n** ibrejn
FS.girl REL 3SGM-eaten-AAE couscous
'the girl who ate the couscous'

For example, to know what data children can rely on for the acquisition of the complementizer, both *wh*-questions (8a) and cleft constructions (8b) will be seen as sources of data about the complementizer *a-*. Therefore, instead of testing *wh*-questions and clefts separately, I will be testing children's knowledge of *wh*-questions and relative clauses in the behavioral study and I will be searching the corpus for instances of all three subtypes of Ā-constructions to have an appropriate picture of the input situation. I will be asking about the 'audible' nature of the complementizer and compare the proportion of *wh*-questions where the *wh*-word makes the complementizer opaque and the ones where the *wh*-word makes the complementizer transparent. I will then look at focus clefts, as they have the same complementizer as *wh*-questions with the goal of seeing what the distribution of opaque vs. transparent strings is there. Specifically, it is necessary to see what the proportion of fronted clefts with a clefted constituent that ends in *-a* followed by the *a-* complementizer is compared to the proportion of fronted clefts that make the *a-* complementizer audible. Making a comparison across constructions of their respective distribution is done with the purpose of seeing whether there are more instances of opaque fronted-element+COMP strings in *wh*-questions or in clefts. This in turn will help us answer the question of what information is available to the child as she is forming a picture of what constructions require a complementizer. If the input shows that a negligible subset of *wh*-questions have audible complementizers, but all focus clefts have an audible complementizer after the clefted element, could this data make the child believe that the complementizer is only obligatory for clefts and not for *wh*-questions? We could only form better learning hypotheses by looking for this distribution in the input.

The learning question here is whether children can learn that there is structural and

grammatical similarity across *wh*-questions, relative clauses and focus clefts in the language. The similarity across these three constructions could be signalled with the presence of the complementizer, whether *a-* or *lli* and that consistently, anti-agreement shows up across all three. What we want to make sure is true in the corpus is that there is a sufficient number of instances and evidence of each construction where the verb displays the AAE morphology when the subject is extracted, which would be positive evidence that these constructions have a feature akin to "Ā-ness" in common. If the input is too impoverished distributionally, our question would be: do children use prior knowledge of the abstract similarity across Ā-construction to help them draw inferences about structure that are hidden by the morphophonology? This question is the basis for the conclusion that the shared AAE is a signal for Ā-ness. Simply observing a shared feature is not by itself sufficient for the child to drive any inference. The inference is driven by the child's prior knowledge that Ā-constructions will all be alike. The obligatoriness of the complementizer and their presence in clefts could allow the child to learn that clefts are present in *wh*-questions even if the morphophonology makes the complementizers opaque in most *wh*-questions encountered in the input. This is based on the child having available an assumption that *wh*-questions and clefts share Ā-structure.

In response to these learning questions, I present a behavioral study on the comprehension of *wh*-questions and relative clauses and an analysis of the possible input through a purpose-built longitudinal corpus of child-directed speech in TB, the first of its kind.

A significant portion of the findings presented here are based on a novel corpus of child-directed Tashlhiyt Berber that is the first of its kind in the following respects. The corpus is longitudinal and it follows the development of the same child from age 2;03,26

to 4;08,5 and her sister who joins the recordings after birth a year later, from 4,07 until 1;10,28 of age. This corpus adds up to about 7 hours of recordings collected between December 2020 and April 2023. The recording spans a total of 11 recording sessions of an average duration of 40 minutes totalling 7982 adult utterances that were coded and analyzed. More details about the profile of the corpus and the background will be provided in Chapter 4, which is where I present the input-related questions and findings in greater detail. Secondly, the corpus was collected to ensure a maximization of the speech of the caregivers, whether directed to other adults in the environment (i.e. adult-directed speech or ADS) or children (CDS). There already exist corpora that were aimed at collecting data on Berber languages, but they do not lend themselves to the research questions formulated here because of the intent and research questions that guided their creation. In some cases, the researchers were more interested in the children as opposed to the caregivers' speech, transcribing the former more carefully and deliberately or because the researcher's focus was on another module of the grammar, as is the case for an existing but not consultable corpus on child-directed Kabyle Berber (Tabti, 2013), that is specifically tagged for sociolinguistic analysis. Relatedly, the transcription is designed to facilitate syntactic analysis. Although there already exists at least one corpus of parent-child interactions in Tashlhiyt Berber, the transcriptions have the two infants' babbling patterns as the main data if not the only data available for analysis (Lahrouchi & Kern, 2018).

The key lessons we glean from these two studies, concerning the learning points in (4) and (6) are:

(9) In the comprehension study reported in this dissertation in chapter 2, by age 4, children's knowledge of *wh*-questions and relative clauses is equivalent. Both *wh*-questions and relative clauses seem to be learned along a similar temporal trajectory. More precisely, the main result is that performance on subject extraction improves with age in the same way in RCs and *wh*-questions, which in turn suggests that children are treating them as structurally alike throughout the time period measured in the study. The construction type is not predictive of children's performance. However, the **extraction type** is. Children are better at subject extraction (the cases marked with the AAE morphology) than object extraction (where only nominal morphology can be relied on to know whether the extracted element is an object).

(10) In the corpus, the distribution of agreement morphology is such that anti-agreement is sufficiently salient in the input. This is due to the feminine singular morphological marker being the most frequent subject agreement marking on the verb. This incidence of the feminine singular morphology (animate) correlates with a high incidence of feminine referents in the discourse, meaning that this is the ideal context in which the occurrence of AAE (whose prefix is isomorphic with the masculine singular or default agreement) is more salient as a morphological marker. Additionally, of all Ā-constructions in the input, about 24% display anti-agreement morphology, evenly distributed with the exception of relative clauses which form the majority. The presence of anti-agreement across the three constructions is a positive sign in terms of signalling to the child that the three constructions have 'Ā-ness' in common. If children perform equally well with AAE in these constructions despite them having

different distributions in the input, then we have positive evidence of them being learned as being issued from the same underlying phenomenon instead of resulting from unrelated processes in the grammar.

(11) Moreover, the majority of *wh*-questions (91%) are of the kind where the complementizer is rendered inaudible by the phonological environment. When it comes to the audibility of the obligatory complementizer, out of the three Ā-constructions exemplified, clefts mostly show evidence of its presence. This is a noteworthy matter because it means that out of the relevant Ā-constructions, *wh*-questions are not a reliable source of information about the presence of the complementizer. This could mean that in the absence of the complementizer in the input, children could either hypothesize that the complementizer is not there or learn from the focus clefts that it might not be there because of a phonological rule, and not because it's a null morpheme.

These findings are rendered interesting by the implications that they have for the child's knowledge state. The explanation proposed here for (9) is that if children recognize the structural similarity between questions and non-interrogative clefts (which are formed through RCs), then children could use the data from non-interrogative clefts for instance, as evidence for the structure of questions i.e., the presence of an obligatory complementizer.

We cannot yet rule out with certainty that, based on the findings in (9), the children succeeded in the task other than by using the morphosyntax, because in general, nothing can be ruled out with absolute certainty. However, the conditions posed by the stimuli were particularly strict by design, in such a way that the child could only possibly rely on

the morphosyntax that she knows to perform well.

Further experiments will be needed to be more certain. But until then, we can put forward the hypothesis that, by age 3 or 4 at the latest, children acquiring TB have understood the grammatical role of AAE in both interrogative and relative clauses. The alternative hypothesis to this more formal explanation is that children succeed due to pragmatic knowledge on what is felicitously being asked of them given the context. The stimuli, however, were designed in such a way that there would be no favouring of an agent or a patient reading, so the idea that children would be relying on pragmatic cues to favour one specific reading is least likely to apply in this scenario. But if a pragmatic account is not accurate, then the experiment suggests that by age 4 at the latest, the issue in (6) about Anti-Agreement distribution's problematicity is no longer an impediment to the successful acquisition of the grammar of *wh*-questions.

Parallel to the question of what the picture in the corpus is, I also ask about the ways in which *wh*-questions, clefts and relative clauses share abstract structure. In Chapter 2, I will present some evidence that addresses questions about the syntax underlying the surface pattern of *wh*-questions in TB, represented in example (1).

The surface pattern is consistent with at least two distinct derivations. The first, (13) treats *wh*-questions as interrogative "Clefts", where the wh-phrase is the "Focus" of the Cleft. In English this would be something akin to: "Which type of couscous is what (or that which) Hanane likes?".

(12) [Man ibreijn][a tra Hanane].

(13) [[Which couscous] [HRC that wants Hanane]].

This first analysis, or **biclausal** analysis, involves two clauses: the clause in which the wh-phrase and the RC are dependents, and then the RC itself. The second analysis (14) , or the **monoclausal** analysis, has just one clause, at the edge of which is the wh-phrase. On this view, the syntax of TB questions is essentially identical to that of English interrogatives, except that the complementizer is mandatorily pronounced in TB and mandatorily silent in English (where the complementizer position is sometimes filled with an audible auxiliary, e.g., in object and adjunct extractions).

(14) [CP Which couscous [C that wants Hanane]].

My behavioral study and corpus study do not determine which analysis is accurate between (13) and (14). However,(13) is favored in extant analyses of Berber languages (Shlonsky, 1987; Stoyanova, 2008) and of many other VSO languages, in Celtic and Austronesian (Adger & Ramchand, 2005; Oda, 2008; Paul, 2001; Potsdam, 2009; Potsdam & Polinsky, 2011; Sheil, 2016). Nevertheless, the data must still be mentioned since the correct analysis, i.e., the adult representation, is the one children must ultimately converge on. Moreover, the two analyses raise different learning questions. For instance, if (13) is correct, the child must learn that the main clause never has an overt verb, overt tense, or any adjuncts, despite the biclausal structure. The two analyses may also have different implications for whether we expect "Reconstruction Effects" for the wh-phrase, where it is subject to grammatical restrictions on phrases in the position of the gap in the RC. As I will show in Chapter 2, in TB, wh-phrases specifically do not show Reconstruction with respect to the Principle C of the Binding Theory, whereas the silent operator *Op* heading RCs does.

The dissertation unfolds as follows. In chapter 2, I will be presenting evidence that TB *wh*-questions, relative clauses and focus clefts share abstract structure as well as the ways in which they share form. I will be doing so by showing the observable parallels in their construction firstly. Namely, the similarities between headless relative clauses and the non-wh-material in the *wh*-question and the parallels between focus clefts and relative clauses. The similarities will be shown by presenting the Anti-Agreement Effect in more detail, a phenomenon that only these three constructions have in common in the language. I will also be showing the differences between these constructions and the possible ways in which these differences could be hurdles for children as they generalize over them. This will be done by showing how the complementizer surfaces across these three constructions. Crucially, whilst the obligatory complementizer -a is shared by *wh*-questions and focus clefts, the relative C head in relative clauses surfaces as *lli* and it is not mandatory. Another difference that will be mentioned is the asymmetry of *wh*-questions and the silent relative clause operator *Op* with respect to Condition C Reconstruction effects. Although the implications of the connectivity effects on the biclausal hypothesis goes beyond the scope of the present dissertation, the data will be provided with the intent of sparking a future conversation on the matter. As I discuss *wh*-questions in TB, I will also describe the distribution of the Construct State, which constitutes additional data in the language that can inform the child about the gap site in filler gap dependencies. More concretely, I will be showing how it could be made use of to establish whether the child is hearing an instance of subject or object extraction.

The question that arises from Chapter 2 is whether children recognize the abstract unity of Ā-constructions in TB. Chapter 3 presents a behavioral study testing the similarity

and differences in the treatment of *wh*-questions and relative clauses by children. I find that children improve on subject questions and subject relatives at the same rate, suggesting that children may treat them as sharing structure.

In Chapter 4, which centers around the new longitudinal corpus, I address three questions. First, I ask whether the children's performance in the behavioral study of Chapter 3 is attributable to frequency effects in the input. We find that this is not the case and the constructions that children seem to perform best in are among the least frequent in the input. Second, can children use the data from the complementizer usage present in focus clefts to learn that the complementizer is obligatory in *wh*-questions? It appears that since the majority of *wh*-questions are of the kind that obscures the complementizer, and the majority of clefts show the reverse distribution, it is likeliest that the child generalizes over to *wh*-question complementizers based on focus clefts. In order to generalize from focus clefts to *wh*-questions, the child must already recognize their structural similarity. Finally, I ask what role anti-agreement plays in signalling the similarity across Ā-constructions in TB, showing that Anti-Agreement is mostly represented in the input in the form of subject *wh*-questions, followed by reduced subject relative clauses, subject clefts and subject relative clauses with an overt *lli* relative pronoun. I will show how there does not seem to be a significant difference across categories in terms of the distribution of anti-agreement with the exception of the relative clause category, depending on how the count is executed. This paints a balanced picture in terms of what kind of construction provides evidence of AAE.

Finally, the dissertation concludes with Chapter 5. The goal of the present dissertation is to present the empirical landscape, formally, structurally and quantitatively, of *wh*-questions in Tashlhiyt Berber, in order to formulate better questions about their struc-

ture and acquisition. Identifying the correct syntactic structure, although only partially addressed in this dissertation, can help formulate more abstract questions about what children's expectations about grammatical structure are, and how they link those expectations with their input in converging on adult mental representations of the grammar. Finally, I will be drawing conclusions that pull together the findings in the behavioral study and the corpus study, and I will make suggestions for possible follow-up studies that can be made to formulate targeted developmental questions.

Chapter 2: Structural and abstract unity of TB Ā-dependencies

Wh-questions, relative clauses and focus clefts share structure both on the surface and abstractly. They share form in the way that they display a complementizer, which is identical for *wh*-questions and focus clefts (*a-*) and different in the case of relative clauses (*lli*). They also share the same word order across the three constructions, whether the fronted constituent is a subject or an object, when the dependency has an argument as a filler. In terms of abstract structure, instances of subject extraction across these three dependency types trigger the Anti-Agreement Effect. This is also the one commonality that can be used in disambiguating word order in subject and object extraction instances, together with the Construct State. Finally, all three are instances of Ā-movement that show evidence of displacement via sensitivity to islands (Ross, 1967). This has been known as a feature of Ā-movement cross-linguistically since Chomsky (1977). I will be presenting these properties in turn with targeted examples in this chapter that is meant to set up the developmental question. The main developmental questions that this chapter is in service of are the following: Does the grammar give enough evidence of the presence of a complementizer across Ā-constructions, to make it possible to learn about their similarity? could children possibly make use of Anti-Agreement and the Construct State to learn to discriminate between subject and object extraction?

2.1 Structural identity

2.1.1 Word order similarity

There are some close surface similarities across relative clauses, focus clefts and *wh*-questions in TB, as can be seen in the upcoming examples

TB is a VSO language with head-initial relative clauses.This is the reason why, across the three constructions, both object extraction (15-17) and subject extraction (18-20) yield the same word order, which is disambiguated through verbal (anti-agreement), and nominal morphology (construct state) as we will explore in the upcoming subsections of this chapter.

Headed Object Relative Clause

(15) ibrejn [lli t-ʃʃa Hanane]
couscous [REL 3SGF-eaten Hanane.]
'the couscous that Hanane ate'

Object Focus Cleft

(16) ibrejn [a t-ʃʃa Hanane.]
couscous [COMP 3SGM-eaten Hanane.]
'Couscous is what Hanane ate'

Object *Wh*-question

(17) Ma [a t-ʃʃa Hanane?]
WH [COMP 3SGF-eaten Hanane]
'What is it that Hanane ate?'

Examples (18-20) are presented to provide a minimal pair with examples (15-17).

Headed Subject Relative Clause

(18) tafruxt [lli i-ʃʃa-**n** ibrejn]
FS.girl [REL 3SGM-eaten-AAE couscous]
'the girl who ate the couscous'

Subject Focus Cleft

(19) Tafruxt [a i-ʃʃa-**n** ibrejn.]
FS.girl [COMP 3SGM-eaten-AAE couscous.]
'It's the girl who ate the couscous (not someone else)'

Subject *Wh*-question

(20) Ma a i-ʃʃa-**n** ibrejn?
WH COMP 3SGM-eaten-AAE couscous.
'Who ate the couscous?'

In (18-20), instead of extracting *Hanane*, I will use 'girl' to satisfy the definiteness requirements of the use of RCs. These examples are furthermore designed to show that the word order is the same in S and O extraction for all constructions. These form a separate construction unlike left dislocation which does not trigger AAE and it's not derived through Ā-movement. These form a separate construction unlike left dislocation which does not trigger AAE and it's not derived through Ā-movement.

(21) Tafruxt, t-ʃʃa ibrejn.
FS.girl 3SGF-eaten couscous.
'As for the girl, she ate couscous'

2.1.2 Anti-Agreement Effect

As seen in examples (18-20), in TB, subject Ā-extraction leads to the suppression of subject-verb agreement (Ouali, 2011; Ouhalla, 1988, 1993). This phenomenon is known as the Anti-Agreement Effect in the literature. This agreement suppression is more visible in the case of feminine NPs. In the case of masculine feature bearing NPs, AAE takes the form of the addition of a morphological marker that supplements the existing default morphology, i.e. *-n*. AAE in Berber is therefore said to take the form of the circumfix i-X-n, marking the participle (Baier, 2018; Ennaji & Sadiqi, 2004; Ouhalla, 1993). Guerrsel and others posit that since the prefix *i-* is homophonous and homomorphic with the masculine agreement prefix, the verb adopts its default agreement form, to which the participial morpheme -n is attached.

While is is not necessarily a particularly salient property of syntax in TB, from a morphological and developmental standpoint, it is worth mentioning what changes the AAE morphology undergoes when the tense of the embedded clause is in the future. In verbs that carry the future prefix ra(d)-, the participial suffix that forms the AAE is dropped. This means that the future anti-agreement form simply takes the form of standard 3SGM agreement. Compare the examples below where you can see an instance of standard feminine singular subject-verb agreement followed by default agreement in the question and focus cleft in (b) and (c):

(22) a. Mam rakkm t-awi.
Mother.yours FUT.you 3SGF-take.
'Your mother will take you'

b. Ma a rakkm jawi?
WH COMP FUT.you 3SGM-take?

'Who is taking you?'

c. Mam a rakkm jawi.
Mother.yours COMP FUT.you 3SGM-take.

'Your mother is the one who is taking you'

This is worth noting since as far as I can tell it has not been documented in the extensive syntactic literature on antiagreement that started with Ouhalla's seminal work in 1989. It could also potentially have some interesting implications developmentally, because of its homomorphy with the masculine agreement. Thus if the extracted subject was masculine, it would look like a normally agreeing form.

2.2 Form Differences

2.2.1 The complementizer, optional vs. mandatory

Now that we have seen an exposition of the similarities, I will be showing one crucial difference that could potentially be a hurdle for children as they generalize over these constructions. Whilst the complementizer a- is obligatory in *wh*-questions and focus clefts, the relative pronoun *lli* can be omitted to form a reduced relative clause, a common adjectival construction in TB.

Here is an example where the *a-* is obligatory, using *manwa gitn* 'which one' though this pattern applies to all *wh*-words.

(23) *Manwa gitn hizʕdad i-ttbujus-n agru?

Which in.them of.FS.monkeys 3SGM-PROG.kiss-AAE FS.frog?

'Which monkey is kissing the frog?'

(24) Manwa gitn hizʕdad a i-ttbujus-n agru?

Which in.them of.FS.monkeys COMP 3SGM-PROG.kiss-AAE FS.frog?

'Which monkey is kissing the frog?'

Lli is however an optional relative pronoun, unlike the complementizer *a-*. Note that talkswat 'the dress' is feminine singular. Let us consider the simplex clause in 10 with standard agreement between the NP and the verb 'be red'.

(25) a. T-zwgga talkswat.
3SGF-is.red FS.dress.

'The dress is red'

b. Talkswat lli i-zwgga-n
FS.dress REL 3SGM-is.red-AAE

'The dress that is red.'

c. Talkswat i-zwgga-n.
FS.dress 3SGM-is.red.AAE

'The dress that is red'

It is perfectly acceptable in the language to pronounce (25c). This means that the child is highly likely to hear in the input a construction where the *lli* complementizer is optional. However, to complete the picture, let us consider some examples of reduced object relative clauses.

(26) a. Tiglaj lli ssnwiħ.
Eggs REL cooked.1SG

'The eggs that I cooked.'

b. *Tiglaj ssnwiħ.
Eggs REL

'The eggs I cooked.'

Reduced object relative clauses where the *lli* is omitted are unacceptable. Therefore, there is an asymmetry here for the child to learn. The complementizers *lli* and *a-* are moreover in complementary distribution. The complementizer which surfaces as *a-* in TB cannot co-occur in the same structure as the relative complementizer *lli* found in subject and object relative clauses, of which the object type is exemplified in (15-17). This suggests that whatever position *lli* occupies, it is likely to be the same as the one that a- occupies.

(27) lhanut lli-s tdda
grocery.store REL-to went3SGF

the grocery store that she went to

(28) Lhanut a-s tdda.
grocery.store COMP-to went3SGF

It's to the grocery store that she went.

The cleft marker which surfaces as *a-* in TB cannot co-occur in the same strucure as the relative complementizer *lli* found in subject and object relative clauses, of which the object type is exemplified in 27. This suggests to us that whatever position *lli* occupies, it is likely to be the same as the one that *a-* occupies.

(29) *lhanut lli a s-tdda
grocery.store REL COMP ALL-went3SGF

The grocery store is what she went to

(30) *lhanut a lli s-tdda
grocery.store COMP REL ALL-went3SGF

The grocery store is what she went to

2.2.2 Ban on identical vowel sequences

An element that complicates the picture for *wh*-questions is that the complementizer *a-*[1] is not pronounced in specific phonological contexts. Like in many Berber languages, adjacent vowels are not tolerated in any morphosyntactic environment (Dell & Elmedlaoui, 2002, p. 189). In TB, multiple phonological rules conspire against a sequence of two consecutive vowels. Three strategies tend to be used across varieties of TB to prevent this occurrence.

1. epenthetic yod insertion.

2. coarticulation rule in which two consecutive vowels are pronounced as one long vowel.

3. one of the two vocoids are realized as a glide.

Dell & Elmedlaoui point out the case of surface aa sequences that arise as a consequence of derivational morphological processes, which I present merely by way of illustration of the complexities of this language's phonological system. The concatenation of morphemes within a single prosodic unit never creates [a+a] sequences. Let us now consider the following example

[1]This complementizer tends to be referred more theory-neutrally in the literature as a cleft-marker (Shlonsky, 1987; Stoyanova, 2008). For an independent test of the CP-hood of clauses headed by *a-* please refer to the excursus on the topic in Chapter 5

(31) ntta a(d) i-ra-n.

he COMP 3SGM-want-AAE.

It is he who wants / He's the one who wants.

This string can acceptably surface as nttayran in fast speech (Dell & Elmedlaoui, 2002). Moreover, since this example is a cleft construction, the emphatic focused pronoun **ntta** receives the main stress. The authors do not clarify whether it is possible to tell which of the two a segments surfaces instead, but since the stress shows up on the explicit pronoun, it is likelier that the a that gets deleted is the one in unstressed position, i.e. the complementizer.

This is relevant to our question of language acquisition because some of the *wh*-words end in *-a* which could mean that for some of the *wh*-questions, the a- complementizer is pronounced or audible, and for others it is not. In order to formulate accurate hypotheses about what the knowledge state of the child is, as well as of what the input is like, we would need to classify *wh*-words based on the ones that make it easier to learn that complementizers are obligatory vs. ones that do not. We will be considering the following *wh*-words in this dissertation, summarized in the chart below. The *wh*-words that end in a sound other than -a, thus making it possible for the complementizer to be pronounced, will be forming 'transparent strings' for the purposes of the acquisitional questions. Conversely, *wh*-words that end in -a (as well as the exceptional *mani*) will be forming 'opaque strings' with *wh*-questions that surface as though the complementizer was optional.

2.3 Island Effects

Resuming the discussion of the similarities across *wh*-questions, relative clauses and focus clefts, it is worth testing whether the commonalities go as deep as the more abstract properties of these constructions. Wh-movement is island sensitive. If TB presents some form of Ā-movement across *wh*-questions, relative clauses and clefts, then we would expect island sensitivity across all three. In TB, as can be seen from example (32b) below, it is not possible to question out of a relative clause. The examples below, atapted from Potsdam's work on Fijian wh-movement (2009) are minimal triplets where a) is the non-wh counterpart of the b) instances. The b examples constitute the island violations that arise upon extraction of the relevant constituent. The c) example is an attempts at clefting out of the relevant construction.

Relative Clause Island

(32) a. Zriħ afrux lli-d-ibusn amuʃʃ.
Saw.1SG FS.boy REL-EP-3SGM.kissed-AAE FS.cat.
'I saw the child who kissed the cat.'

b. *Ma tzrit afrux lli-d-ibusn ?
WH 2SG.saw.2SG FS.boy REL-EP-3SGM.kissed-AAE ___?
*'What do you see the child who kissed?'

c. *Amuʃʃ ad zriħ afrux lli-d-ibusn
FS.Cat COMP saw.1SG FS.boy REL-EP-3SGM.kissed-AAE
*'It's the cat that I saw the child who kissed'

Example (32c), which is an instance of Complex NP Constraint, shows an illicit

extraction of a focus constituent from within the same RC. This is further evidence that movement out of a RC is blocked, as is expected for Ā-constructions.

Adjunct Islands

Here is an example of an adjunct island with a clefted phrase being extracted out of the adjunct clause *[llih issnwa]*'after he cooked'.

(33) a. T-ffuɣ [lliħ i-ssnwa imnsi]
3SGF.left [after 3SGM.cooked dinner.]
'She left after he cooked dinner'

b. *Imnsi a t-ffuɣ [lliħ i-ssnwa]
Dinner COMP 3SGF.left [after 3SGM.cooked]
*'It was dinner that she left after he cooked'

TB focus clefts pattern with wh-movement in the same way that relative clauses do.

2.4 Reconstruction Effects

One seemingly puzzling aspect that characterizes TB is the asymmetry between *wh*-questions and relative clauses with respect to connectivity effects. We will now go over this fact to show another respect in which Ā-constructions do not pattern alike in the language.

2.4.1 Reconstruction by way of Principle C violation

Reconstruction is the mechanism by which a syntactically moved constituent is interpreted in its pre-movement position. I use precedence here as a metaphor, to avoid suggesting that there is temporality to the mental representation of movement. Reconstruction effects arise

when an expression is connected to two different syntactic positions, one where the expression is phonologically realized and one where it is visible to some grammatical principle (Krifka & Schenner, 2018, p. 6). Principle C of binding theory states that an R-expression is free everywhere, i.e. it must never be A-bound (Chomsky, 1981; Reinhart, 1983). To test whether a constituent is interpreted, we look at speakers' judgments about co-reference between a pronoun and an R-expression in its domain. If there is reconstruction, we should see a principle C effect (the R-expression is bound). If there is no reconstruction, then we expect no principle C effect.

As a preliminary step before using Principle C violations (or a lack thereof) as diagnostics for movement, let us verify that Principle C effects arise independently in TB.

(34) Issn$_i$ is tgzmt ttswwera nKhalid$_{*i}$.

3SGM.know INT 2SG.tore.2SG photo of.Khalid

(He$_i$) knows that you destroyed a picture of Khalid$_{*i}$.

In this example, it is not possible for the null pro in the matrix clause to corefer with the R-expression Khalid in the embedded clause. In this configuration, the matrix null pro c-commands the R-expression Khalid, thus constituting a violation of condition C.

Some further evidence of the independent presence of condition C in TB can be observed in the four-way paradigm below, where we can look at backwards anaphora and cataphora to test condition C effects.

(35) a. Iknna iffoɣ$_{i/j}$, ar-ikmmi Amir$_i$ hbrra.
As.soon.as 3SGM-left, PROG-3SGM-smoke Amir outside.

'As soon as he$_{i/j}$ left, Amir$_i$ smoked a cigarette outside.'

b. Ar-ikmmi Amir$_i$ hbrra, iknna iffoɣ$_i$.
PROG-3SGM-smoke Amir outside as.soon.as 3SGM-left.

'Amir$_i$ smoked outside as soon as he$_i$ left.'

c. Ar-ikmmi$_{i/*j}$ hbrra, iknna iffoɣ Amir$_j$.
PROG-3SGM-smoke outside, as.soon.as Amir 3SGM-left.

He$_{i/*j}$ smoked outside, as soon as Amir$_j$ left.

d. Iknna iffoɣ$_i$ Amir$_i$, ar-ikmmi$_i$ hbrra.
As.soon.as 3SGM-left Amir, PROG-3SGM-smoke outside.

'As soon as Amir$_i$ left, he$_i$ smoked outside.'

The critical example out of this four-way paradigm is (35c). If the R-expression [Amir] is blocked from being coreferential with the matrix pronominal subject in c, but can be coreferential with the non-c-commanding pronominal in d, then that demonstrates the presence of principle C in the language. This is borne out.[2]

Now that we have established our independent generalization, let us see what co-reference patterns tell us about the base position of the wh-phrase.

For the purposes of reconstruction, we want to see whether the wh-phrase behaves as though it were in a position identified in the four-way paradigm of examples, namely, whether it cannot be bound by an antecedent null pronominal in a c-commanding relation to it. If the wh-phrase acts like it is bound, that would suggest that the fronted wh moved to its pronounced position from a lower argument position. This is how reconstruction

[2]The judgments in 35 were elicited by first providing each sentence separately as context and then asking the question "Who left?" in the case of (35a) and (35c) and "Who is smoking" in the case of (35b) and (35d). The possible answers that were provided were then summarized by means of the coindexation indices.

would be taken as evidence of movement.

Let us consider the following example:

(36) Issn Amir$_i$ manta gitnt h ttsawwr nKhalid$_j$ a-ihml$_{i/j}$.

3SGM.know Amir which in.them in photos of.Khalid COMP-3SGM.like.

Amir$_i$ knows which picture of Khalid$_j$ he$_{i/j}$ likes.

Both the interpretations in which Khalid and the embedded position subject of the embedded clause are coindexed, and the interpretation where they are not, are acceptable.

Nevertheless, there is a preference for Amir being the antecedent of the null pronoun in embedded position. To test whether this preference arises due to pragmatic or other factors unrelated to grammar, I elicited the judgments of the same sentence with a feminine (gender mismatched) matrix antecedent. In this case, Khalid can be interpreted as coreferential with the embedded null pro without the subtle issue that arose in the previous example.

(37) Tssn Malika$_i$ manta gitnt h ttsawwr nKhalid$_j$ a-ihml$_{k/j}$.

3SGF.know Malika which in.them in photos of.Khalid COMP-3SGM.like.

Malika$_i$ knows which picture of Khalid$_j$ he$_{k/j}$ likes.

The judgment for this sentence is that the embedded clause null *pro* could be Khalid or it could be someone else entirely. Since coreference between Khalid and the silent pro in the embedded clause is possible and available, it follows that the R-expression Khalid in the constituent is not bound by anything anywhere, thus enabling principle C to be satisfied in Tashlhiyt Berber 'picture of' constructions. This would, in turn, suggest that

the wh-phrase, [manta gitnt h tsawwr n Khalid], is not reconstructed in the argument position.

This data points towards the wh-phrase not moving to the position it is pronounced in in the language, which is in line with the hypothesis that the wh-phrase is base-generated in sentence-initial position, and does not move from a lower position. Upon testing a relative clause, however, we find a different outcome. The example below is an object relative clause modifying the object of the matrix clause.

(38) Izra Amir$_i$ ttswwira nKhalid$_j$ lli-d-iħml$_{i/*j}$.

3SGM.saw Amir photo POSS.Khalid REL-EP-3SGM.like.

Amir$_i$ saw the picture of Khalid$_j$ that he$_{i/*j}$ likes.

The interpretation where the nominal-embedded R-expression [Khalid] and the silent masculine singular pronoun are coreferential is blocked. If only the matrix R-expression [Amir] can be referred to, this suggests that Amir binds the silent pronoun in embedded position. Khalid however cannot be coreferential with the pronoun. In its hypothesized base position, as the argument of *likes*, Khalid would be within c-commanding range of that pronoun. Thus, if Khalid and the pronoun were coindexed, that would constitute a principle C violation.

The anti-reconstruction of the wh-phrase in TB is consistent with the hypothesis that TB *wh*-questions are not movement-based, i.e. the wh-phrase does not move to the position in which it is pronounced, but it was generated there. Considering however how the relative clause head does reconstruct, this asymmetric connectivity picture appears puzzling. One way of possibly making sense of these results is by looking at some extant

analyses of V1 Austronesian and Celtic languages that typologically share a non-negligible number of properties with TB [3]. Reconstruction effects arising within the relative clause only appears consistent with the analysis of *wh*-questions in Malagasy for instance where the only detectable Ā-movement is found in relative clauses but not for *wh*-words (Paul, 2001; Potsdam, 2009; Potsdam & Polinsky, 2011). Although the implications of these findings are beyond the scope of the present dissertation, the sources of the asymmetry in TB in particular are worth exploring in the future.

The reader is encouraged to refer to Potsdam's work on Fijian wh-movement (2009) which includes a detailed review of the results of wh-movement tests across Polynesian languages doing an in-depth typological survey the type of which is unearthing similarities that connect the sentence-initial base-generated nature of *wh*-words to what has been observed in Celtic languages as well (Adger & Ramchand, 2005; Oda, 2008; Sheil, 2016, inter alia).

[3]TB is a language that presents the following characteristics, which it shares with Irish as studied by Oda (2008) and some Austronesian languages like Maori, Niueian and Malagasy (Potsdam, 2009; Potsdam & Polinsky, 2011):

1. VSO canonical word order
2. Consistent predicate-first word order, in non-verbal clauses too.
3. Ā-related phenomena in the form of Anti-Agreement
4. VP raising analysis of the word order (Ouhalla, 1988; Ouhalla et al., 1994)
5. Overt complementizers that follow the wh-phrase

2.5 Construct State

An important component to complete the picture on how to disambiguate between subject and object extraction is the construct state morphology. The construct state is a special marking that shows up on subjects in postverbal position. TB is a VSO language, so the fact that the subject has construct state morphology on it when it's in canonical position can be thought of as a way of signalling where the subject NP belongs before any special movement has taken place. All other types of NP are in the free state, except for the complements of specific prepositions and the NPs selected by numbers. Here is a schema of the CS alternation. Masculine singular nouns that begin in *a-* have the first sound undergoing a change to *w-*.

	FS	**CS**	**English**
FEMININE	taglitt	→ taglitt	queen
	tafruxt	→ tafruxt	girl
MASCULINE	agllid	→ wgllid	king
	afrux	→ wfrux	boy
	azru	→ wzru	stone

Table 2.1: Construct State (CS vs. FS) alternations for masculine and feminine singular

As can be seen from the table (2.1) above, feminine singular nouns that begin in ta- show that the first vowel undergoes deletion in the construct state.

(39) a. T-ffuɣ [lliħ i-ssnwa imnsi]
3SGF.left [after 3SGM.cooked dinner.]
'She left after he cooked dinner'

b. *Imnsi a t-ffuɣ [lliħ i-ssnwa]
Dinner COMP 3SGF.left [after 3SGM.cooked]
*'It was dinner that she left after he cooked'

All other nouns that begin with a different sound, usually borrowings from other languages are opaque and do not show a clear CS marking. Moreover, pronouns also do not display the construct state transparently. Another instance of opacity in this respect is the plural agreement marker. In the case of the masculine, the plural is marked with i-X-n which remains unchanged and the same goes with the feminine plural circumfix ti-X-n. This means that the CS is a disambiguating marker only in the specific context of singular nouns, as well as a narrow subset of noun phrases.

The question that thus arises is whether this information in the form of AAE and the CS is available in the input in such a distribution that allows the child to learn the structure of subject and object Ā-extraction. This is important information on the way to the complete acquisition of *wh*-questions in the language.

(40) Azʕdud lli i-ttħbu-n a-gru

FS-monkey REL 3SGM-covering-AAE FS-frog

'The monkey that is covering the frog.'

(41) Azʕdud lli i-ttħbu **w-**gru

FS-monkey REL 3SGM-covering CS-frog

'The monkey that the frog is covering.'

As we can see, when a subject DP is in its canonical postverbal position, it takes the CS form, or the marked form, here in the embedded position in (41). However, when the subject is in its preverbal position upon extraction, it is in the Free State form, as you can see in (40). This is also the form of the Object, whether it is in its usual post-subject gap

position in (40), or in its pre-verbal position upon extraction in (41), the object relative clause.

Before moving to the next section, it is worth pointing out that the CS in TB appears when certain types of prepositions take DPs as complements. I will illustrate this with the table below. The prepositions in parentheses are the ones originally listed in Guerssel's description of the sentences as they are in his variety of Berber (Guerssel, 1992). The ones I use in the examples are from TB, which is more relevant to the study at hand.

In the figure (2.1) above, you can see how the same pattern as in the structural CS arises, i.e. that the *a-* or *i-* prefix that is not part of the root gets elided in the feminine, hence the strikethroughs in the dative, inessive and elative examples. In the masculine counterparts, the vowel prefix is replaced by the bilabial proximant w, which is exemplified in the instrumental, comitative, genitive and allative.

I highlighted the genitive to show that this is one instance in which Berber languages overlap with Semitic languages in one use of the Construct State form, due to the evocativeness of the term 'construct state'. This is however not the kind of CS that we are concerned with then it comes to the corpus and the children's knowledge. We want to know whether they can make use of the information about the position of the subject. Therefore, the CS marked on the complements of special prepositions will not be relevant in this pursuit. We will still look for these instances in the corpus in Chapter 4 and compare the proportion of PP object CS to structural CS to show where the evidence comes from.

To summarize, here is table (2.2) with the set of instances in which the CS and the FS arise. The FS is the 'elsewhere' condition. The only relevant CS instance, as mentioned in this subsection, is the VSO subject DP marking.

Case	Item	Examples
instrumental	**s**	s-**w**kʃʃud INS-wooden stick With the wooden stick
dative	**i**	i-t**ɐ**fruxt DAT-girl To the girl
comitative	**d**	d-**w**frux COM-boy With the boy
genitive	**n**	tammnt n-**w**glif honey GEN-honeycomb Honey from the honeycomb
superessive/ inessive	**ħ (x-dy)**	fl-ħ-t ħ-t**ɨ**gmmi left-1SG-pro IN-home I forgot it at home
elative	**mn (zy)**	n-uʃkæd mn-T**ɨ**znit. 1PL-arrived ELA-Tiznit We arrived from Tiznit.
allative	**dar (ɣer)**	ar dar-**w**madl-ann. until ALL-mountain-DEM all the way to that mountain.

Figure 2.1: TB prepositions that take CS nominals as complements

CONSTRUCT STATE	FREE STATE
VSO subject DP	VSO Object DP
complements of some prepositions	Left-dislocated DPs
	Ā'-extracted DPs
	complements of some prepositions

Table 2.2: Distribution of Berber CS

An important point about the construct state now is to ask how frequent are the vowel-initial nominals in the input, compared to the consonant-initial nominals? And most

critically, how do the frequencies interact and how much data on the visible CS alternations learners would be able to make use of in their intake? The presence of this special marking in post-verbal position subjects can provide a helpful signal to analyze the preverbal DP as an object in the case or object relative clauses. We will revisit these crucial questions in Chapter 4.

2.6 Conclusion

We have seen an overview of TB grammar that can help us identify better the developmental puzzle that the children acquiring the language are facing. TB subject *wh*-questions and object *wh*-questions show the same word order that can however be disambiguated through anti-agreement morphology and the construct state. However, (1) anti-agreement morphology is isomorphic to masculine singular morphology and in the future tense (2) anti-agreement morphology can be identical to the masculine morphology. Moreover (3) only a subset of NPs can display overt CS morphology, which means that the signal may be present only if a specific distribution is present in the input. In the case of issue (3) the construct state would be a useful signal only if the input shows a sufficient occurrence of nouns that show the marking transparently in order to draw the inference that the CS has to do with the position of the subject along the clausal spine.

ldren know through the behavioral study in Chapter 3 and what children encounter in the input through the corpus presented in Chapter 4.

The subsection about the differences is meant to show that although *wh*-questions, relative clauses and focus clefts are the same construction underlyingly (as we will see,

instances of Ā-movement), some differences could potentially pose problems as the child is learning to make generalizations across these constructions. Although the a- complementizer is obligatory for *wh*-questions and relative clauses, the *lli* complementizer is optional. A problem would arise if the child were to only hear opaque wh-strings and reduced subject relative clauses and infer that the optionality is shared by both construction types. If the child however learns to track what the possibilities are, the issue issue might not arise. Although the child may hear both subject RCs where the relative pronoun is pronounced and omitted, the child would never encounter instances of ma where the complementizer follows.

In many of the languages that share the typological syntactic profile of TB, *wh*-questions have been analyzed to involve a biclausal structure where the wh-phrase portion constitutes an independent clause that takes a relative-clause-like 'remainder' as its complement. There are ways in which TB could be one of these languages and Chapter 5 is dedicated to exploring this possible analysis. However, it appears thus far that the learning questions raised would hold independently whether the language shows a biclausal profile to *wh*-questions or a monocloausal one.

Chapter 3: The comprehension of *wh*-questions and anti-agreement

This chapter is ultimately meant to answer two questions that were raised in Chapter 1 and 2. 1) Do children recognize the abstract unity of Ā-constructions in their language? If that is the case, we would expect them to perform equally well across *wh*-questions, relative clauses and focus clefts, or in comparable ways. If we were to design a study where we test one of these constructions per condition and we were to run a statistical analysis, then construction type would not be a predictive factor. 2) Do children know the morphology that will allow them to distinguish subject *wh*-questions from object *wh*-questions? If they do, then we would expect them to perform well in the subject condition. The predicted picture for the object condition is more complicated and that will be discussed in the portion of this chapter dedicated to the parser.

3.1 The study

As foreshadowed in Chapter 1, the present study situates itself in the acquisitional literature by exploring previously untested ways in which a filler gap dependency is signaled, i.e. via anti-agreement and the construct state. In the psycholinguistic literature, the closest phenomenon that has been studied is wh-agreement. One such example is an adult psycholinguistic study on wh-agreement in Chamorro (Wagers et al., 2015), which inves-

tigates the role of this property in the processing of subject and object Ā-extraction. In this study, the authors found that the presence of wh-agreement seemed to have a facilitatory effect on the processing of some wh-constructions, which was less observable when wh-agreement was not present. This is relevant under the unifying account that conceives of anti-agreement as a typology of wh-agreement (Baier, 2018). In the presence of anti-agreement, the speaker who has full competence in the grammar gets evidence that the extracted NP is a subject, and when anti-agreement is absent, the child would get two sources of evidence of the extracted NP being an object, i.e. subject-verb agreement on the verb and the construct state on the embedded NP, which is a feature that only subjects in post-verbal position display. Let us consider below the *wh*-question version of a relative clause example previously presented.

(42) Man afrux [a i-bus-**n** amuʃʃ]?
WH boy [COMP 3SGM-kissed-**aae** FS.cat].
'Which boy kissed the cat?'

(43) Man amuʃʃ [a i-bus **w**frux]?
WH cat [COMP 3SGM-kissed **cs**.boy].
'Which cat did the boy kiss?'

Example sentence (42) is an instance of subject *wh*-question, with the signature AAE morphology on the verb. Example sentence (43) is an object *wh*-question where subject-verb agreement are intact and the subject is marked by the CS, showing that the object was extracted in this instance.

The complementarity of these cues and the presence of multiple cues in the case of the longer dependency (ORCs and object *wh*-questions) make this language an interesting testing ground. This study asks whether TB-acquiring children are sensitive to this richness of signals to the gap position and whether this could make TB-acquiring children perform well in the comprehension of filler-gap dependencies at the ages involved. We tested this knowledge using relative clauses and *wh*-questions where the NPs in extracted and embedded position are both fully lexical (as opposed to indexical) and animate in every condition, in the form of a pointing task paradigm for older children adapted from a preferential looking task by Perkins and Lidz (2020).

3.1.1 A few words on the subject-object asymmetry

Though it will not be an essential component of this dissertation's discussion of *wh*-questions, it is still worth mentioning a few words on the subject-object asymmetry literature (See Lau and Tanaka 2021 for a comprehensive review).

There is a divergence of opinions about the developmental timeline of filler-gap dependencies, based on the gap site. Some posit that subject $\bar{A}$-extraction is acquired first and object $\bar{A}$-extraction only subsequently (Friedmann et al., 2009). Some posit that they are learned simultaneously (Stromswold, 1995). In an analysis of the corpora of 12 English-speaking children, contrary to previous reports (Ervin-Tripp, 1970; Tyack & Ingram, 1977), object questions were found to be produced at the same age as subject questions (Stromswold, 1995). The other logical possibility would be for object $\bar{A}$-extraction to be acquired before subject $\bar{A}$-extraction, but there are no strong proponents of it (Lau &

Tanaka, 2021).

In the few more recent studies, included in the overview above, and which probe the knowledge of filler-gap dependencies by infants, the subject-object asymmetry doesn't arise for 20-month-olds. The underlying mechanisms for *wh*-questions and relative clauses should already be in place, but the differential performance, i.e. their RC-specific dip - might be due to surface features of relative clauses that make the dependency less clear than is the case for *wh*-questions. This struggle with the shape of the RC complementizer is compatible with a scenario where the children represent adult-like filler-gap dependencies. The position that I maintain and the assumption on the matter throughout this dissertation is that the subject-object asymmetry is a deployment issue at its core.

The contribution of the parser

I will be outlining what I assume to be the bounds of the parser for the remainder of this chapter. It has emerged from previous literature that the child parser shares the property of processing incoming speech incrementally with the adult parser. Children parse speech and resolve syntactic ambiguity incrementally (Trueswell et al., 1999, for a review see Omaki and Lidz, 2015). This goes hand in hand with an active filling strategy, for instance. On the other hand, unlike the adult parser, children have been shown to struggle with revising an initial misparse (Lidz et al., 2017; Pozzan & Trueswell, 2016; Trueswell et al., 1999) possibly due to immature domain general cognitive abilities, i.e. executive function (Novick et al., 2005; Ovans, 2022). This entails that they cannot integrate incoming information that acts as a cue to disambiguate between multiple possible interpretations, unlike adults (Omaki & Lidz, 2015, p. 163). The adult parser relies on mature control processes that can inhibit

multiple competing interpretations and go with the one that is most relevant to the context and most compatible with the available cues (Novick et al., 2005, p. 273). To summarize, a) the child parser assigns form and meaning representations to the incoming speech stream incrementally and b) it does not rely on mature inhibitory control mechanisms, so it's stuck with the first interpretation that is assigned during incremental parsing, even after being presented with disambiguating cues. Assuming an active filling strategy, where the hearer attempts to assign a filler to the earliest gap in the dependency, we would expect children to perform well when attempting to assign an extracted subject to its gap and conversely perform worse when assigning an extracted object to its further away gap in the dependency.

3.1.2 Methodology

The comprehension of subject and object RCs and *wh*-questions was tested using a pointing task paradigm. The dependent measure in the study consisted of pointing to the agent of the action in the trial scene visible in the video. Due to the age range of the children we chose to involve (older than 24 months), this was a more effective measure. Finer-grained measures, such as eye-tracking, are not necessarily more precise for this specific population when dealing with a split screen paradigm (two videos simultaneously on the screen) as there may not be reliable correspondence between the behavior of the child and their attention state at a given moment (Crain & Thornton, 1998)[1].

[1]One possible critique that could be advanced against this design choice is that the measure was not fine-grained enough to establish whether children meant the agent or patient in the side of the screen that they would point at. Some researchers have brought up the importance of also identifying the exact participant the child is pointing at even within the target-like side of the screen, to see if there is a clear agent-patient distinction (Arnon, 2005; Lau & Tanaka, 2021). However, this design requirement is based on the assumption that the child's pointing is deliberately directed at the precise participant they understand

Two factors were considered for this experimental design, namely construction type and extraction site. The two levels for the first factor were relative clause and *wh*-question, whereas the two levels for the extraction site were subject and object. This was also reflected in the four main conditions that children were tested on. The design was between subjects.

Here are exemplified the materials by condition.

Subject Condition

(44) a. Mliji-za azʕdud lli i-ttħbu-**n** agru.
Show.me-EMPH FS.monkey REL 3SGM-PROG.cover-AAE FS.frog.
'Find the monkey that's covering the frog.'

b. Manwa gitn ħizʕdad a i-ttħbu-**n** agru?
Which in.them of.FS.monkeys COMP 3SGM-PROG.covering-AAE FS.frog?
'Which monkey is covering the frog?'

Object Condition

(45) a. Mliji-za azʕdud lli i-ttħbu **w**gru.
Show.me-EMPH FS.monkey REL 3SGM-PROG.cover CS.frog
'Find the monkey that the frog is covering.'

b. Manwa gitn ħizʕdad a i-ttħbu **w**gru?
Which in.them of.FS.monkeys COMP 3SGM-PROG.cover **cs**.frog?
'Which monkey is the frog covering?'

to be the answer to the instruction, which may not always be the case necessarily.

The complementizer which surfaces as *a-* in TB cannot co-occur in the same structure as the relative complementizer *lli* found in subject and object relative clauses, of which the object type is exemplified in (15-17). This suggests that whatever position *lli* occupies, it is likely to be the same as the one that a- occupies.

Some of the design considerations of this study were informed by what has previously helped children comprehend Ā-dependencies without an asymmetry across the board (Gagliardi et al., 2016; Perkins & Lidz, 2020) due to the following methodological precautions: the inclusion of multiple verb trials to allow the child to get at ease with the task. A between-subject design was also implemented to prevent interference from the requirements of different conditions. Felicity conditions for the use of a relative clause or a *wh*-question were also respected, in line with the seminal literature on what constitutes a felicitous use of a relative clause or *wh*-question (Crain & McKee, 1985).

In this study, the dependent measure was the instances of the act of pointing towards the agent divided by the number of trials, which corresponded to separate verbs. For each stimulus, participants would be given two opportunities to respond via pointing. If, after the repetition, children pointed to a different character from their first attempt, then that would be considered their final choice. The thought behind this was that by the time the second opportunity to respond would arise, children had had more exposure to the stimulus and were potentially able to make a more confident decision.

3.1.3 Participants

A total of 85 (41 male) neurotypical toddlers and preschoolers were involved, between the ages of 2;3;26 and 5;0;27. The mean age was 3;6;26.

The best-represented group of children in each condition was the middle range, between 36 and 48 months. 16 children were in the 24-36 month range (5 M), 52 children were in the 36-38 month old range (26 M), 17 children were in the 48-62 month range, (10 M).

Pooling all ages together, the following are the numbers of children per condition. OQ: 22 (10 male), OR: 20 (17 male), SQ: 21 (11 male), SR: 22 (13 male). The more represented age group in each condition is the range between 36 and 48. The following is the breakdown of their specific pool size. OQ: 14, OR: 10, SQ: 13, SR: 14. The starting number of participants before exclusions was 104. A total of 11 children were excluded due to a bias towards one side of the screen. An additional 2 children were excluded for two reasons: we did not manage to find details about their age but they also happened to mostly have non-valid responses, i.e. they would not provide a response to the question for more than half the trials. Finally, 4 were not included in the final analysis as they did not engage with the task at all. The somewhat high exclusion rate was due to several possible factors, for instance the unfamiliarity of some of the children with the equipment, novelty arising from interacting with a stranger in the community as the experimenter was, or lack of understanding of the instructions.

Participants were recruited in Morocco throughout the summer of 2022, between early June and late July. In the first month, the children were recruited from a total of

12 schools and kindergartens selected from the Chtouka-Aıt-Baha province, a region where the predominant language spoken is TB[2]. In the second month, which was after schools closed and the summer vacation began, I recruited children in their homes. A total of 6 towns and villages in the proximity of Aıt Milk, Ait Melloul, and Biougra were visited.

One of the main requirements for participation and inclusion was for the children to be exposed to TB at home, more specifically that both parents speak TB to each other and to their child. The Souss-Massa region in southern Morocco is a historically Tashlhiyt-dominant region. However, work opportunities in the rising south have made it possible for Moroccans from the Arabic-dominant areas to move there. Therefore, it has been increasingly common in the past ten years to find children whose first language is not a Berber one in the formerly Berber-majority regions of Morocco. When running the study, I did not involve children whose parents speak Arabic to them at home, even when their kindergarten teacher is a Tashlhiyt-speaker. The language profile of each child was established when conversing with either parent or with the child's nursery teachers acting in loco parentis. It sometimes occurred that children were sent to take part in the experiment even when they did not speak TB. Sometimes, however, some teachers would provide a mixed list of registered pupils including children who were exclusively Arabic-speaking. In those cases, I would generally find out that the children did not speak TB even before the beginning of the experimental session, since they would not be responsive to the ice-breaker games fully led in this language. We obtained from IRB an exemption from

[2]Ecole primaire Hassan Ben Mbarek in Ait Izza, Ajial Almustaqbal in Douar Taous as well as association-led pre-kindergartens in the villages of Imzelen, Ait Iguize, Afouzar and Kasbat-N-Tallaght. Some of the children were also recruited in the city of Biougra from the following centers: Wfuq, Ecole Privee "Al-Kalima", Ecole Privee "Assedk", "Ligue Marocaine pour la Protection des Enfants" and Mme Malika Talʕbubit's private daycare. Many thanks to all of the schools involved and the Aderdor and Toual families, who are among the people that made the relations with these institutions possible.

collecting written consent. Requiring a signature on a form could potentially induce anxiety about the intent behind the data collection which could lead participants to withdraw. However, children were always asked verbally whether they wanted to take part before the experimental stimuli were played and we only went ahead when they expressed assent. They would be asked if they wanted to watch a cartoon and respond to some questions by pointing at the screen of the laptop while watching the cartoon.

3.1.4 Materials

The videos used as stimuli were adapted from the videos used in a previous study that probed the comprehension of relative clauses and *wh*-questions (Perkins & Lidz, 2020). The characters involved in the events on the screen were a frog, a chimpanzee and an orangutan. All three happen to be animals to control for the animacy variable brought up in the section where we discuss distribution effects as a source for the subject-object asymmetry. The stimuli show three characters who are in the process of performing different actions on each other. The total of six trials corresponded to the verbs hug, bump, kiss, cover, causative laugh and lift. We filmed two new videos for cover and lift by featuring the same three characters and replicating the identical lighting and background settings, taking into account modifications due to the necessary props that were required for the new scenes. The auditory stimuli accompanying the videos were recorded by an adult female native speaker of TB using child-directed pitch and prosody. The stimuli for each separate condition were recorded in the same session, ensuring that the intonational contour would stay consistent throughout the same typology of instructions.

3.1.5 Procedure

TB-acquiring toddlers and children were scattered across the kindergartens of the Chtouka–Aıt-Baha province in little groups ranging from 3 children to about a dozen at a time. Sometimes it was also necessary to travel directly to the villages where the children resided with their families to supplement the number of participants. It was not always possible to maintain the same levels of background noise or quiet and there wasn't availability of the same type of furniture on which to prop the equipment and where to seat the participants. Despite the high variability of environments in which the experiment took place, there was an attempt at maintaining the steps as identical as possible across experimental sessions and settings and adapt them to the children involved and circumstances each time. The laptop with the stimuli, a MacBook Air 10,1 with a 13,3" screen, would be placed on a table, in such a way that the videos would be at eye level, directly in front of the children. Depending on whether the table was high or low, children would be seated on a chair in kindergartens or cross-legged on a carpet in the home. The screen would be exactly at arm's length to make it easier for the children to point at the target character of their choosing as they heard the audio instructions. Each session was video recorded with a Nikon DSLR 3500 camera on a tripod positioned at an angle right behind the child. The angle was chosen in such a way that the child's pointing with their hand on the screen was visible, but only one side of their faces was visible. The position of the investigator was next to the child and slightly behind them to minimize distraction. In some instances, when the background was particularly noisy, the entirety of the session was audio recorded using a Zoom H2n Handy recorder to supplement the built-in microphone of the DSLR

camera.

3.1.6 Time course of the experiment

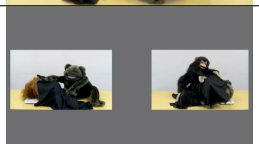
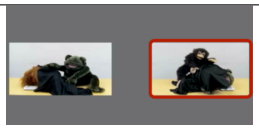
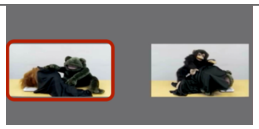
Phase	Visual Stimuli	Auditory Stimuli
Familiarization	 	Smoqql-za! Kra ar-itthbu kra. Tzritt? Kra ar-ithbu kra. (6.8s) x2 *Look! Somebody's hugging somebody. See? Somebody's hugging somebody*
Salience		Smoqql! Hati urtn saswa! (6.8s) *Look! They are not the same!*
Test		
Subject *wh*-question Condition		Manwa gitn ħizʕdad a i-ttħbu-**n** **a**gru? *Which monkey is covering the frog?*
Object *wh*-question Condition		Manwa gitn ħizʕdad a i-ttħbu **w**gru? *Which monkey is the frog covering?*

Table 3.1: Trial timecourse

The total duration of the experiment was 5.2 minutes, but the session would range between this time and 10 minutes total due to the icebreaker activities that children would be engaged in before the experimental video started playing. Before beginning the viewing of the video, children were given printed black and white drawings and sheets with stickers to stick on the pages. Once this interaction was completed, the children were given a brief

introduction to what they were about to watch and instructed about how to respond to the voice that they would hear.

At the very beginning of the video, the puppets of the characters that would be taking part in the different activities are introduced as a frog, a brown monkey and a black monkey. Each of the characters would show up separately and take up the screen for a total of 6.8 seconds each.

They would hear the corresponding lines for the other two characters during their introduction. After that, there would be a separate familiarization video for each of the six trial verbs, shown in table (Table 3.1). Children would see one of the puppets and hear audio that is intended to illustrate how the puppet will be called for the rest of the trial. The familiarization phase would be characterized by two videos in which the characters are engaged in the activity corresponding to the verb. In the case of cover, the child would first see a video of the frog covering one of the monkeys with a blanket. After a blank screen, the child would see a different video in which one of the two monkeys would be covering the frog. At the end of this step, the child heard a participant-neutral statement meant to draw their attention to the covering event without priming them about who out of the puppets initiated the action. Subsequently, the two videos were shown in a split screen set up so that the children would follow the two events simultaneously. This change was accompanied by a statement also intended to be neutral and non-priming. Then, 6.8 seconds later, as the children are watching the two videos playing simultaneously on the screen, they would hear the test sentence, exemplified in the third column in table (Table 3.1).

The matrix verb *show* was chosen since it was more suited to the action of pointing in the task in the present study. After this sequence, there would be a 4 second blank screen

during which the children would hear the test sentence without any visual accompaniment. After this 4-second period, the video of the two looped scenes would resume. This was the point during which the children were expected to point at the participant as was requested in the auditory instructions. Between this trial and the following trial there would be a brief 4 second break with a video of colorful toys rotating on the screen or falling and bouncing with catchy music playing in the background or a still image of an infant's face accompanied with an audio of a baby's laugh. This would keep the child's attention focused on the screen in between trial changes. The sequence described above for cover would thus resume with the remaining 5 verbs and corresponding scenes. The order of the scenes was shuffled to counterbalance them and the roles of the animals (agent vs. patient) were also reversed for the same purpose.

3.1.7 Hypotheses and predictions

For a child to perform in an adult-like manner, it would be necessary for them to have both mature grammatical representations and display mature deployment mechanisms. Under the hypothesis where children have adult-like competence and adult-like performance, one would expect them to point to the agent of the action in the depicted event significantly more in the subject condition than in the object condition, since they can look for and identify the anti-agreement morphology as well as the construct state morphology and integrate it in their parse as the information comes in.

Timeline 1 below illustrates (abstractly) how we would expect the online parsing of subject *wh*-questions to proceed, assuming a parser that is immature, i.e. has a less reliable

syntactic revision mechanism. The incoming pieces of linguistic information are presented on the left of the line, and on the right there are the corresponding hypotheses that get formed about NP attachement as new information is presented.

Let us replicate below the subject *wh*-question condition stimuli example.

(46) Manwa gitn hizʕdad a i-ttbujus-n agru?

Which in.them of.FS.monkeys COMP 3SGM-PROG.kiss-AAE FS.frog?

'Which monkey is kissing the frog?'

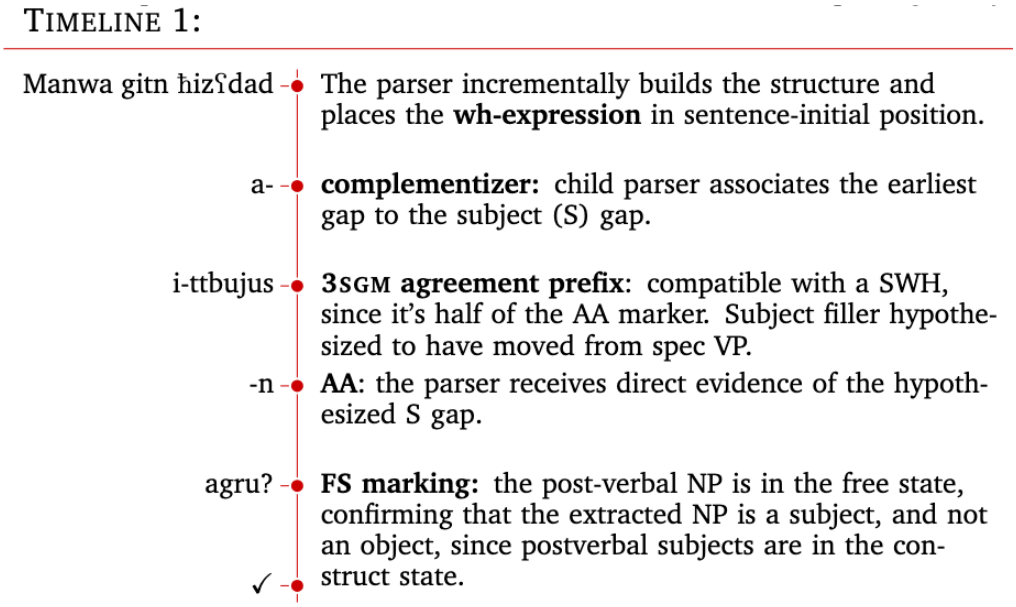


Figure 3.1: Predicted online parsing of a subject *wh*-question

Timeline 2 illustrates the parsing of object *wh*-questions. I will replicate below the *wh*-question children would hear during one of the object condition trials.

(47) Manwa gitn hizʕdad a i-ttbujus wgru?

Which in.them of.FS.monkeys COMP 3SGM-PROG.kiss CS.frog?

'Which monkey is the frog kissing?'

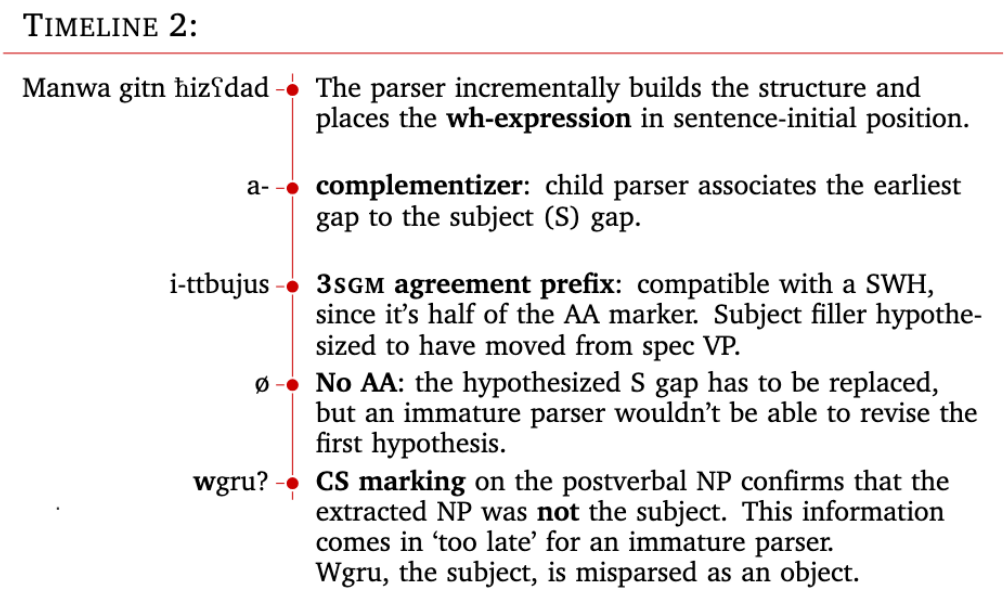


Figure 3.2: Predicted online parsing of an object *wh*-question

This hypothesis 'timeline' goes through the auditory stimulus directed to the children during the Object *wh*-question condition, here glossed and translated. The stage at which the verbal agreement information and the lack of anti-agreement are reached looks the same, hence the single tree.

In the case of Tashlhiyt Berber, a child listening to the stimuli during the object condition (47) would come across two sources of disambiguating information about the gap site. One would be the absence of the anti-agreement suffix on the verb, which confirms that the extracted NP is an object. The other discriminating piece of information would

be the construct state morphology on the postverbal NP in embedded position. Since this marker only shows up on postverbal subjects, it could provide information, together with subject-verb agreement, that the extracted NP is an object.

It's only possible to obtain more direct evidence of the children's specific knowledge of anti-agreement and CS if they can rely on a parser that is sufficiently mature to help them revise an initial misparse. The misparse revision skill is especially crucial to TB since both anti-agreement and the construct state show up in the more embedded position in the relative clause or *wh*-question.

3.1.8 Results

Coding was done once live and then twice offline by rewatching the videos of the experimental session by a coder not blind to the conditions. A further offline coding with a coder who is blind to the conditions will be carried out. We first obtained the percentage of points to the agent of the event depicted in the video by counting the number of trials in which the children pointed at the agentive animal puppet over a total of valid trials. A trial was valid when the child responded, whether target-like or not. Validity here is meant in terms of engagement with the task at hand. The results were then plotted using age as a continuous variable and dividing the responses into four facets, each corresponding to one of the conditions. Since the points are in terms of how many pointings were directed towards the agent, adult-like performance would be to have a high incidence of points in the subject conditions (SQ and SR) and a low incidence in the object conditions (OQ and OR).

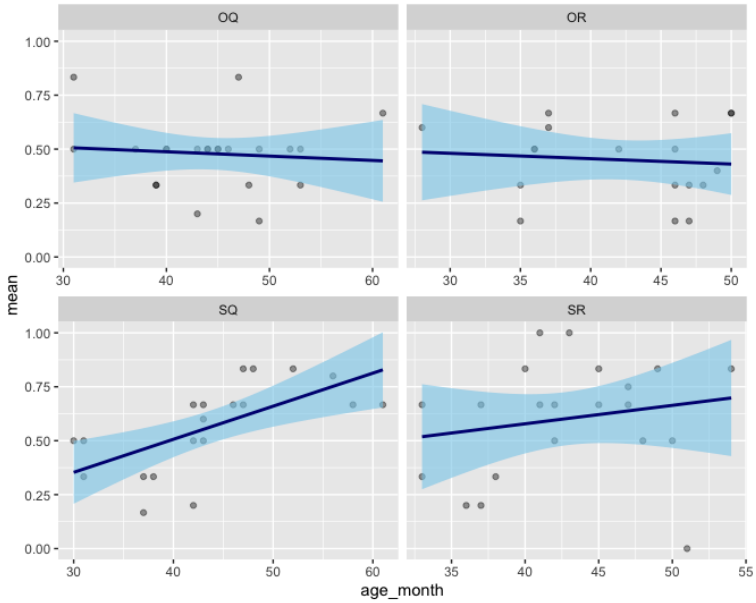


Figure 3.3: Percentage of pointings to the agent in each condition, by age. Each condition is represented by a facet of the plot, where OQ and OR are respectively the object *wh*-question and the object relative clause conditions. SQ and SR stand for subject *wh*-question and subject relative clause. On the x axis there are the ages of the children in months, treated as a continuous variable. On the y axis there are the percentages of agent responses, originally over a total of 6 trials. The pale blue shading indicates the aggregate standard error. Each dot corresponds to an individual child.

If we were to run our statistical analysis using age as a continuous variable, then statistical power would not be a problem since the children who took part in each condition added up to around 20. However, if we look at the specific numbers of children from each group, power might be an issue. Even the better-represented age group, 36-48, has less than 14 children taking part in each condition which might make it hard to reach any clear conclusions.

3.1.9 Analysis

To see what the possible predictors of the children's performance were, we fit a Generalized Linear Model in R Studio using the glmer model engine, which is part of the lme4

package[3] (Bates et al., 2015). We wanted to see how the condition (subject or object), the construction type and the age of the children can potentially contribute to agent responses. Agent responses here are intended as the percentage of pointings towards the agent in any given condition, regardless of whether the agent is the target response. The formula of this specific model is agent $\sim$ is_subject * is_question * log(age_month) + (1 | subject). This provides instructions for a model in which the occurrence of the response variable, in this case, the agentive response, is modeled to see what can predict it based on the possible interaction between the variable (is a question or a relative clause), the variable condition (is subject or object) and the log age of the participants in months. This formula does not just calculate the interaction between these three terms, but also all the possible combinations between them. Here, age is treated as a continuous variable.

We found two significant interactions, reported in figure Table 3.2. The condition (subject or object) was highly predictive of the rate of points to the agent. The other significant interaction that was found is the one between the gap-type condition and the children's age in months. The older children were, the more likely they were to point at the agent in the subject condition.

	Estimate	**Std. Error**	**z value**	**Pr**$(>\|z\|)$
(Intercept)	-2.9416	2.2344	-1.317	0.1880
is_subject	-4.9045	2.2344	-2.195	0.0282*
is_question	-1.3553	2.2345	-0.607	0.5442
log(age_month)	0.8130	0.5954	1.366	0.1721
is_subject*is_question	-1.4587	2.2341	-0.653	0.5138
is_subject*log(age_month)	1.3702	0.5954	2.301	0.0214*
is_question*log(age_month)	0.3576	0.5954	0.601	0.5481
is_subject*is_question*log(age:month)	0.3664	0.5953	0.615	0.5383

Table 3.2: GLM Model fitting results

[3]Many thanks to Utku Turk for fitting the model and for the plot in figure Figure 3.3

Whether the condition was a question or a relative clause was not a significant predictor, across the board, i.e. for both the subject and the object conditions (p= 0.5). Taking a closer look at the object condition, the average performance of the children is situated between 40% and 53%. You can compare this range to chance in the plot in figure Figure 3.4, where the chance level is marked with a bar filled in red. Only in the subject condition do we see performance above chance.

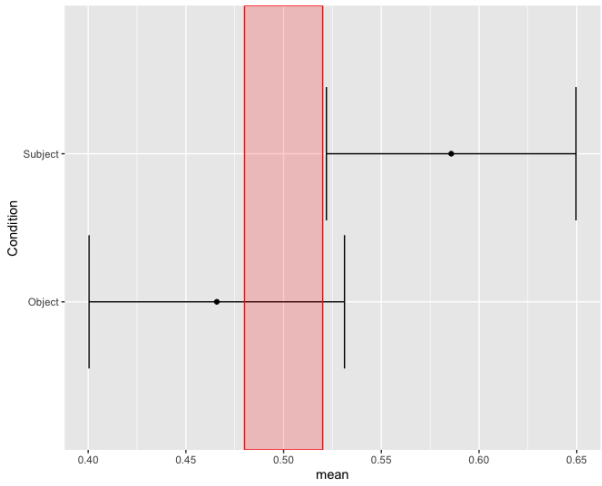


Figure 3.4: Mean performances in subject and object condition compared to chance level

We also ran another analysis to test whether there were any trial effects [4], although we did not expect any in particular. This was also done by fitting a GLM model with the following formula: glm(answer ∼ as.factor(verb),data= dataframe), which models how well the specific verb of each trial can predict the response variable, i.e. the percentage of points to the agent, here labelled *answer*. In the bar plots below, we can see how the only significant relation was between the verb laugh (causative) and the responses in the subject *wh*-question condition. The output of the generalized linear model also confirms that the only condition in which there the relation between one of the trial verbs and the

[4]Many thanks to Andy Yu for the code and the analysis.

children's score is significant is the SQ condition. The results of this one in particular are reported in figure 8. The analysis was carried out by using the verb laugh (causative) as the benchmark. The results in table 2 (figure 9) are all in relation to this trial. Comments on this will follow in the discussion.

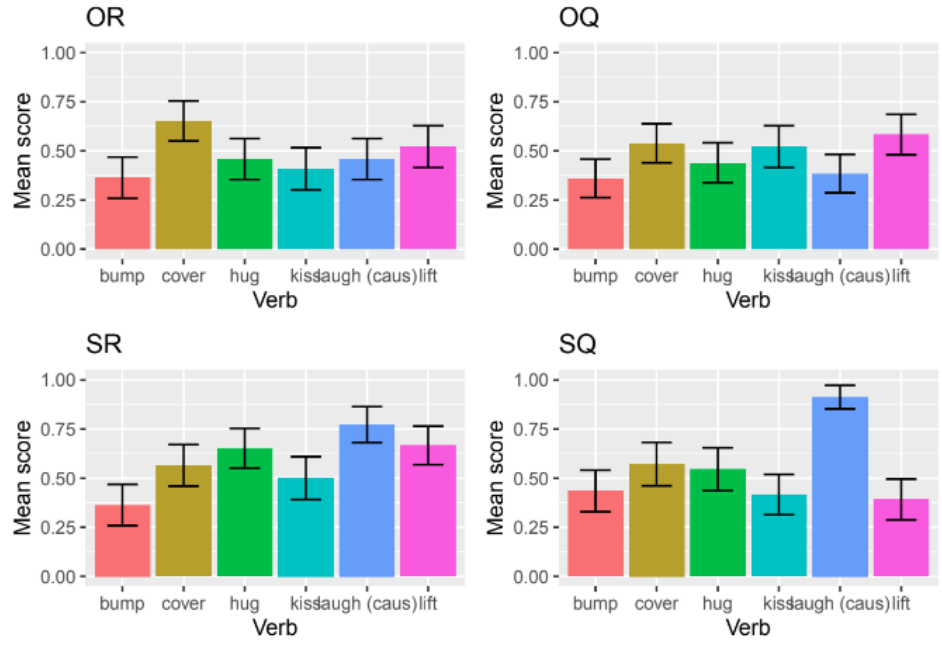


Figure 3.5: Mean scores by trial verb for each condition

	Estimate	**Std. Error**	**z value**	**Pr** $(>\|z\|)$
(Intercept)	2.3514	0.7398	3.1784	0.0015
verbbump	-2.6137	0.8510	-3.0713	0.00213
verbcover	-2.0637	0.8612	-2.3962	0.0166
verbhug	-2.1690	0.8548	-2.5376	0.0112
verbkiss	-2.6878	0.8478	-3.1705	0.0015
verblift	-2.7932	0.8543	-3.2696	0.0010

Table 3.3: Subject Question Condition results by trial

3.1.10 Discussion

A key takeaway from this study is that there was no difference in performance between relative clauses and *wh*-questions or, more precisely, the construction type was not predictive of performance. This is compatible with the notion that by the ages measured here, children treat these constructions alike because they represent them alike. This could potentially answer the question raised in Chapters 1 and 2 about whether children can recognize the abstract unity of Ā-dependencies in TB.

Another piece that can help children converge on the fact that Ā-dependencies are unified is their knowledge of anti-agreement. This is what we find with respect to this morphological marking. Children perform well in the subject condition, increasingly better as a function of age, and they perform at chance overall in the object condition, regardless of age. The subject condition pattern is compatible, in principle, with a subject bias as well as with acquired knowledge of anti-agreement. It is however necessary to make sense of the results in the two conditions in relation to each other. In a scenario in which children know the grammar of anti-agreement, they would be expected to know that its presence signals extraction of the subject and that its absence signals an object extraction. The question now is about the nature of the factor that prevents children from applying this hypothesized knowledge in the object condition.

The pattern of responses is in line with a subject-first bias that is architecturally entailed in an incremental parser, with no need to be independently formulated (e.g. an agent prominence bias, a frequency-based expectation, etc.). An immature parser that follows an active-filling strategy would lead to good performances in the subject conditions

and non-at-ceiling performances in the object conditions due to the failure to revise an initial subject misparse. This kind of parser would be expected to effectively mask children's knowledge of the grammar of object extraction.

If TB-acquiring children's responses were guided by a total subject-bias, this should have been reflected in the object condition in such a way that children would systematically pick the agent regardless of real gap site. However, upon closer inspection of the children's performance in the object condition, we do not see the same inclination to interpret the first NP as the subject. If the subject bias had been the only guiding principle behind children's responses (and no knowledge of the morphosyntax), then we would need to find an explanation for what mitigates the effect of this bias in the object condition.

This discussion would not be complete without mentioning the trial effects analysis. The effect of laugh (causative) could only be seen for one specific condition, i.e. SQ. The predictive power of the verb laugh (causative) of the performance of the children in SWH The predictive power of the causative verb laugh on the children's successful performance in the SQ condition raises the question of whether the relation is a causal one. Namely, one might ask: did the fact that children almost always pointed to the correct character in the screen when the trial was 'causative laugh' be due to the fact that the causative gives an advantage in this specific task? The causative construction is a productive pattern formed with the addition of the s- prefix to the root of the verb in Tashlhiyt Berber (Ennaji & Sadiqi, 2004, p. 89). There is no inherent subject-centricity about the causative, which would lead to the expectation that it be an effective cue in the subject condition. This effect is only present in one of the two subject conditions, which is puzzling in light of the fact that overall, there is no construction type effect ($p = 0.5$). Furthermore, no effect of the

only causative verb is found in the object condition. This pattern might not be meaningful and further tests are needed to see how a fully causative set of trial verbs would compare with non-causative trial verbs in a possible study.

3.1.11 Future directions

Aside from a specifically targeted production study to see if children at this age do represent the construct state and anti-agreement, the results would need to be compared further with the corpus of child-directed Berber that forms the basis of the findings of the final portion of this dissertation. Now that we have seen how children treat *wh*-questions and relative clauses (equivalently) and how they perform with anti-agreement marking (better as a function of age), it is time to ask whether this behavioral data is due to frequency in the input of the constructions they excelled in.

We will be seeing that in the upcoming chapter which is about the corpus of child-directed TB purposely built for this dissertation, object *wh*-questions are much more frequent than subject *wh*-questions. A pattern of the sort, on a frequency-based hypothesis where what's most frequent in the input predicts performance, children would have been much better in the object-condition than in the subject condition. In light of this finding that we will expand upon subsequent to the presentation of the corpus, based on this behavioral study, we suggest that the measurably worse performance of the children in the object extraction condition was not due to object constructions being more infrequent in the language.

Chapter 4: A corpus of child-directed Tashlhiyt-Berber

Now that we can confirm that children are competent comprehenders of the AA morphology by age 4, we can move onto the next steps of our investigation.

The corpus in this chapter is meant to achieve two broader goals, to start: a) document characteristics of the input available to TB-acquiring children, to subsequently; b) identify how children use the input to draw inferences about the structure of TB.

Here are the specific questions we want to ask:

1. Is the children's performance in the previous chapter due to frequency in the input of the constructions they succeeded at?

2. If overt complementizers are as rare as we expect in the input (given the universal preponderance of 'who' and 'what' crosslinguistically, and 'who' and 'what' ending in *-a* in TB), can children use the data from the complementizer usage in focus clefts to learn its obligatoriness in *wh*-questions?

3. What is the role that anti-agreement plays in signalling the similarity of Ā-constructions?

4. What is the presence of the CS in the input and could that be useful information for children's online processing of object *wh*-questions?

The question in (1) is meant to ask whether children performed best in the construc-

tions that are most frequent in the input and performed worst in the constructions that are least frequent. Thus, we want to see whether subject extraction instances are most frequent in a corpus and if specifically subject *wh*-questions and subject relative clauses are the most frequent. In terms of (2), we want to see whether children's knowledge state by the time they acquire AAE and the CS includes knowledge of the obligatory complementizer that is masked by independent phonological rules. One might wonder where the idea that children would be making a connection between focus cleft complementizers and complementizers in *wh*-questions. Here are two observations that support this idea. We know that children show evidence of hypothesizing a layer above TP from at least age 3 and will produce overt *wh*-words and complementizers in constructions where they are not required in adult varieties of English and Korean. This is the case of medial *wh*-questions (Liter et al., 2022; D. McDaniel et al., 1995; D. S. McDaniel, 1986; Thornton, 1990) inter alia and REL pronoun 'kes' overuse in Korean (Lee, 1991). The other idea is that children might have a preexisting notion of what Ā-constructions are, or more specifically, that they are expected to pattern alike. The way that this alikeness is expressed would vary from language to language but if this expectation is part of the toolbox that the child approaches the language with, then it would be a natural consequence to pool together the evidence that the child hears of a complementizer in *wh*-questions with the same one in focus clefts, especially when another outward expression of the similarity in the form of anti-agreement is available in the input. Questions (3) and (4) are motivated by the facts presented in Chapter 2.

The ideal scenario would be for the child to be exposed to a sufficient proportion of *wh*-questions with an audible [a] complementizer, alongside instances of anti-agreement

taking place across all Ā-construction types.

Here is what we would need to look for in the input to determine the possible answer to the developmental question. In the case of *wh*-questions, it would be important to code and analyze all the possible Ā-constructions that appear in the input.

1. Varied co-occurrences, i.e. the child would see enough instances of the transparent WH+COMP strings to show that the *a-* occurs with *wh*-words in general, and not just some.
2. Occurrences of the *a-* complementizer in clefts, where the *a-* must follow a clefted fronted element that ends in a vowel other than a so that the child infers that the inaudibility of the C head is due to a phonological rule and not a morphological one.
3. Occurrences of anti-agreement across all Ā-construction types, so that the child correctly lands on the generalization that AA is an Ā-related phenomenon and not specific of only some constructions, e.g. RCs or WH questions only but not clefts.
4. Instances of Construct State compared to instances of Free State across all nominal constructions, paired with a study of what instances of CS are overt and transparent, i.e. more deployable for the purposes of subject tracking and filler tracking more in general in Ā-constructions.

The key aspect here is *diversification* of the input, rather than frequency of occurrence, although both properties are functional to successful development of target-like representations. By diversification I mean that there should be enough distinct and discrete datapoints that the child doesn't end up making the wrong generalization about the distribution of a property like anti-agreement. (e.g. only RCs and cleft have AA because the child has only ever heard object *wh*-questions in the input).

We know from an extensive body of work that frequency correlates positively with performance of children for a vast array of construction types and "modules" of the grammar (Stromswold 1995, Huttenlocher et al 2002, Lieven 2009 inter alia). It is however still unclear whether frequency impacts the learning mechanisms themselves or whether it

impacts the practice of the child post-acquisition of a construction. Nevertheless, we would expect that the more exposed children are to phenomena such as anti-agreement constructions and a variety of *wh*-questions, the better they will produce them and internalize them in their mental grammar.

The 'ideal quantity' of the various elements is however challenging to determine in absolute terms. There would need to be sufficient cues to show that the *wh*-word is shaped like a cleft, with its corresponding complementizer. There would have to be a noticeable amount of cleft constructions that can make it to the intake[1] as well as relative clauses and headless relative clauses in the input alongside *wh*-questions so that it would be easier for the child to draw parallels between the underlyingly identical structure of the three constructions. It is however unlikely for clefts and RCs to be as numerous as *wh*-questions. RCs are definitely known to be significantly less common in spoken language than in written text, for instance (Kush, Sant and Strætkvern 2021). Also, note that the issue of the complementizer not being visible in *wh*-questions where the *wh*-word ends in a vowel also shows up for cleft constructions where the clefted NP ends in a vowel, or specifically in *-a.* Therefore, the input would have to have a qualitatively helpful proportion of clefting examples where the *a* is equally visible, otherwise it would make no sense to draw the parallel across constructions. This is an important note to keep for the analysis of the corpus data we will be illustrating in the Results section of this chapter.

[1] A common use of *intake* in the literature conceives of intake as the data internalized through attention and perception, i.e. what the child has attended to, which is then available for further computations in the mind. This is known as the perceptual intake (Omaki & Lidz, 2015). The other possible and related definition is of the intake as data that the language acquisition mechanism selectively extracts from the input on the basis of specific biases that might be guiding the child's learning (Lidz & Gagliardi, 2015; Omaki & Lidz, 2015). In this dissertation, I will mostly be making reference to perceptual intake unless otherwise stated, since a corpus limits what we can say about what the child has actively or selectively taken to be part of how they are making inferences about the language.

4.1 Methodology

Behavioral studies can help us probe what a child's competence and performance might look like at a given point in time, but corpora of caregiver-child interaction, especially ones that were collected longitudinally, are, for a linguist, a valuable source of information about (a) the experience of the child and (b) the kinds of utterances that children produce. When comibined with explicit theories of grammar and inference, corpus data can help us identify possible learning paths, which can then be tested against further behavioral data. We obtain evidence about the timeline along which a child's grammar develops, both from a competence and performance standpoint depending on the questions we ask. There is also immense versatility to a corpus, depending on the original goals that led to its construction.

Unlike, however, much of the previous work on children's productions ((Brown, 1973; Demuth, 2022; Snyder, 2007; Stromswold, 1995) inter alia), the present corpus was planned and conceived to catch a glimpse of what is part of the child's ambient noise in a setting. As much as possible, we attempted to clear up the noise in the relevant sections and systematically transcribe the Tashlhiyt Berber utterances in the ambient speech, to extricate analyzable utterances from overlapping speech so as to be able to make more accurate predictions.

4.1.1 Corpus Profile

The present corpus of child-directed TB was collected over the course of three years, from the winter of 2020 to the spring of 2024. More specifically, the first video was filmed

December 21st 2020. The child, whom we shall call "Rima" was 2 years, 3 months and 26 days old at the time. By the time of the last video included in this analysis, or tenth session, she was 4 years, 8 months and 5 days old. Her sister "Nala" joins the recordings as an addressee and a babbler from the second session onwards. The age of the target children is written in the standard notation in the language acquisition literature thart follows the following format: Y;MM(,DD). Note that the delimiter between years and months and between months and ages are, respectively ';' and ',' . Therefore, 4,07 could only mean 4 months and 7 days and 4;7 means 4 years and 7 months old.

Child	**# utterances**	**Age span**
Rima	2677	2;03,26-4;08,05
Nala	562	04,06-1;10,28

Table 4.1: Child Utterances and Age Span

Session date	**Duration**	**# Utterances**	**Rima's age**	**Nala's age**
Dec 21st, 2020	00:41:23	950	2;03,26	NA
Oct 10th 2021	00:31:28	1157	3;01,15	4,07
Mar 17th 2022	00:23:43	739	3;07,09	9,15
Apr 3rd 2022	00:31:42	797	3;07,09	10,00
Apr 10th 2022	00:42:49	1523	3;07,16	10,08
Apr 13th 2022	00:51:38	1144	3;07,19	10,12
Jan 15th 2023	00:43:02	970	4;04,21	1;07,13
Jan 18th 2023	01:02:11	1393	-	1;07,16
Jan 23rd 2023	00:37:16	1213	4;04,29	1;07,21
Apr 30th 2023	00:39:19	1598	4;08,5	1;10,28

Table 4.2: Corpus overview

Other than the sessions summarized in table 6.2, only one video, dated February 22nd 2024, was excluded from the analysis because upon transcribing it, the audiofile was found to be corrupted and therefore unusable. The total transcribed, coded and analyzed sessions amount to 403 minutes, circa 6.8 hours of utterance-dense caregiver-child interactions. The

mean duration of the 10 recordings is 40 minutes and 30 seconds.

The density of the recordings is calculated by looking at the mean number of utterances per minute, reported in the table below. The utterance number in the is calculated including the children's utterances as well. For a specific breakdown, there are some further charts below.

Sessions	**# Utterances**	**utterance/min**
1	950	23
2	1157	37
3	739	31
4	797	25
5	1523	36
6	1144	22
7	970	23
8	1393	22
9	1213	33
10	1598	41

Table 4.3: Utterance number divided by session

The utterance number was calculated by using a command in CLAN, MLT, which stands for Mean Length of Turn, that pools together all utterances per speaker tier (MacWhinney 2024, 26). As the name suggests, it also outputs the number of words per utterance for each speaker's utterance, but that is a function that is not being used here. It is possible to either pool together all utterances without specifying what tier we are looking for, in which case we would obtain the number of all utterances. For this specific purpose, the formula below was ran within the CLAN interface.

(48) mlt +t* @

When specifying the tier, we can pick out the specific number per speaker, as we do in the example below:

(49) mlt +t*GRAN @

This command will also include all unintelligible utterances (transcribed as xxx using the CHAT transcription convention) of the speaker as well as utterances that were not transcribed in full (www) or are irrelevant to the input because pronounced in a room where none of the children are or pronounced by the investigator. As you can see from the results of the count, we can get a clearer picture of what was taking place in the recorded sessions and compare this with daily life in the household on the basis of an informal interview conducted after the corpus was built.

The sessions followed a similar structure with the exception of the January 15th one which was filmed at dinnertime and was characterized by a significant amount of overlapping speech and adult-directed speech.

To make the setup as naturalistic as possible, visits were arranged on the basis of the children's routines, to ensure both minimum disruption and optimal recording conditions. These sessions still reflected the weekly activities that the family would engage in. The parents and the children's grandmother did not want the children to be spending most of their time watching cartoons, so they encouraged the use of illustrated books for playtime. There are oftentimes, throughout the various recordings, requests on the part of the children to watch cartoons that were only seldom granted.

4.1.2 Procedure

The video and audio recording sessions were carried out in the children's home as well as in the home of relatives as was the case when they were visiting their aunt. Before every

recording session, children were asked if they consented to being recorded or beginning to play in a specific area of the house in terms of lighting and space. Their unwillingness to talk was interpreted as a revocation of consent and therefore followed immediately by an interruption of the recording, which explains the high variability of the durations listed in table 3.1. When adults came into the set up they were also asked if it was fine for us to record them and if, after an explanation of the goals, they consented to only the audio and not the video, they would not be filmed at all or the camera would be pointed towards an angle that would exclude the member of the family from the line of sight. This is why, oftentimes, the mother and some of the uncles are only audible and not visible in most of the videos.

The preferred method for recording the video-sessions was a Nikon DSLR 3500 camera on a tripod positioned at an angle right behind the child. The camera's recording timer allowed it to only record in increments of 20 minutes at a time, therefore the videos that lasted longer than 20 minutes always consisted of the editing of multiple clips together. The December 2020 and all the January 2023 sessions were filmed using this device operated by the principal investigator. These recordings were constrained by the presence of the investigator on Italian soil which could take place only during the winter break. Recordings were made impossible during the summer break because the children and their families would often travel abroad for a considerable chunk of it that overlapped with the time that the principal investigator would be able to travel to Italy. The session from October 10th 2021, the only one for which only an audio recording is available, was recorded using an LG phone belonging to the grandmother of the two girls, whereas the rest of the video sessions were video-recorded by means of an iPhone 11 belonging to the aunt of the two girls. The

videos and audios were sent to the main investigator through the Box cloud storage service using a VPN for added encryption.

There was minimal involvement in the conversations on the part of the investigator present. Since the speech of the investigator does not constitute regular input in the children's daily life, and since the investigator is only a heritage speaker of TB, speaking would not have constituted a contribution to the database. Here are the few exceptions to this rule of no intervention. The very first videorecording, December 21st 2020, features the highest number of utterances by the investigator, since it was mostly a trial, but was then included into the analysis due to the high quality and representativeness of the set up. Sometimes, mid-session it was necessary to communicate with the family about whether or not the children were allowed to touch objects or equipment brought in by the investigator.

Other than this, interactions were avoided to prevent any form of interference to the naturalistic goals of the set-up. As many measures were taken as possible, given the means available, to make the setting and data collection naturalistic[2].

The presence of an investigator (albeit one that is related to the participating family) sitting next to an unusual device such as a DSLR camera mounted on a professional tripod might not be the picture of naturalness, which is also why it was advantageous that the

[2]There are 8760 hours in a year. If we multiply this by the years of a child before schooling begins, an age by which most generative developmental linguists would argue that adult-like grammar has been achieved, we obtain about 52560 hours, a generous estimate that does not account for the fact that infants will sleep for about 18 hours a day for the first half year of their life. This also excludes in-utero exposure.

Attempting to catch a glimpse of the linguistic developmental process by relying on a 7-hour corpus of child-directed speech is like being tasked with hypothesizing the falling trajectory of confetti scattered across a stage, only on the basis of their color, material, and shapes, with the help of a single fixed spotlight.

Talking about representativeness of the input given this state of affairs is an idealization at best. These approximate numbers are provided as a way of illustrating the sampling issue. In terms of naturalness or ecological validity, that is difficult to achieve without a LENA or similar portable device that can follow a child and minimize the extraneousness brought about by the presence of an investigator in a room. However, the way in which we went about collecting this preliminary data could be a helpful starting point for our purposes.

majority of the sessions were filmed by relatives present in the home with their personal devices which often led the children to forget they were being recorded.

There are, on the market, much more less intrusive ways of recording speech. This procedure was made on the basis of the means available at the time during the pandemic, and subsequent recordings were also dictated by the circumstances. The hope in future is to obtain even more naturalistic input to achieve more ecological validity of this method of data collection.

4.1.3 Linguistic and migratory background of the relatives

Rima, and subsequently Nala who was born one year into the project, live together with their parents as well as their paternal grandparents in a condominium in the northern Italian countryside. The main caregivers, i.e. the adults who spend most of the time with the children are the paternal grandmother and the mother of the girls who however does not figure prominently in the video sessions. The aunt has the children over at her apartment at the weekends sometimes which is the setting for the April 2022 sessions. The fact that the grandmother heavily features in most of the recorded sessions is representative of the chunk of time that she spends conversing with her granddaughters in daily life since she is the adult who plays with them when the mother is taking care of chores and cooking around the house. The reason why the mother does not feature particularly from the second session onwards is because she has Nala, who at that point is only 4 months old to take care of.

The presence of the paternal aunt, the girls' father's sister, is however not represen-

tative of the children's daily life. Despite featuring prominently across the recordings (she plays, reads and eats with the girls in 6 out of 10 recording sessions), she only visits the children once a month or more frequently depending on whether they need to be babysat, in which case they will visit her at her apartment a few km away in the same town. Nevertheless, her input is included in the analysis because she speaks the exact same variety of TB as the majority of the family, which means that her input could be considered an extension of the father and the grandparents'.

Rima and Nala's father moved to Italy aged 17, and their aunt, moved to Italy aged 20. The grandmother has been living in Italy since age 35. The adult family members visit the TB-dominant region annually. These aspects of the family's "upkeep" of TB are important to specify given that the setting is distant from the source country. The biographical details above are meant to stress that the adult family members stopped being immersed in TB daily and consistently since an age that is considered to be well past the critical period of language development. These are also necessary to keep in mind for an eventual study in which we compare the quality of the input with a family that speaks TB in a non-diasporic context.

4.1.4 Recruitment of the family

Note that the behavioral experiment was conducted in Morocco, in a Tashlhiyt-dominant region, whereas the corpus was collected in a predominantly TB-speaking household that resides in Italy. It is possible that the quality of the input recorded and transcribed for the corpus analyzed here was different from what would have been obtained had we collected

a corpus of child-directed TB in the source country. However, here are some observations that mitigate this difference and that motivates the validity of this corpus in a comparison with a behavioral measure of children's competence in the language.

Firstly, the recordings were conducted starting from an age at which schooling in the dominant language (Italian in this case) has not yet started. The goal was to conduct a longitudinal study where it is possible to get an idea of how input changes over time and obtain a sufficient amount of representative samples between ages 2 and 5. The recordings could not have extended past the ages of 4 or 5 for the main child, because that is around the time that children start getting enrolled in kindergarten in Italy. Since TB is a minority language in Morocco as well, it is a matter of time before children start getting systematically exposed to Arabic and/or French. Children would still mostly speak TB outside if they are growing up in the Souss-Massa region. This is not to minimize the possible differences that would arise from TB being in contact with Italian vs. with Moroccan Arabic, the latter of which being significantly closer typologically and phonologically.

4.1.5 Transcription

Once back from the videotaping trips to Italy, the principal investigator took care of carrying out a full transcription of the videos and recordings that were possible. Since the family members had to be relied on for some of the recordings on the basis of their availability and availability of equipment, sometimes, we would receive short clips from different sessions instead of a full recording. Those sessions would not be included in the final transcript since, by themselves, they were too small a sample compared to the ideal sample recording

of 30 to 60 minutes.

Reliability of the transcription has not been verified by a second transcriber due to the unavailability of Tashlhiyt Berber speakers who could do this work. The files will, however, be made available with their audio-linked counterparts for anyone who would be willing to contribute their time and attention to the project. To mitigate the errors of the sole transcriber available, the transcriptions were revisited and checked a few months following the first transcription and checked a third time a few months after the second check.

The transcription was carried out using IPA, resulting in a broad transcription of the parental speech and narrow transcription for the children's early articulations. Although there already exist a few conventions for reporting Amazigh language examples in the linguistic literature (Guerssel & Hale, 1987; Ouali, 2011) and growing internet databases in TB (Outahajala et al., 2010), IPA was used with the purpose of making the corpus of easier access for researchers who are not Berberologists as there is growing interest in the languages outside this realm. Although arguably more laborious, using IPA could potentially pay back in terms of ease of access to the resources available on the language.

Another advantage of using IPA as opposed to orthography is "theoretical neutrality" (Snyder, 2007). Namely, by avoiding to transcribe a word using standard orthography, there is no imposition on the part of the transcriber of their own interpretation of a word that is pronounced at that specific developmental stage. However, whenever possible, the %com or %eng (comment and English translation respectively) tiers were added in the .cha transcription file, with the best guess of what word the child was attempting to pronounce, taking into account the discourse as well. Since the main purpose of the present corpus

was a syntactic analysis of Ā-constructions present in the input of the child, special care was mainly taken for the transcription of adult speech and child speech in the context of these constructions, to evaluate competence. Overall, however, there was not as much of an interest in transcribing the children's productions, which is why any contributions from the linguistic community to improve the children's output reports are welcome.

The transcription was organized using the CHAT system conventions, the most up-to-date manual (MacWhinney 2024). The transcription files are all in .cha format and hosted in the CLAN software which is the standard for CHILDES corpora. As the transcription was underway, speech was segmented into tiers that corresponded to each speaker participating verbally in the conversation. If someone never said anything but showed up in the "scene" their actions would be described in a dedicated %com: or %sit: tier only if the action could not be determined from the video alone.

The speech stream was segmented into separate utterances on the basis of the prosody, using in turn the dedicated notation for interruptions, whether self-created or externally induced, as well as the notation for overlapping speech. The entirety of the corpus is also media-linked, meaning that clicking on any given utterance allows playback that retrieves both the audiofile and, when possible, the video file as well

Originally, the hope was to videotape sessions of caregiver-child interactions at least once a month, for a duration of 1 to 2 years. However, it was not always possible since some of these recordings were dependent on the availability and media skills of the children's relatives in those moments of the year in which the principal investigator could not be present in person.

The ultimate goal with the corpus presented here is to donate it to CHILDES and

make it available both to the linguistic community and whoever would need to make use of this corpus for their work. Whilst the original aim was to arrange a substantive source of data to make precise hypotheses about the input and intake of a child acquiring Tashlhiyt Berber syntax specifically, this corpus can be useful for far beyond what it was built for. It could be helpful to observe up close the babbling patterns, or more in general the articulatory development in a language with TB's idiosyncratic phonological properties. It can also be helpful in the context of heritage language research to look at the early stages before a child's competence in their native language begins attriting to make space for the future dominant language, seeing as the children involved here reside in Italy and will be enrolled in a kindergarten by the time the sessions come to a close. Given the detailed %sit or situation descriptive tiers, that provide the context for the image when the video is not accessible, the corpus could also be used potentially for sociolinguistic work.

Before the construction of this corpus, the one publicly available corpus of Tashlhiyt Berber is the one collected by Lahrouchi and hosted in PhonBank (Lahrouchi & Kern, 2018), i.e. the extension of the TalkBank database (which CHILDES is part of) which focuses on children's phonological development. This corpus is specifically designed to provide a timeline of babbling of infants. Therefore, transcriptions are only provided for the children's production and other than the audio files, there is no consultable transcript for the parental interactions. This is the only available corpus of Tashlhiyt Berber, the same variety that is spoken in the corpus I report on in the present dissertation.

One of the first corpora of its kind in Kabyle Berber or Taqbaylit, was constructed by Tabti (2013) with the goal of following a child's stages of language development before "full" production competence, which the author calls *prélangage*. Tabti's Kabyle corpus includes

child and adult-Berber where words are transcribed for the adult speech also. This was mainly conducted to derive a sociological analysis on the basis of words and word usage. The words collected were then classified into different thematic categories, i.e. family members, food, animals, objects, body etc... And then listed only in their unique instances, without repetition. While this corpus lends itself particularly well to lexical analysis, it does not allow a more fine-grained morphosyntactic analysis in its current available-to-use form. There is also extensive work on baby talk by Gutova (2015) on babytalk in Senhaja de Srair Berber.

However, until this instance there was never a corpus of child-directed TB that had the primary goal of investigating the input from a morphosyntactic standpoint.

4.1.6 Coding

I first established three separate macro columns: A'-constructions, declarative clauses and nominal categories. Utterances are tiered in CLAN, i.e. to each utterance corresponds a code assigned to each speaker. Upon extraction of a speaker's tier, the utterance is segmented into all the verbs and for each verb, their agreement features are coded.

For instance, if the goal was to code declaratives, let us consider the following example, taken from the transcript for the December 21st, 2020 session:

We also segmented sentences like this into two separate units, since the target here is the subject-verb agreement morphology in those environments in which it is licensed tnnaim tra tifiji tra tifiji.

CHILDES now uses universal dependencies for tagging the transcribed speech and

since there were no sufficient resources to add TB to the Universal Dependencies database, I came up with a system of tagging things manually through Excel spreadsheets.

The glossing of the transcription was not theory-neutral, in that I did not follow the convention for glossing anti-agreement in the WALS, but simply used AAE as an abbreviation. More specifically, for the partially glossed portion of the transcript, I was able to carry out an analysis by means of the freq and combo lines and I subsequently double-checked the results through a manual search that was aided with Excel.

4.1.7 Analysis

In the analysis, we will include utterances that are directed at other participants in the conversation and not just what is directed to children. This is informed by the extensive literature on overheard speech and children's receptiveness to what has been said by adults to each other as well as the observation of this behavior among the children involved in the corpus themselves.

We coded the proportion of A'-construction relative to all the utterances, including exclamations and interrupted clauses. Then we compare the agreement patterns with declarative clauses.

The following elements were **included** in the count of declarative clauses:

1. Negation. Although negative concord in TB is marked through a vowel alternation change in the negated verb,these instances were counted because what was relevant was the presence of the masculine singular prefix as well as the feminine singular prefix which were the target that mattered

2. Polar questions, as they keep the agreement intact as well as the word order of a declarative clause

Here is an example[3] of the vowel alternation where a standard agreeing perfective verb ends in *-a*, whereas it ends in *-i* when there is negative concord (Ennaji & Sadiqi, 2004).

(50) *GRAN: Illa diddi@c
%xcod: 3SGM.be ouch.
'%eng: there's ouch'

(51) *GRAN: oho ur dars illi diddi@c oho.
%xcod: no NEG on.her 3SGMbe.NEG ouch no.
'%eng: no there is no ouch on her, no'

In examples taken from a transcription in CLAN, %xcod indicates the glossing tier where all morphemes and segments are defined and the %eng tier indicates translations. Tiers formatted asterisk followed by acronym such as GRAN, CHI etc are speaker tiers assigned to a participant in the conversation being transcribed.

The following elements were excluded from the declarative clause count:

1. Tenseless/aorist clauses
2. Imperative
3. Disfluencies
4. Self interruptions before the completion of a predicate were also excluded.

[3]Note that the %xcod glossing convention reported here follows Leipzig glossing rules instead of the CLAN or CHILDES internal system now based on Universal Dependencies. In the present dissertation, I will be using this convention to improve legibility of the various examples adapted from the corpus.

5. Repetitions

Instances of immediate duplication of a question within the same "utterance"/prosodic unit were counted as one. For instance, if one of the adult caretakers asked:

(52) Matʃʃit matʃʃit?
WH-COMP-2SG.eaten WH-COMP-2SG.eaten?
'What did you eat, what did you eat?'

An example like (52) was coded as one question. On the other hand, if the structure was identical but the verbs were different, the questions were coded as separate as is the case for examples like the one below.

(53) Ntta ad ʃʕʕiħ ntta ad ʃʃiħ.
DEM COMP crave.1SG DEM COMP ate.1SG
'That is the one I craved, that is the one I ate'

4.1.8 Results

The upcoming numbers relating to Ā-constructions in the input have to be put in the context of the total sum of utterances pronounced by all the familiar adults in the various video recording sessions. The total of these utterances is 7982, including all repetitions. Out of all of these, I extracted the following subject fronting Ā-dependencies. Unlike the case for the utterance calculations, these were obtained through an initial manual coding in Excel. Also note that the percentage is calculated not on the basis of the total utterances, but rather using all instances of A'-constructions as the total, i.e. 1164.

	Subject extraction				
Clause type	**SWh**	**SCI**	**SRC**	**Redux SRC**	**Total**
Counts	101	60	43	79	283
Proportion	8.8%	5.2%	3.7%	6.8%	24.3%

Table 4.4: Subject extraction counts and proportions by clause type.

	Object extraction			
Clause type	**OWh**	**OCl**	**ORC**	**Total**
Counts	238	49	21	308
Proportion	20.4%	4.2%	2%	26.5%

Table 4.5: Clause type distribution in object extraction

Adjunct extraction				
Clause type	**AWH**	**ACI**	**ARC**	**Total**
Counts	325	62	14	401
Proportion	28%	5.3%	1.2%	34.4%

Table 4.6: Adjunct extraction data by clause type.

Across all question types, here is a breakdown of the types of *wh*-words that children are exposed to in about 7 hours of input we have fully transcribed and analyzed. First, a separation was made between the wh-phrases that make the *a-* complementizer transparent and the wh-phrases that make it opaque, or that create the context for elision.

The total is the percentage of opaque vs. transparent strings of *wh*-word + complementizer is on the basis of 664, the total number of *wh*-questions across all extraction types (subject, oblique and object). Isolates and fixed expressions are included in this total count. However, since the question at issue is the co-occurrence of the *wh*-word with the CP it takes as a complement, the actual total number of utterances with the string wh+COMP with a full remainder is **642**.

	Types of *wh*-words	
Clause type	**Opaque WH+~~a~~**	**Transparent WH+a**
Counts	605	37
Proportion	91.1%	5.6%

Table 4.7: Types of *wh*-words in clause types

As can be seen here, the transparent strings where the *wh*-word doesn't end in a vowel and therefore makes the *a*- complementizer visible are only 5.6% out of the *wh*-questions. Whereas the opaque forms constitute 91.11% of the total *wh*-questions.

4.1.9 Breakdown of the *wh*-words

	Opaque ~~wh+COMP~~		
Clause type	**Ma**	**Mani**	**Manza**
	who/what	where	where
Counts	524	63	18
Proportion	79.1%	9.5 %	2.7%

Table 4.8: Opaque strings breakdown

One note is in order about the fact that ma in TB means both who and what. In the child language literature, we see examples of analyses where there is a distinction made between subject who and object who, as well as subject what and object what (Stromswold 1995). I will not make this further breakdown because ma is ambiguous between who and what, or rather, there is no who vs. what distinction in Berber at the level of the wh-phrase. If I were to make a subdivision, I would take for granted that all *ma*- strings followed by anti-agreement are the subject, but it would not be clear whether the speaker "meant" what or who.

Going back to the initial percentages of the opaque vs. transparent forms, the result of

Clause type	Transparent wh+ COMP				
	mnnaw	**mnʃkk**	**managw**	**manwa gitn**	**manik**
	how many	how much	when	which	how
Counts	1	2	8	10	16
Proportion	0.2%	0.3%	1.2%	1.5%	2.4%

Table 4.9: Clause type counts for Transparent wh+ COMP

this initial count shows that the vast majority of *wh*-words present in the input is of the kind that "drowns out" the signal of biclausality that could be provided by the complementizer. This will be expanded on in the discussion.

4.1.10 Focus cleft constructions

	Clefted element + COMP	
	Opaque (ends in -a)	Transparent
Counts	36	135
Proportion	21%	79%

Table 4.10: Counts and Proportions of Clefted Elements

In the corpus, a total of 171 focus clefts was counted. The distributions in table (Table 4.10) shows the reverse pattern of what we found to be the case for *wh*-words in table(Table 4.7). Almost 80% of the instances of focus cleft followed by complementizer are of the transparent case, where the clefted consistuent ends in a consonant or a vowel other than *a-*. The empirical reason behind this is that the most frequently clefted constituents are demonstratives, possessives and location adverbs whic all end in consonants, such as *wwann* 'that one' *xtann* 'that.F one' and *ʕid* 'here'.

I report below some relevant examples from the corpus.

(54) a. ɣainn a-mu giħ lʕin .
DEM COMP-to 1SG.gave eye

'That's what I saw'

b. wwad a-tbbi.
DEM COMP-3SGF.tore.

'That's the one she tore up'

Utterances (54a) and (54b) were produced by the children's grandmother (*GRAN:) during the January 23rd, 2023 and January 18th, 2023 sessions respectively. The rare times in which a clefted NP ended in -a were the instances where what was clefted was a first name, a place name or a borrowing from another language. The example below was adapted from an example found in the corpus.

(55) Bologna a-s tdda

Bologna COMP-to 3SGF.went.

'It's Bologna that she went to. '

4.1.11 Declarative clauses agreement

In this section, we will be looking at the percentages of each type of subject-verb agreement instance across all declarative clauses and compare that with the types of phi features associated with heads in subject extraction constructions.

Agr	**1SG**	2SG	3SGM	3SGF	**1PL**	**2PLM**	**2PLF**	**3PLM**	**3PLF**
Count	174	349	476	603	106	7	8	82	26
Prop.	9%	18%	25%	31.2%	5.5%	0.4%	0.4%	4.2%	1.3%

Table 4.11: Person, number and gender agreement occurrences in declarative clauses.

The vast majority of the subject-verb agreement instances take the form of the 3SGF marker, which makes up 31.2% of all agreement instances in declaratives, followed by 3SGM which makes up 1/4 of subject-verb agreement instances. Another number not present here is that the vast majority of the masculine instances are inanimate, whereas the reverse is true for the feminine agreement where the subject tends to be an animate subject. Out of 283 inanimate subjects, less than 10% were feminine instances.

The empirical motivation for this distribution is the mostly feminine participants and referents in the discourse as well as the presence of mostly women during the recording sessions. The masculine subjects mentioned tend to be either the children's father, uncle, grandfather, who however do not feature heavily in these sessions, and mosly animals, toys and characters from books and cartoons.

4.1.12 Construct State morphology in the input

Recall from the Chapter 2 presentation of the Construct State that the only transparently marked forms of the CS are masculine and feminine singular nouns starting in *a-* and *ta-* respectively. Therefore, in order to ascertain the possibility of learning about this morphology, we will only be considering instances of the transparent morphology in the input as well as the instances of the construct state that specifically mark post-verbal position of the subject, unlike the other form of construct state examples in the language of some specific prepositions that always take complements in the CS, which is an unrelated phenomenon for our purposes.

The CS and the FS forms were calculated on the basis of a series of nominal expres-

sions that include fully specified lexical nouns and pronominal clitics. The total number of nominal expressions counted in the input was 1639. Out of this total there were 143 instances of CS, both overt and opaque, i.e. the case of borrowings from other languages.

Nominal Marking	**FS**	**CS**	**Total**
Counts	1496	143	1639
Proportions	91.3%	8.7%	100%

Table 4.12: Instances of FS and CS across nominals

The CS constitutes less than 10% of nominal markings and this number includes the opaque instances of CS where the vowel/glide alternation in the case of the masculine and the vowel elision in the case of the feminine are not transparent. This pattern is reminiscent of the opaque vs. transparent *wh*+COMP string alternation which shows how the language is rife with opacity in specific contexts crucial to learning.

CS Cases	**MSG**	**FSG**	**Total**
Counts	122	21	143
Proportions	85%	15%	100%

Table 4.13: CS instances breakdown - Masculine and Feminine cases

The data in table (Table 4.13) is a further breakdown of the specific CS instances that have been found in the input and separates the masculine from the feminine singular instances, the latter of which are in the minority.

CS Cases	**PP Object**	**Opaque**	**Total**
Counts	108	60	143
Proportions	75%	42%	100%

Table 4.14: CS instances breakdown - CS objects of prepositions and opaque cases

Immediate alternation	English	FS # instances	CS # instances
agajju-wgajju	head	10	3
afus-wfus	hand	9	7
asmmed-wsmmed	cold	2	7
ahlig -whlig	belly	5	1
asgrs-wsgrs	shopper bag	2	2
aɣrum-wɣrum	bread	20	2
adbib-wdbib	medical doctor	1	2
amarg-wmarg	nostalgia	1	2
ataj-wataj	tea	1	2
azʕdod-wzʕdod	monkey	26	7
amuʃʃ-wmuʃʃ	cat	1	7
attarru-wttarru	trash can	2	3
agru-wgru	frog	3	2
anzar-wnzar	rain	1	2

Table 4.15: Contiguously pronounced free and construct state alternations of same noun - Masculine

Immediate alternation	English	FS# instances	CS # instances
tadartt-tdartt	little foot	3	1
tafruxt-tfruxt	girl	7	4

Table 4.16: Contiguously pronounced free and construct state alternations of same noun - Feminine

A few unforeseen results

Something that emerged from the analysis of the corpus is finding a significant number of "non-canonical" or rather "non-textbook" instances of the various constructions. In table 6.2 containing an overview of the subject extraction instances, we can see how out of all the relative clauses, two thirds were reduced relative clauses without the *lli* complementizer. Out of all the SRCs, only half this amount displayed a *lli* complementizer. On the other hand, almost 100% of the ORCs had the *lli* complementizer, excluding the headless ones.

The other unexpected datapoint is that there are constructions with the *lli* relative complementizer that are not RCs. These are examples in which *lli* is a demonstrative.

AUNT: Tssnt m-ad-skrħ, lvidjawat lli ʕdlħ-tn kullu.
2SG.know.2G WH-COMP-done.1SG videos DEM edit.1SG-3PL.M all.
You know what I've done, those videos, I edited them all.

Figure 4.1

Here,*lli* does not cover the function of a relative pronoun because if it did, there would not be a resumptive pronoun after edit, but rather a gap. One way in which this *lli* is a relative pronoun is if we hypothesize that the relative clause is silent, an instance of ellipsis, where the entire relative clause gets dislocated to the left periphery and the portion below the *lli* goes unpronounced:

Lvidjwat lli-~~s-am-nnih,~~ ʕdlħtn kullu
Videos REL-~~EP-you-said.1SG~~, edit.1SG-3PL.M all.
The videos that ~~I told you about,~~ I edited them all.

Figure 4.2: Utterance from corpus, March 17th 2022 session

There are relative clauses where **ma**, the wh-phrase, is the complementizer, such as some indirect *wh*-questions, as well as relative clauses that have *a-* as the relativizer, which is the case of headless or light-headed relative clauses.

Maʕʕ (why) is an adjunct *wh*-word that takes an oblique relative clause as a complement. Unlike the rest of the *wh*-words, maʕʕ, and its Moroccan Arabic borrowing counterpart, ʕlaʃ, sometimes used by TB speakers, it does not have *a-* as a complementizer but *lli* + ħ, which is an inessive preposition. I have included it in the data analysis, and coded and counted all of its instances (18 in the entire corpus, excluding instances in which the *wh*-word is pronounced in isolation), to check if there was sufficient information that the child could use to make a more general inference about the two-clause structure of all

wh-questions in TB.

The 'why' case shows that regardless of the specific complementizer if the children can know that there is optionality, and just like some relative clauses might not have *lli*, but rather *a-* or even *ma-* as a complementizer, then they could likewise induce or expect that there may be an "implicit" complementizer. This leads me to a reflection on the various unexpected data points in the input. Relative clauses and clefts may or may not have an audible head. If the child knows this and comes to understand that the wh-construction is underlyingly built on a cleft, and a relative clause, then there could potentially be the expectation that this presence/absence of complementizer is an option for *wh*-questions too. This state of things would "normalize" or make less salient that she only hears some of the *wh*-words with the complementizer to follow and not others.

4.2 Discussion

4.2.1 Relation between corpus frequency and performance in *wh*-questions and RCs

In the sample of the input, looking at the results summarized in tables (Table 4.4), (Table 4.5) and (Table 4.6), the most frequent Ā-construction are *wh*-questions. The least frequent were relative clauses if we exclude reduced relative clauses that are more like adjectival constructions as seen in Chapter 2. Clefts are the second most frequent constructions. In terms of extraction type, oblique *wh*-questions come in first, followed by object extraction and subject extraction is the lowest proportion with only 24% compared

to obliques (34.4%) and object extraction (27%). The pattern observed here suggests in two ways that frequency does not matter to performance. It applies neither to extraction type nor to construction type. Firstly, children performed well in the subject condition as a function of age, despite subject constructions being least frequent. More specifically, subject *wh*-questions are half the amount of object *wh*-questions in the corpus, whereas there are twice as many subject relative clauses (43) as object relative clauses (21), including only RCs with *lli,* the explicit pronoun in the study materials. Secondly, to the point about construction type, although *wh*-questions are most frequent, there was no effect of construction type on their performance, so the children were equally good at the task whether there it was the RC or the WH condition.

We can thus answer the first question raised at the beginning of Chapter 4 on whether the performance on *wh*-questions and RCs in the behavioral task is due to frequency. Of course, since there is no overlap between the children participating in the coprus and the behavioral study, and the corpus is circumscribed, there needs to be a disclaimer that for the answer to be applicable more globally, this sample needs to maintain the same proportions even upon expansion of the duration of the sessions and even if we were to consult a corpus of the parent-child interactions for the children who took part in the behavioral study.

4.2.2 Distribution of the complementizer

In the case of TB input, we are faced with direct positive evidence of the complementizer, but only for a minority of *wh*-words. If the pattern we saw here holds upon expansion of the corpus, i.e. by expanding the corpus to the amount of hours of ambient speech experienced

by a child during her first 5 years of life, and we obtain the same percentage of instances of transparent wh+COMP strings (5.6%), then it will likely be insufficient. This is because the percentage of instances is low to the point that it could potentially be indistinguishable from noise. Noise is the amount of errors, disfluencies, false starts, interruptions inherent to spoken production, that are always present in some measure in spoken language. Noise refers to an amount of errors that is negligible and would not constitute enough data to lead to the inference that the error is actually part of the grammar. Now, there isn't a universal percentage of noise that we can make reference to, so we should attempt to make a test by looking at the present corpus and searching for the percentage of occurrence of a specific error, which could be signalled by a repetition or an immediate correction follow-up mid sentence. We could do this for a couple types of errors and then compare that in proportion to the same number of utterances that formed the total for the percentage of transparent *wh*+COMP strings. If this percentage is comparable, could that mean that the instances of audible complementizer *a-* could be mistaken for noise.

However, co-occurrences between specific *ma*-words and an overt complementizer might be informative enough in the input, no matter how little of it there is, provided that the child is sensitive to this piece of information. The non-randomness of the *ma*-words that occur with an audible complementizer might make it less likely for their co-occurrence to be conceived of as noise.

Nevertheless, construction frequency in the input can affect learnability, regardless of the child's ability to distinguish target sentences from noise (Kush et al., 2021). What this could entail in the case under analysis for us is that the child could maybe undergo at first through a phase in which she hypothesizes that the language has *wh*-questions without a

complementizer given the sheer presence of the *ma*-questions and, the more exposure she gets as a function of time, the more material she has to converge on the target grammar. Pearl (2022) specifies that noise is not an inherent property of speech, but is a property of the intake, from the perspective of the learning child. For some specific cases, the noise does not solely arise as a consequence of the errors and disfluencies audible in the ambient speech, but they can also arise over the course of the children's development and processing abilities which could potentially cause children to "perceive correct data incorrectly" (2022, p. 422).

An alternative approach to this would be to more closely analyze the available distribution. If the complementizer is overt 5% of the time, the issue is whether there is something systematic about the 5% of instances. Given that the amount occurs only in contexts in which the wh-phrase ends in a consonant, then that could be an easily detectable signal if the child is attuned to the phonology.

Based on the transparent strings encountered in the input as sampled in this corpus, *wh*-words all end in a consonant, including the ones that end in a glide, i.e. *mnnaw*'how many' and *managw* 'when', as can be seen in table (Table 4.9). Whereas the opaque strings all feature vowels. This is a pretty clear-cut distinction that the child could make use of to draw the correct inference about the status of the complementizer, namely that it is obligatory. The presence of this internal consistency and systematicity would therefore circumvent the need to do an analysis that compares the rate of occurrence of a non-systematic genuine noisy error.

Now, facing this puzzle, I will replicate here the second question raised at the beginning of this chapter, namely: could the children be using the data from the comple-

mentizer in the focus clefts to learn its obligatoriness in *wh*-questions? Considering table (Table 4.10), about 80% of instances of focus clefts are followed by an audible complementizer. If the children represent a layer above TP and expect that position to be filled, and if they have otherwise independently inferred the unity of Ā-constructions via anti-agreement for instance, then it would be reasonable to expect that the child would apply what she has learned about the complementizer in focus clefts over to *wh*-questions, especially if the data is systematic in the input, however scarce.

In relation to the question about whether the children's knowledge unifies the Ā-dependencies encountered under one construction type, it is now time to discuss the results of the AA distribution in the input.

4.2.3 Distribution of the agreement marking

The data in table Table 4.11 represents the idealized scenario discussed in Chapter 1 in reference to what possible environment would make the anti-agreement marking more salient. The preponderance of feminine agreement in the language would make the presence of masculine singular marking for Ā-extracted subjects particularly noticeable in the speech stream. If we combine this evidence with what emerges from the behavioral study, not only do children perform well in the anti-agreement but they also seem to get a good amount of evidence for it in the input across all Ā-constructions.

Therefore, it is possible to conclude that by ages 3-4 as represented in the input, children know the grammar of anti-agreement. The fact that they perform well in different construction types independent of their respective frequency is also suggestive that

anti-agreement might be achieving its hypothesized job of unifying these constructions or reinforcing this connection for the TB-acquiring child.

4.2.4 Distribution of the CS

As foreshadowed in Chapters 1 and 2, we set out to investigate whether the frequency of visible CS alternations in the input could be sufficient for learners to make use of to acquire this morphological marker that in turn serves the parsing of surface-ambiguous Ā-dependencies. From the results of the corpus analysis, it appears that there is very limited data to use.

Qualitatively, we wanted to know whether there is the right kind of information in the input to help the child make the inference that the *agru-**w**gru* alternation for instance is something to look out for. Out of all the overt CS alternations, we identified 14 specific instances of nouns for which both the Free State and the Construct State instantiation were pronounced within the same utterance or the same time stamp in the transcript, thus providing particularly useful positive evidence of how the CS works. Although there were multiple repetitions for most of these instances in table (Table 4.15, the total overt and alternated CS markings, 49 instances, were less than half of the total CS instances (122). Qualitatively, however, this could be a good enough signal since it does fulfil the requirement of being systematic, in the same way that the audibility of the complementizer is systematic in the case of the 5% of *wh*+COMP strings where the *wh*-word is consonant-final. It might be that amid the seeming scarcity of data about the CS, the quality of it is what makes the CS available to the child in her intake.

If we combine the result of the corpus analysis to the behavioral study, we see that the information obtained about performance and ambient speech is still not conclusive when it comes to the construct state. Based on the parser considerations made in chapter 3, it is not possible to definitively conclude whether children are aware or not of the construct state morphology, because this marker comes 'late' in the online dependency if we reconsider a replication of the object *wh*-question stimulus below:

(56) Manwa gitn hizʕdad a i-ttbujus **w**gru?
Which in.them of.FS.monkeys COMP 3SGM-PROG.kiss CS.frog?
'Which monkey is the frog kissing?'

Recall that the online parsing prediction for (56) was that by the time the child hears the masculine agreement prefix on the verb 'kiss', she'd go with a subject extraction interpretation which is dictated by the hypothesis that the parser is greedy (filling the gap as soon as possible) and immature (struggles to revise). Therefore, for deployment reasons, even if the child were to know that the w- sound change marks the CS and the presence of a subject in post-verbal position, the task used does not make the knowledge state of the child about this construction evident.

It might be that the CS was what was mitigating the children's performance in the object condition. As a matter of fact, the coding of the results of their points in the pointing task paradigm was based on the children's second choice the second time they were allowed to answer the same question. Alternatively, the CS morphology might come too late after the anti-agreement morphology to be deployable by the child. Either way it is necessary to design a study in which children could be tested in a way that circumvents the parser

in the future.

4.3 Conclusion

We set out to answer three broad questions with this chapter, also in relation with the results in chapter 3. Here is the brief overview of the findings.

1. We found that performance on *wh*-questions and RCs in the behavioral study cannot be due to frequency of occurrence. The vast majority of Ā-dependencies are *wh*-questions and only 1/3 of subject RCs shows a *lli* complementizer.

2. Since focus clefts have a higher proportion of transparent [clefted element]+a strings, TB-acquiring children could be making use of this data to infer the biclausality of *wh*-question in their language.

3. Anti-agreement may or may not be playing a role in signalling the similarity across Ā-constructions, but children are competent in it by age 4 and they hear it in ways that are comparable across constructions in the input.

4.4 Future Directions

We explored the acquisition of *wh*-questions in TB from a purely formal standpoint, and this approach also guided the type of information that we looked for in the corpus. Although it is important to recruit help from all fronts when investigating the acquisition of a specific construction or set of linguistic properties, language development does not happen in a social vacuum, a fact that requires discourse context to be part of our studies. One possible

departure from the purely formal approach of only looking for the signal would be to ask whether children could in principle answer subject and object *wh*-questions by being particularly attuned to context, as a way of complementing the work that was presented in this dissertation concerning the possible cues in the input and their deployment. For instance, to know what context is available, a future study could be an analysis of a number of utterances before and after every Ā-construction in the input and then turn these into stimuli for a force task (Orita, 2015). Then, the number of lines before the utterance would be presented to participants in a behavioral task followed by alternatives to choose from in terms of the most contextually relevant *wh*-question, relative clause or cleft.

Since the behavioral task featured in chapter 3 might have partially masked the children's knowledge of the construct state, a study that could circumvent the parser could involve a production task. This could complement the kind of information about the knowledge state that we obtain from a comprehension study.

Chapter 5: Conclusive Remarks

5.1 Summary

In this chapter, I will trace back our steps to summarize the various findings that characterize the contribution of this dissertation.

We set out to study how children learn the grammar of *wh*-questions in a language in which word order of *wh*-questions, clefts and RCs could give rise to ambiguity between subject and object extraction instances. Therefore, for the TB-acquiring child, knowledge of the verbal (AAE) and nominal (CS) morphology is critical in a language of the sort, in order to make inferences about the filler and the gap.

(57) a. Manwa gitn ħifrxan a i-bus-**n** amuʃʃ?
WH in.them in.boys COMP 3SGM-kissed-**aae** FS.cat
'Which boy kissed the cat?'

b. Manwa gitn ħifrxan a i-bus **w**muʃʃ?
WH in.them in.boys COMP 3SGM-kissed CS.cat
'Which boy did the cat kiss?'

Alongside this issue of surface ambiguity is the issue that *wh*-questions are characterized by an obligatory complementizer that the phonology makes audible only in specific and rare circumstances. The rule against sequences of vowels makes the string of the po-

tentially most frequent wh- word in Berber *ma* eclipse the a- complementizer. This could potentially become a problem if *ma* was a particularly prominent *wh*-word in the input, because it could potentially entail the lessening of the signal constituted by instances of *wh*-questions

This set of properties is part of what makes TB stand out as a special language in the morphosyntactic literature and one worth studying from an acquisitional standpoint.

We will now review in turn the findings in relation to the three areas we asked about at the beginning of the dissertation:

1. Is the performance on *wh*-questions and relative clauses due to frequency?

2. Can children use data from complementizer usage in focus to learn its obligatoriness in *wh*-questions?

3. What role does anti-agreement play in signaling the similarity of Ā-constructions

4. (raised in Chapter 2) Is the CS signal informative enough in the corpus that we could plausibly posit that the child tracks it to infer extraction of the object?

5.1.1 Ā-dependencies: alike in the behavior, unalike in distribution

We probed the children's knowledge of the properties in this preamble by testing how children performed in the comprehension of reversible action subject and object *wh*-questions and relative clauses as reported in Chapter 3. We found that their knowledge of *wh*-questions and relative clauses is comparable, because upon fitting a glmm model we found

that the construction type was not predictive of the children's performance. Recall the findings about the COMP-based and connectivity-based asymmetries between *wh*-questions and RCs in Chapter 2. Although the outward expression of *wh*-questions and relative clauses shows divergences in the optionality of the COMP for instance, this does not seem to be reflected in the children's treatment of them.

The parallel performance of children in *wh*-questions and RCs seems remarkable also when we consider how the distribution of these constructions is in the corpus. Therein, we showed that *wh*-questions constituted the vast majority of Ā-constructions. Relative clauses of the kind tested in the behavioral study, i.e. of the type that show the optional complementizer constitute less than half of the instances of relative clauses that do show the relative pronoun *lli*.

5.1.2 The obligatory-but-sometimes-masked complementizer story

Although TB *wh*-words are always followed by an obligatory complementizer, independent phonological properties of the language, i.e. a ban on sequences of idential vowels, potentially make *wh*-questions with the most prominent *wh*-word, *ma*, ' who' and 'what', look like there is no complementizer. This specific puzzle raises the question of what learning mechanisms and what available data can help the child learn *wh*-questions that are always followed by an obligatory complementizer when the language makes the complementizer invisible in most cases.

By looking at the corpus we found that *wh*-questions of the kind that make the complementizer audible, i.e. end in a consonant are among the rarer in the input, 5.6%

to be exact, against the 91.1% of wh-phrase instances that make it inaudible. Moreover, although there is high variety across the 5.6%, i.e. this percentage is made up by five separate instances of *wh*-words that end in a consonant. In contrast to that, The opaque cases are mostly made up of the same exact *wh*-word, ma translated as 'what' or 'who', which makes up 79.1% of the total of *wh*-words in the input.

In the behavioral study, in the *wh*-question condition, children were faced with manwa gitn 'which one of', a *wh*-word that bares the complementizer and is only featured 1.5% of the time based on the results in Chapter 4, children performed well independent of this aspects. This raises the question of whether children can use the data from complementizers in this 5.6% alone or if they draw from other information in the input.

If children were to use the overt complementizer data from the wh-instances alone, it would be little and possibly noise levels of infrequent, but at least qualitatively there is a systematic nature to the 5.6%. As remarked all *wh*-words in these instances end in consonants and all the *wh*-words in the opaque cases end in a vowel.

This takes us to the other question we asked in Chapter 1 about whether data from the complementizer in focus clefts could also contribute to this pattern recognition in the input that tells them that the complementizer is there, it is just masked by phonology. We looked for the data about the complementizer in focus clefts and found that the ratio of opaque to transparent is the reverse of the *wh*-questions case. We found that cleft constructions make up, across extraction types (subject, object and oblique), 14.7% of A'-dependencies, 171 in count. More than 75% of cleft constructions show a consonant-ending clefted constituent, where the vast majority of clefted constituents were demonstratives, possessives and other indexicals which is attributable to the highly salient discourse referents by all

the participants and the use of the clefts as contrast clefts.

If the child naturally expects that there are parallelisms to be found across Ā-constructions, then it would make sense for a child to draw from the commonalities of where the Despite the qualitative superiority of the clefts input, this alone would not be the answer to the question of how the child learns about the obligatory cleft/wh-complementizer, since it has to be combined with what the child has observed in the skewed wh+COMP data. The child could be piecing together the complementary information provided by the 5.6% of transparent strings in the *wh*-questions with the 21% (36 out of 171) of focus clefts with opaque complementizer strings and infer that what these instances have in common is that there is an association between the audibility of the complementizer and the sound that the fronted constituent ends in. The children would possibly need to have encountered instances of some of the same words in environments that do not mask the word-final sound to reinforce this inference.

Detection of a Subject-Object Asymmetry

Related to the notion of frequency, from the results of the corpus analysis, object *wh*-questions are much more frequent than subject *wh*-questions. Although this dissertation was not set up to answer this question, it is worth mentioning that at first glance, these preliminary numbers go against the expectations of the contribution of frequence to the subject-object asymmetry for instance (Lau & Tanaka, 2021). Frequentist approaches tend to attribute the subject-object asymmetry to the preponderance of subject extraction in the input. There are studies in which frequency is considered to be pretty central and important predictive of the earliest relative clause typology that gets acquired (Diessel and

Tomasello 2001, Diessel 2004, Kidd et al 2006, Roland). So, by extension, if a language shows significantly more instances of object-Ā-extraction in the input, children should be expected to perform as well if not better at this extraction type. This is however not what we found in the behavioral study. Children perform at chance in the object condition despite the preponderance of object Ā-constructions in the input. This would suggest that children's performance in the study might not reflect their real competence. The at-chance results are, instead, consistent with a picture in which children's performance was due to an active-filling parser with immature revision in the face of discriminating data. If children attend to what is most frequent in the input, then there would be no a priori reason why TB-acquiring children would not have knowledge of this salient extraction site in the age ranges involved here. Therefore, the object condition results are compatible with a situation in which the children know the grammar of anti-agreement, as well as the construct state, but have a parsing mechanism that is not sufficiently mature to make use of the verbal and nominal information in embedded position to revise their misparse about the role of the extracted NP.

5.1.3 The Anti-Agreement story

The picture for the acquisition of anti-agreement is rosy both behaviorally and in terms of its presence and signal quality in the input. In the pointing task paradigm study, children show a significant improvement in subject *wh*-questions and relative clauses as a function of age. Children were much better at subject extraction than object extraction and object extraction was at chance, which suggested that children's success in the subject condition

was unlikely to be due to extralinguistic strategies such as default salience of agents in events, an explanation often used in the subject-object asymmetry literature.

In terms of the corpus, the majority of standard subject-verb agreement instances are feminine singular, 31% of the total, with the masculine singular as a runner up with 25% of subject-agreement instances. As discussed, this was the ideal scenario for anti-agreement to be particularly salient in the speech stream. Moreover, there was less than 1% of instances of future anti-agreement, which is indistinguishable from the masculine standard agreement, though even if it has a more sizeable presence, the proportion of feminine referents discussed daily in the input in the presence of the two baby sisters in the corpus would not have had a detrimental effect to the acquisition of AAE.

Recall how, in Chapter 1, we asked what role anti-agreement plays in signaling the similarity of Ā-constructions. It seems that rather than playing an active role in signalling to the child that the relevant constructions to the AAE share Ā-ness, anti-agreement could mean empirical confirmation in the language that these constructions share an abstract quality in general that then the child would need to combine with what they know about classic hallmarks of Ā-movement such as island constraints for instance to then derive the complete picture.

5.1.4 The Construct State Story

We saw in Chapter 2 that the construct state is a nominal marking specific to canonically post-verbal subjects, both in matrix simplex clauses and in embedded clauses. We asked in Chapter 2 whether the construct state is a signal the child could be using to keep track

of the extraction site in an object-fronting filler gap dependency. This is because when the object is extracted, the child could get evidence of this fact by identifying the CS marking on the last NP in the dependency to retroactively infer that the first element is an object. This signal is to be combined with the presence/absence of AAE. We have found that in the input, the construct state of the useful kind, i.e. the marker on the subject that can help the child track the gap is a particularly limited signal. There are 143 instances of the construct state in the input out of a total of 1639 nominal forms, 8.7% of the total. Out of these constructions, 15% are opaque and 75% is made up of the PP complements that were presented in the chart in Chapter 2.

We have thus far argued, in this Chapter, for ways in which small percentages of a construction in the input can be salvaged by the internal consistency and systematicity coupled with innate knowledge of abstract similarity. The complication in the nominal morphology case is that 75% of the occurrences of CS cannot be transferred to the function of the CS subject. The alternate occurrence of CS in the language, where the complements of specific prepositions and the complements of numbers are in the CS, do not offer much help in inferring the position of other NPs in the dependency. This is a function unlike the subject CS which is relevant to subject and object Ā-extraction. Moreover, the CS signal in the input is partially opaque meaning that of the CS examples that are useful for our questions, a sizeable proportion are borrowings, so the vowel to glide transformation in the masculine and the vowel elision in the feminine singular are not available. This was the case of words such as *mama*, *baba* and *xalti*, respectively ‘ mom’, ‘ dad’ and ‘ auntie’, borrowed from French and Arabic historically and often used in subject position in the relevant instances.

The CS input is also uninformative as significantly skewed towards the masculine CS, with the masculine singular constituting 85% of CS instances and the feminine singular the remaining 15%.

In terms of the behavioral study, because of the stimuli design, which was based on the comprehension of an auditory stimulus that is incrementally attended to, it might be the case that in combination with what is known about the child parser, the CS signal might come late and not be deployable by the child. We would need more direct ways less taxing on the working memory of a child to unmask the specific knowledge that they have of this marker. A child production-focused study of the corpus of child-directed TB would also allow us to learn more about what the child knows.

All in all, there is not sufficient information from the input or the study to tell what the role of nominal morphology plays for the child who is converging on adult-like strategies of tracking the extraction site of the object and the subject. For this reason, priority in the future directions portion of this chapter will be placed on discussing a more direct way of probing the knowledge of the CS in children.

5.1.5 Wrapping it up

In conclusion, in light of what we outline about the grammar in Chapter 2, the behavioral study in Chapter 3 and the corpus analysis in Chapter 4, the findings are compatible with a learning theory where the child has an abstract knowledge of the $\bar{A}$-dependency category that would allow them to face the challenges posed by unauspicious input. Keeping track of the sistematicities alone without having any notion of the abstract $\bar{A}$-dependency category

would not result in the behavioral study results, because under the alternative view in which a child infers underlying similarities through input exposure, the child would possibly be thrown off by an overt post-wh-complementizer that is so rare in the environment. In the corpus, there was a considerable scarcity of *wh*-words that display a complementizer. If this frequency pattern is confirmed for RCs, clefts and *wh*-questions in a significantly more expanded corpus, in light of the particular scarcity of many of the we suggest that children's ability not to treat the three constructions as separate would constitute yet another point of evidence in support of the nativist theory of language development, contra constructionist approaches which would hypothesize the inverse trajectory where the child first hypothesizes that closely related constructions that are superficially dissimilar should be treated as separate phenomena (Goldberg, 1995).

5.2 Future directions

5.2.1 Construct State Acquisition

As was alluded to in the future directions of Chapter 4, a way of circumventing the parser complications that obscure the children's knowledge of the grammar would be to devise a study where their cognitive inihibition and control abilities are taxed as little as possible.

The study will be focusing on the child's knowledge of the construct state morpheme and its role in the structure. We would have to test how helpful this cue is by itself in the simplest clause possible.

(58) a. T-umz tfruxt ikzin.
3SGF-held CS.girl FS.dog.

'The girl held the dog'.

b. T-umz *pro* tafruxt
3SGF-held *pro* FS.girl.

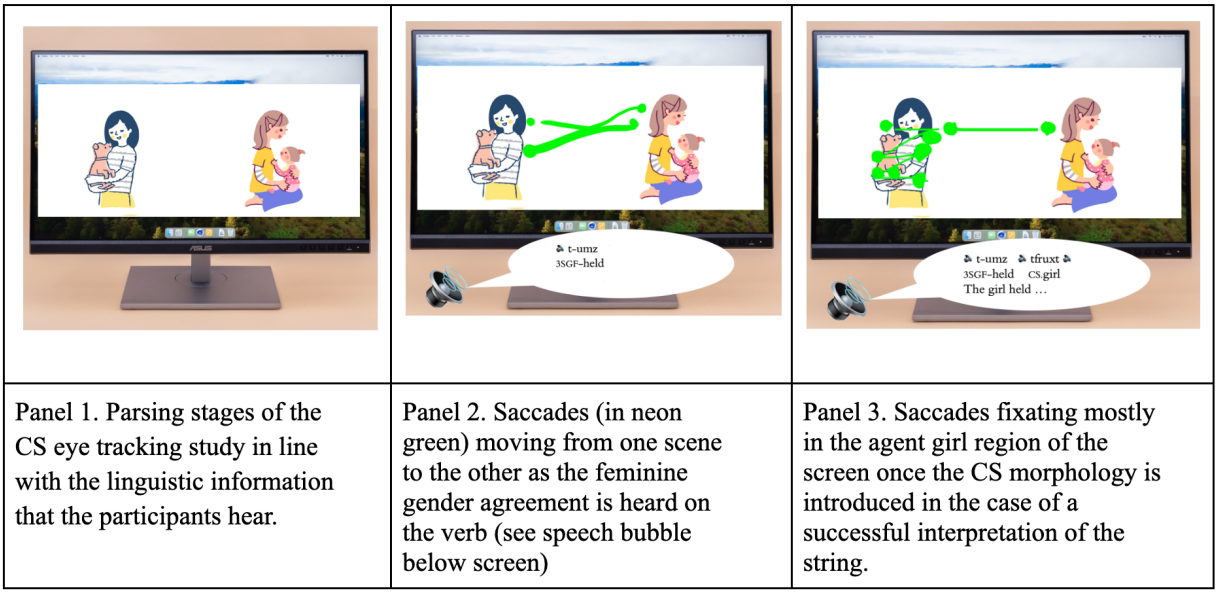
'She held the girl.'

Tafruxt is the free state form of girl, and tfruxt is its construct state form, that, as you recall, only shows up if the NP is a post-verbal subject (or the complement of some prepositions, e.g. n-tfruxt, of the girl). To test the knowledge of the child of this construction, we need to make use of the most ambiguous possible string. To achieve so, I propose to test the children on the 'tumz tfruxt' vs. 'tumz tafruxt' segments of the two sentences in (58a) and (58b), and track their online interpretation of the string as they are looking at two scenes where, on the one side 'girl' is an agent and the syntactic subject and on the other side, a 'girl' is a patient and the syntactic object.

Figure 5.1: Visual stimuli example

Since object drop is ungrammatical in TB (as well as other Berber languages), to

control for acceptability, we need to introduce something like white noise at the end of string (58a) before the object is pronounced. To counterbalance the examples, it will have to be introduced after string 2 as well. This is shown as the loudspeaker emoji the end of each string in the stim on panel 3 of (5.2). I also add it in between the verb and the NP to control for the 'pro' presence in example (58b).



Panel 1. Parsing stages of the CS eye tracking study in line with the linguistic information that the participants hear.	Panel 2. Saccades (in neon green) moving from one scene to the other as the feminine gender agreement is heard on the verb (see speech bubble below screen)	Panel 3. Saccades fixating mostly in the agent girl region of the screen once the CS morphology is introduced in the case of a successful interpretation of the string.

Figure 5.2: Experimental timecourse

In the visual world paradigm, children will be presented with a screen with two images simultaneously, that depict two separate scenes (5.2). On the one left hand side a young girl is depicted holding a dog. On the opposite side, a woman is holding a young girl (measures will be taken to ensure that the designs of the characters make it as clear as possible that the holder of the dog and the baby being held are both plausible candidates for the 'young girl' label). Then, a prerecorded sentence will start playing. The verb will be pronounced first and since it carries feminine agreement, the hearers are expected to look at both sides. panelure 2 depicts the hypothesized saccades (movement of the eye fovea between two or

more points) in this context. The NP pronounced first after the verb in Berber could be the subject or the object. The subject is fully optional, but the object is not. However, for the sake of this experiment, we will be inserting noise after the first NP so that the only thing the speakers can rely on is their knowledge of the nominal state alternation to know whether to look at the agent or the patient of the 'holding'. Suppose the next NP shows the construct state, which reliably marks the subject. In that case, we expect the participants to turn their attention in full to the young girl holding the dog (panel 3). We would use the reaction times in milliseconds of this study as a measure of Berber-acquiring children's integration of morphological information as the speech stream is coming in.

Appendices

.1 APPENDIX A - Glossed Stimuli

Example of full trial from RCO (Object Relative Clause) List 1A session. This is to show what screens and stages come before the familiarization phase, which is whether the schema of a trial starts from in Chapter 3.

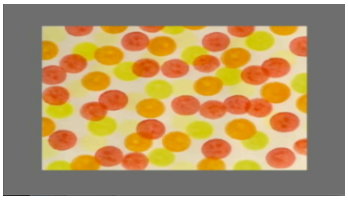


Figure 3: Opening screen

The first sequence is a screen with increasing coloured circles that fill up the background as classical music plays.

Figure 4: Blank screen

This screen is followed by a visual introduction of one of the three characters in motion.

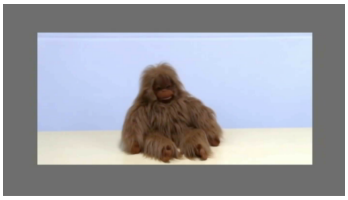


Figure 5: Introduction of brown monkey

This screen is accompanied by the following utterance pronounced by a native speaker:

(59) Smoqql! azʕdod aqhwi aynn! Tunnit zʕdod aqhwi?

Look! FSmonkey brown DEM! Seen-2SG FSmonkey brown?

'Look! It's a brown monkey! Do you see the brown monkey?'

Figure 6: Blank screen

After a brief blank screen, the chimpanzee is introduced with the following exclamation.

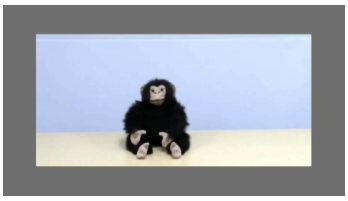


Figure 7: Introduction of black monkey

(60) Smoqql za! Azʕdod asggan aynn! Tunnit azʕdod asggan?

Look EMPH! FSmonkey black DEM! Seen-2SG FSmonkey black?

'Hey look! It's a black monkey! Do you see the black monkey?'

Figure 8: Blank screen II

After a brief blank screen, the frog is introduced with the following exclamation.

Figure 9: Introduction of frog

(61) Smoqql za! Agru jaynn! Tunnit agru?

Look EMPH! FSmonkey black DEM ! Seen-2SG

'Wow look! It's a frog! Do you see the frog?'

This video square dissolves into grey again and then the upcoming scene of hugging begins.

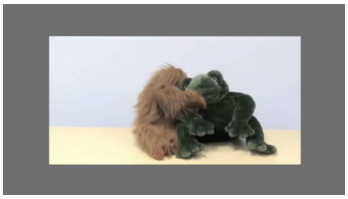


Figure 10: Brown monkey hugs frog

Smoqql za! Kra a-ittʕnnag kra. Tzrittn? Kra a-ittʕnnag kra.

Look! Hugging! Somebody's hugging somebody

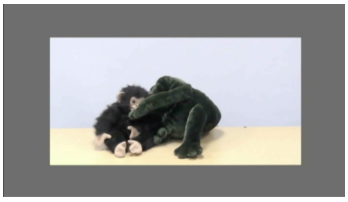


Figure 11: Frog hugs black monkey

.2 APPENDIX B - Representative transcription samples from some corpus sessions

Types of Anti-Agreement utterances found in the input

1. **Subject Cleft and Contrastive focus**

(62) Kmmi a ur i-ri-n!

You.SGF COMP NEG AAE-want-AAE

'YOU don't want to'

(63) Xtad a i-skar-n "bomm"!

DEM-SGF COMP AAE-make-AAE

'THIS ONE goes "boom!"'

(64) Jan ka a i-ga-n winn-m.

one only COMP AAE-be-AAE of-POSS.SGF

'Only one (of them) is yours.'

1. **Adjectival constructions (subtype of SRCs)**

(65) lħbaqq i-ʒʒa-n

basil AAE-be.perfumed-AAE

'fragrant basil/ basil that is fragrant'

(66) xtad lli i-mzzij-n

DEM-SGF REL AAE-be.small-AAE

'the small one/ the one who is small'

(67) tiglaj lli i-nwa-n-in

egg-PLF REL AAE-be.cooked-AAE-PL

'eggs that are cooked'

Bibliography

Adger, D., & Ramchand, G. (2005). Merge and move: Wh-dependencies revisited. *Linguistic inquiry*, *36*(2), 161–193.

Arnon, I. (2005). Relative clause acquisition in hebrew: Towards a processing-oriented account. *Proceedings of the 29th Boston University Conference on Language Development (BUCLD)*, 37–48.

Baier, N. B. (2018). *Anti-agreement*. University of California, Berkeley.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belkadi, A. (2010). *Pronominal and clitic systems in taqbaylit berber*. University of London, School of Oriental; African Studies (United Kingdom).

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of arabic dialects*. Oxford University Press.

Borer, H. (1988). On the morphological parallelism between compounds and constructs. *Yearbook of morphology*, *1*(1988), 45–66.

Boukhris, F. (1998). *Les clitiques en berbère tamazighte, unpublished ph. d* [Doctoral dissertation, dissertation, University Mohamed V, Rabat].

Bowers, J. (1993). The syntax of predication. *Linguistic inquiry*, *24*(4), 591–656.

Brandi, L., & Cordin, P. (1989). Two italian dialects and the null subject parameter. In *The null subject parameter* (pp. 111–142). Springer.

Brown, R. (1973). Development of the first language in the human species. *American psychologist*, *28*(2), 97.

Calabrese, A. (1987). Focus structure in berber: A comparative analysis with italian. *Studies in Berber Syntax, Lexicon project working papers*, *14*, 103–120.

Caponigro, I., Torrence, H., & Maldonado, R. Z. (2020). *Headless relative clauses in mesoamerican languages*. Oxford University Press.

Chaker, S. (1985). Adverbe. *Encyclopédie berbère*, (2), 157–163.

Chomsky, N. (1977). On wh-movement. *1977*, *65*.

Chomsky, N. (1981). Lectures on government and binding. *Studies in generative grammar*, *9*, 1995.

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic inquiry*, *8*(3), 425–504.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. *syntax*, *1*, 506–569.

Chung, S. (1981). Transitivity and surface filters in chamorro. *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 311–332.

Citko, B. (2004). On headed, headless, and light-headed relatives. *Natural Language & Linguistic Theory*, *22*(1), 95–126.

Collins, P. C. (2002). *Cleft and pseudo-cleft constructions in english*. Routledge.

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. *Proceedings of NELS*, *15*, 94–110.

Crain, S., & Thornton, R. (1998). Investigations in universal grammar.

Dell, F., & Elmedlaoui, M. (2002). *Syllables in tashlhiyt berber and in moroccan arabic* (Vol. 2). Springer Science & Business Media.
Demuth, K. (2022). The acquisition of sesotho. In *The crosslinguistic study of language acquisition* (pp. 557–638). Psychology Press.
Drouin, J. (1996). Les formes participiales en berbère: Essai de dialectologie comparée. *Littérature orale arabo-berbère*.
El Hankari, A. (2021). *Aspects of the morphosyntax of tarifit berber*. Cambridge Scholars Publishing.
El Mountassir, A. (1999). *Initiation au tachelhit: Langue berbère du sud du maroc*. L'Asiathèque-maison des langues du monde.
Engdahl, E. (1983). Parasitic gaps. *Linguistics and philosophy*, 5–34.
Ennaji, M., & Sadiqi, F. (2004). A grammar of amazigh. *Fès: Faculty of Dhar El Mehraz*.
Ennaji, M., & Sadiqi, F. (1986). The syntax of cleft sentences in berber. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *10*(1), 53–77.
Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current directions in psychological science*, *26*(5), 451–457.
Ervin-Tripp, S. (1970). Structure and process in language acquisition. *Monograph series on Language and Linguistics*, *23*, 313–344.
Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J. S., & Bates, E. (2006). Macarthur-bates communicative development inventories. *Database record*.
Flynn, S. (1988). Second language acquisition and grammatical theory. *Linguistics: The Cambridge Survey*, *2*, 53–73.
Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of a-bar dependencies. *Lingua*, *119*(1), 67–88.
Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, *23*(3), 234–260.
Gensler, O. D. (1993). *A typological evaluation of celtic/hamito-semitic syntactic parallels*. University of California, Berkeley.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
Green, G. M. (1971). Notes on clefts and pseudo-clefts and other related matters.
Guerssel, M., & Hale, K. L. (Eds.). (1987). *Studies in berber syntax* (Vol. 14). The MIT Press.
Guerssel, M. (1984). *Descriptive syntax of berber*. The MIT Press.
Guerssel, M. (1992). On the case system of berber. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *37*(2), 175–196.
Gutova, E. (2015). Baby talk in the maghreb. *Arabic Varieties: Far and Wide. Proceedings of the 11th International Conference of AIDA–Bucharest*, 281–290.
Haegeman, L. (2020). *Introduction to government and binding theory*. Wiley-Blackwell.
Hamburger, H., & Crain., S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language development, vol. 1: Syntax and semantics* (pp. 245–274). Hillsdale, Lawrence Erlbaum.
Hewett, M. (2022). How to tell base-generated and movement-derived resumptives apart matthew hewett (mhewett@ uchicago. edu)/the university of chicago lsa 2022, january 6, 2023 1. introduction (1) resumption in tunisian arabic restrictive relatives & wh-questions.
Hewett, M. R. (2023). *Types of resumptive ā-dependencies* [Doctoral dissertation, The University of Chicago].
Heycock, C., & Kroch, A. (1999). Pseudocleft connectedness: Implications for the lf interface level. *Linguistic inquiry*, *30*(3), 365–397.
Higgins, F. R. (1979). *The pseudo-cleft construction in English*. Garland.

Hovdhaugen, E. (1989). *A handbook of the tokelau language* (Vol. 77). Oxford University Press.
Huang, C. J. (1982). Move wh in a language without wh movement. *The Linguistic Review*.
Hyams, N., Ntelitheos, D., & Manorohanta, C. (2006). Acquisition of the malagasy voicing system: Implications for the adult grammar. *Natural Language & Linguistic Theory*, *24*, 1049–1092.
Keenan, E. L., & Manorohanta, C. (2001). A quantitative study of voice in malagasy. *Oceanic linguistics*, 67–84.
Kenstowicz, M., & Wahba, W. (1983). Wh-in-situ constructions in egyptian arabic. *Current approaches to African linguistics*, *2*, 261–81.
Krifka, M., & Schenner, M. (2018). *Reconstruction effects in relative clauses* (Vol. 75). Walter de Gruyter GmbH & Co KG.
Kush, D., Sant, C., & Strætkvern, S. B. (2021). Learning island-insensitivity from the input: A corpus analysis of child-and youth-directed text in norwegian. *Glossa: a journal of general linguistics*.
Lahrouchi, M., & Kern, S. (2018). From babbling to first words in tashlhiyt language acquisition: Longitudinal two-case studies. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *63*(4), 493–526.
Lasnik, H., & Saito, M. (1984). On the nature of proper government', linguisticinquiry15, 235-289. *Lasnik23515Linguistic Inquiry1984*.
Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa: a journal of general linguistics*, *6*(1).
Lee, K.-O. (1991). *On the first language acquisition of relative clauses in korean: The universal structure of'comp'*. Cornell University.
Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist.*, *1*(1), 333–353.
Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.
Liter, A., Grolla, E., & Lidz, J. (2022). Cognitive inhibition explains children's production of medial wh-phrases. *Language Acquisition*, *29*(3), 327–359.
Lust, B. C. (2006). *Child language: Acquisition and growth*. Cambridge University Press.
Makhad, H. (2004). The syntax of events and temporality: A comparative approach. *Unpublished Ph. D. dissertation, University Mohamed V, Rabat*.
McDaniel, D., Chiu, B., & Maxfield, T. L. (1995). Parameters for wh-movement types: Evidence from child english. *Natural Language & Linguistic Theory*, *13*(4), 709–753.
McDaniel, D. S. (1986). *Conditions on wh-chains*. City University of New York.
Mettouchi, A. (2009). 13. mood and modality in berber. *Modals in the Languages of Europe: A Reference Work*, *44*, 431.
Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 756.
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281.
Oda, K. (2008). V1 and wh-questions: A typology. In *Verb first: On the syntax of verb-initial languages* (pp. 107–133). John Benjamins Publishing Company.
Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, *22*(2), 158–192.
Omari, N. (2001). The syntax of negation in tashelhit: A comparative approach. *Unpublished Ph. D. dissertation, Mohamed V University, Rabat, Morocco*.
Orita, N. (2015). Computational modeling of the role of discourse information in language production and language acquisition.

Ouali, H. (2011). *Agreement, pronominal clitics and negation in tamazight berber: A unified analysis*. A&C Black.
Ouali, H. (2020). Verb morphology, tense, and aspect in tamazight. *McGill Working Papers in Linguistics*, *26*.
Ouhalla, J. (1988). *The syntax of head movement: A study of berber* [Doctoral dissertation, University of London].
Ouhalla, J. (1993). Subject-extraction, negation and the antiagreement effect. *Natural Language & Linguistic Theory*, *11*(3), 477–518.
Ouhalla, J. (2005). Agreement features, agreement and antiagreement. *Natural Language & Linguistic Theory*, 655–686.
Ouhalla, J., & Hankari, A. E. (2015). Wh-cliticisation: The derivation of operator-variable links and wh-words in berber. *Corpus*, (14), 235–262.
Ouhalla, J., Lightfoot, D., & Hornstein, N. (1994). Verb movement and word order in arabic. *Verb movement*, *41*(72), 10–1017.
Outahajala, M., Zenkouar, L., Rosso, P., & Martí, A. (2010). Tagging amazigh with ancorapipe. *Proceeding of the Workshop on Language Resources and Human Language Technology for Semitic Languages*, 52–56.
Ovans, Z. L. (2022). *Developmental parsing and cognitive control* [Doctoral dissertation, University of Maryland, College Park].
Paul, I. (2001). Concealed pseudo-clefts. *Lingua*, *111*(10), 707–727.
Pearl, L. (2022). Poverty of the stimulus without tears. *Language Learning and Development*, *18*(4), 415–454.
Perkins, L., & Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, *27*(1), 98–115.
Potsdam, E. (2009). Austronesian verb-initial languages and wh-question strategies. *Natural Language & Linguistic Theory*, *27*, 737–771.
Potsdam, E., & Polinsky, M. (2011). Questions and word order in polynesian. *Topics in Oceanic morphosyntax*, *27*, 121–153.
Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636–643.
Reinhart, T. (1983). Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy*, 47–88.
Rizzi, L. (1997). The fine structure of the left periphery. *Elements of grammar: Handbook in generative syntax*, 281–337.
Ross, J. R. (1967). Constraints on variables in syntax.
Schallert, O., & Bidese, E. (2021). Doubly-filled comps in alpine varieties: A contrastive study on embedded questions in germanic and romance. *STUF-Language Typology and Universals*, *74*(1), 129–161.
Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early understanding of subject and object wh-questions. *Infancy*, *4*(3), 423–436.
Sheil, C. M. (2016). *Scottish gaelic clefts: Syntax, semantics and pragmatics*. University of California, Berkeley.
Shlonsky, U. (1987). Focus constructions in berber.
Shlonsky, U. (1992). Resumptive pronouns as a last resort. *Linguistic inquiry*, *23*(3), 443–468.
Snyder, W. (2007). *Child language: The parametric approach*. OUP Oxford.
Soltan, U. (2011). On strategies of question-formation and the grammatical status of the q-particle huwwa in egyptian arabic wh-questions. *University of Pennsylvania working papers in linguistics*, *17*(1), 114–124.
Stoyanova, M. (2008). *Unique focus: Languages without multiple wh-questions*. John Benjamins Pub. Co.

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition*, *4*(1-2), 5–48.
Tabti, R. (2013). Corpus sur le prélangage de l'enfant kabyle etude sociolinguistique. *Iles d imesli*, (5), 97–134.
Thornton, R. J. (1990). *Adventures in long-distance moving: The acquisition of complex wh-questions*. University of Connecticut.
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.
Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of child language*, *4*(2), 211–224.
Van Gelderen, E. (2013). *Clause structure*. Cambridge University Press.
Wagers, M., Borja, M. F., & Chung, S. (2015). The real-time comprehension of wh-dependencies in a wh-agreement language. *Language*, 109–144.
Wahba, W. A.-F. B. (1984). *Wh-constructions in egyptian arabic*. University of Illinois at Urbana-Champaign.