

The maximal size of infinitives: a truncation theory of finiteness*

Deniz Satık, Harvard University
deniz@g.harvard.edu

This paper argues for the following finiteness universal: an infinitive cannot co-occur with a high complementizer (such as *that* in English). Although such an observation may seem trivial, assuming Rizzi (1997)'s articulated CP allows one to redefine *that*. In a vein similar to Wurmbrand and Lohninger (2019), I propose that infinitives can come in different sizes. This paper combines Pesetsky (2021)'s arguments that finiteness is a matter of clause size together with truncation theories of infinitives such as Shlonsky and Soare (2011)'s to argue for a novel understanding of finiteness, proposing precise and falsifiable definitions for finite and nonfinite clauses. Based on a crosslinguistic survey of several different languages belonging to many different language families, I present a theory of finiteness under which a clause is defined as nonfinite iff its ForceP/CP2 layer has been truncated, and finite iff it is untruncated. Although derivational theories of finiteness predict this generalization, infinitives come in at least eight different sizes crosslinguistically. Beyond arguing for this finiteness universal, this paper also discusses the cartographic predictions that result from maximal size of infinitives in a given language. Under this definition of finiteness in terms of the truncation of the C domain, I will argue that the surprising phenomenon of finite control does not exist.

Keywords: finiteness, complementizer, infinitive, clause size, left periphery

1 Introduction

One of the most poorly understood notions in generative grammar is the notion of finiteness. For descriptive grammarians, this is relatively simple: finiteness is seen as a property of the verb. As Nikolaeva (2007) points out, in Latin, the finite/nonfinite distinction was originally just the presence or absence of agreement of the verb, though other properties were later considered to be relevant for finiteness as well—the most important of which is tense.

This works straightforwardly to analyze finiteness within a European context, but as we will soon see, such a definition of finiteness cannot be extended crosslinguistically. Landau (2013) lists a number of languages with inflected infinitives, such as Turkish, Brazilian Portuguese, Basque, Hungarian and Welsh which have nonfinite complements that are inflected for agreement. An example from European Portuguese is provided in (1) below from Raposo (1987):

- (1) Será difícil [eles aprovar-em á proposta].
It will.be.difficult they to.approve-3PL the proposal
'It will be difficult [for them to approve the proposal].' European Portuguese

*Thanks first and foremost to Susi Wurmbrand and David Pesetsky for extensive help. I would also like to thank Jonathan Bobaljik, Jim Huang, Jay Jasanoff, Stefan Keine, Idan Landau, Ur Shlonsky and Hoskuldur Thrainsson for helpful comments. All errors are mine. This paper is dedicated to Erdoğan Satık.

One, as Raposo (1987) does, may claim that agreement is not the relevant property for the finite/nonfinite distinction: instead, the distinguishing property is tense. But this does not work either. In Tamil, as McFadden and Sundaresan (2014) points out, we see the opposite scenario with a gerundival participle in (2) below, in which the embedded clause is embedded for tense, but lacks agreement, yet appears to be nonfinite given its inability to stand alone:

- (2) Raman_i [EC_i Seetha-*vae* naaleeki paar-pp-adaagae] so-nn-aan.
 Raman EC Seetha-ACC tomorrow see-FUT-GER-ACC say-PST-3MSG
 ‘Raman_i spoke of [EC_i seeing Seetha tomorrow].’ Tamil

Another property that has been commonly assumed to distinguish finite clauses from nonfinite clauses is whether the clause licenses overt subjects, such as by Chomsky (1977a). For example, A-movement out of a finite clause is not possible, as in (3a), but it is from a nonfinite clause, as in (3b). Although in the past such a distinction was tied to Case and agreement, in more recent proposals such as by Pesetsky (2021) it is tied to clause size: (3a) involves a clause as large as CP, which precludes the possibility of subject extraction, whereas (3b) involves a clause that is smaller than CP, which allows the possibility of subject extraction:

- (3) a. *David_i seems [that t_i likes exfoliation].
 b. David_i seems [t_i to like exfoliation].

Ultimately, I will adopt a similar line of reasoning. Regardless, it seems *prima facie* possible that subject licensing is related to finiteness, especially in languages like Mandarin which have no inflectional morphology whatsoever, and hence, no tense and agreement. Such languages have remained puzzling for theories of finiteness for decades. As has been noted by many in the literature on Mandarin, clausal complements of verbs such as *like* cannot have an overt subject or a null pronoun that does not refer to the matrix subject. In other words, we seem to observe a controlled PRO in the complements of such sentences, as in (4) from Ussery et al. (2016) below. This indicates there might be a finite/nonfinite distinction in Mandarin after all:

- (4) Xiaoming xihuan (*ta) chi shoushi.
 Xiaoming like he eat sushi
 ‘Xiaoming likes to eat sushi.’ Mandarin

McFadden and Sundaresan (2014) raises further challenges for this line of reasoning, however, based on evidence from languages such as Tamil, Sinhala, Modern Irish and Middle English which have clauses that are clearly nonfinite—that lack tense and agreement—yet allow subjects to be licensed, as in the Modern Irish example in (5) below.

- (5) Ghoillfeadh se orm [tu me a ionsai].
 would.bother it on.me you.ACC me INF attack
 ‘It would bother me for you to attack me.’ Irish

As Raposo (1987) points out, even inflected infinitives in European Portuguese allow overt pronominal subjects—which Raposo ties to agreement. Regardless, McFadden and Sundaresan undermine the correlation between subject licensing and finiteness, not just for simpler models of subject licensing via Agreement in the GB and Minimalist framework like Raposo (1987)’s, but also for Landau (2004) and Szabolcsi (2009), who assume a more complex relationship between tense, agreement and subject licensing in clauses.

Another potential distinguishing property, briefly alluded to above in our discussion of Tamil, is the ability of a clause to stand alone. This seems difficult to reconcile with the existence of imperatives like *Catch her!* which, even in languages with very rich inflectional morphology, have little inflection, and yet can stand alone. Therefore, although I have simplified the empirical terrain somewhat, many works, such as Nikolaeva (2007), have concluded that there is no single morphosyntactic definition or single semantic function associated with finiteness.¹ As such, works like Wurmbrand et al. (2020) claim that different morphosyntactic categories are responsible for finiteness in different languages—such as agreement in the South Slavic languages.

Although I agree with this conclusion, I will argue that there is a single syntactic property that nonfinite clauses crosslinguistically have in common, providing further evidence for Pesetsky (2021)’s claim that finiteness is a matter of clause size. In this paper, I will propose that there is in fact at least one specific clausal projection which all nonfinite clauses lack. In particular, I would like to bring the attention of the reader to a seemingly trivial fact: an infinitival clause can never co-occur with *that*, which is often referred to as a finite complementizer:

- (6) Caitlin seems (*that) to be pretty.

I will argue that (6) is true of all nonfinite clauses. Such an observation, at this stage, is plainly circular: it is trivially true that a finite complementizer cannot head a nonfinite clause. But what is a finite complementizer? Why is *that* associated only with finite embedded clauses? We can answer these questions if we adopt works which split up the CP domain following Rizzi (1997), we can change our conception of what *that* actually is. And this will allow us to bypass this circularity and make a non-trivial crosslinguistic generalization.

Following Rizzi (1997), I split up the C domain in a manner which is schematized below. Further details will be provided in section 3 of this paper, but I will first note that I have eliminated Rizzi’s labels of ForceP and FinP, and replaced them simply with CP2 and CP1. As we will see, this splitting-up is justified by the possibility of double complementizer constructions crosslinguistically, and the existence of complementizers which seem higher and lower in the C domain:

- (7) **CP2 (high)** > IntP > FocP > TopP > WhP > **CP1 (low)** > PropP > TP

Villa-Garcia (2012) provides an illustrative example from Spanish, where there are two complementizers *que*, and the topic precedes one but follows the other:

- (8) Susi dice **que** a los alumnos (**que**) les van a dar regalos
 Susi says that DAT the students that cl. go to give presents
 ‘Susi says that they are going to give the students presents.’ Spanish

I define a *high complementizer* as a complementizer that heads CP2. It precedes topics and focus-marked elements. Thus, the notion of CP2 can be used to define finiteness in a non-circular manner. I argue that high complementizers never appear with nonfinite clauses. A complementizer that heads CP1, on the other hand, is a *low complementizer*. It often appears with nonfinite clauses, but it need not. It follows topics and focus-marked elements, but only if the TopP and FocP layers have not already been truncated, which is almost always the case.

¹Many morphosyntactic categories have been suggested to be responsible for finiteness in the literature: mood, tense, aspect, person marking, illocutionary force, nominal morphology on the verb, and markings that mark dependent clauses in certain languages. Given that a full discussion of these properties would take us out of the scope of the paper, the reader is referred to Nikolaeva (2007) for further discussion.

It is in fact possible to distinguish between these complementizers even in English: I will uncontroversially claim that *that* is a high complementizer. *For* may be a low complementizer. Although many such tests will be presented throughout this paper, I will provide a simple illustrative example. For example, notice that, as Haegeman (2012) points out, topicalization is possible in the embedded clause complements of non-factives, and in this case *that* precedes the topic:

- (9) I said that Manufacturing Consent_i, Chomsky wrote t_i.

This indicates that *that* is a high complementizer in Rizzi's system. On the other hand, infinitives in English never allow topicalization or focalization. For authors who follow Rizzi's framework, this has been taken to indicate that English infinitives seem to be deeply truncated compared to English finite embedded clauses.

- (10) * Chomsky claimed Manufacturing Consent_i, to have written.

The lack of topicalization and focalization in infinitives is by no means a universal, although they appear to be rarely attested. For example, Hebrew infinitives seem to display almost the entire range of the properties of the C domain, allowing *why*-embedding, topicalization, focalization and more, according to Shlonsky (2014):

- (11) ani roce [et ugat ha pereg]_i lenasot t_i.
 I want DOM cake the poppyseed to.try
 'I want to try the poppyseed cake.' Hebrew

And yet, Hebrew infinitives crucially cannot be headed by the high complementizer *še*:

- (12) ani roce (*še) lenasot et ugat ha tapuxim.
 I want (*that) to.try DOM cake the apples
 'I want to try the apple cake.' Hebrew

More revealingly, there are languages—at least Icelandic, Mandarin, Serbian and Spanish—which have complementizers that behave as high complementizers in finite clauses in fact cannot behave as a high complementizer in nonfinite complements.

For example, what has been called the infinitival marker in Icelandic, *að*, appears only with control complements. But another element, *áð*, behaves like *that*, appearing with finite embedded clauses. They have different properties: the former does not allow topicalization at all as seen in (13a), while the latter allows it, following *að*, as in (13b). To account for this phonetic identity, I will propose that *að* is the phonetic form when either CP2 or CP1 is filled in the clausal domain.

- (13) a. * Risarnir lofa [að [á morgun]_i éta ríkisstjórnina t_i].
 the-giants promise to to-morrow eat the-government
 'The giants promised to eat the government tomorrow.'
 b. Risarnir segja [að [á morgun]_i éti þeir ríkisstjórnina t_i].
 the-giants say that tomorrow eat they the-government
 'The giants said that they will eat the government tomorrow.' Thraínsson (1993)

Therefore, the main crosslinguistic generalization that I will argue in this paper, is in (14) below.²

²The scope of this paper is to cover only the clausal size of infinitives; as such, I will only focus only on infinitives. For the most part, I will not discuss gerunds or other kinds of nonfinite clauses in this paper, leaving it to future research. However, it is likely that conclusions that I make concerning infinitives can also be made concerning gerunds as well.

- (14) **Infinitive Size Generalization (ISG):** No infinitive projects CP2.
No infinitive can co-occur with a high complementizer.

This allows for a definition of finiteness in terms of the truncation of the C domain, and hence clause size. This paper is an investigation on the clause size of infinitives more generally, beyond the generalization made in (14) above.

As such, I will also argue that the clause sizes Pesetsky (2021) proposes for infinitives misses many empirical generalizations concerning the size of infinitives, indicating that infinitives can be larger than what Pesetsky proposes. I show that there are at least eight sizes that are attested crosslinguistically. I will also conclude that all control complements are truncated in the C domain at least to some degree, and hence, finite control does not actually exist, contra Landau (2004) and others. This paper presents and discusses data from several different languages belonging to many language families to make these generalizations.

This paper is structured as follows. Section 2 presents Pesetsky (2021)'s derivational theory of finiteness, dubbed Exfoliation, which posits a single size for all infinitives. Section 3 introduces the reader to Rizzi (1997)'s structure for the C domain. Section 4 shows that infinitive size can come in six different maximal sizes crosslinguistically and I discuss the generalizations that result from it. Section 5 discusses the consequences of my theory of finiteness: consequences on Exfoliation, whether factives are truncated, and whether finite control exists. Section 6 concludes.

2 Exfoliation: towards a derivational theory of clause size

Given that I will argue that finiteness is a matter of clause size, this first background section is dedicated to introducing the reader to the recent derivational theory of finiteness in Pesetsky (2021). Although I depart from Pesetsky in some specific areas that I discuss in section 5, I am largely in agreement with his attempt. This section will provide the necessary background to understand the truncation theory of finiteness that I create and defend in this paper.

The idea that finiteness is a matter of clause size far outdates Pesetsky's work. Bouchard (1984), Koster (1984) and Hornstein and Lightfoot (1987) all argue that the extraction of an object correlates with the size of the embedded clause; more recently, Müller (2020) has proposed a similar theory to Pesetsky's. But the attempt by Pesetsky is the most well-developed. As he notes, there is a great variety of clause types found in the languages of the world. Here are some examples from English, in which the embedded clauses are italicized:

- (15) a. I think *that Caitlin mixed hot sauce into my salad.* *finite*
 b. I prefer *for Caitlin to put hot sauce in my salad.* *infinitive*
 c. I suggest *that Caitlin put hot sauce in my salad.* *subjunctive*
 d. I remember *Caitlin putting hot sauce in my salad.* *gerund*

At least in English, raising in English is only possible from infinitives:

- (16) a. Caitlin seems ~~Caitlin~~ to have solved the problem. *infinitive*
 b. *Caitlin seems that ~~Caitlin~~ has solved the problem. *finite*

The core questions that Exfoliation seeks to address are: why do nonfinite clauses exist in the first place, and why do the properties of the subject position in nonfinite clauses differ from their

finite counterparts? One central puzzle to consider arises with raising-to-object/ECM constructions: it has often been considered, since Vergnaud's letter to Chomsky and Lasnik, that the driving factor for raising-to-object constructions is Case assignment, and all nouns need Case.

Let us first discuss Case assignment. Vergnaud, more generally, notes that the distribution of nominals is restricted in a way that CPs/PPs are not:

- (17) We are sure [_{CP} that the world is round] vs. * [_{DP} the world's roundness].

Under Case-driven accounts of raising-to-object constructions, the subject of the nonfinite clause in (18a) is not able to get Case in its base-generated position, so it needs to move up, perhaps to Spec,VP of the matrix verb. There, it is assigned accusative Case. A similar line of reasoning drives the assigning of nominative Case to the matrix subject in raising-to-subject constructions in (18b). In (18c)-(18f), we see that elements which cannot assign Case lead to unacceptability:

- (18) a. Caitlin believes him_i [_{t_i} to be smart]. *raising-to-object*
 b. Caitlin_i seems [_{t_i} to be smart]. *raising-to-subject*
 c. * It seems Caitlin to have solved the problem. *unaccusative matrix verb*
 d. * It was believed Caitlin to speak Irish well. *passive matrix verb*
 e. * Caitlin is aware Madeline to be the cutest. *adjective*
 f. * Caitlin's belief it to have been raining. *noun*

But this makes an incorrect prediction. If elements that don't need Case, like CPs and other elements Pesetsky discusses, we would predict certain structures like the ones below to be grammatical. We obtain the same contrast regardless of their inability to be assigned Case:

- (19) a. Caitlin considers [that the world is round] to be a tragedy. *raising-to-object*
 b. [That the world is round] seems to be a tragedy. *raising-to-subject*
 c. * It seems [that the world is round] to be a tragedy. *unaccusative matrix verb*
 d. * It was believed [that the world is round] to be a tragedy. *passive matrix verb*
 e. * Caitlin is aware [that the world is round] to be a tragedy. *adjective*
 f. * Caitlin's belief [that the world is round] to have been raining. *noun*

The Case approach to this puzzle is on the wrong track. Under a derivational theory of clause size like Exfoliation, these ungrammatical examples do not follow from Case. Under Exfoliation, all clauses are born finite and are reduced in structure to nonfinite via a process of subject extraction. While raising-to-object and -subject constructions allow (18a) and (18b) because they involve subject extraction, (18c)-(18f) are ruled out because they involve illegal in finivization, or subject extraction: these constructions simply do not have a subject extraction probe.

Let us now get into the technical details of this account.³ Pesetsky makes a very strong claim: ultimately, all nonfinite clauses are created via a process of subject extraction, even control constructions which do not prima facie involve subject extraction, putting aside movement theories of control like Hornstein (1999)'s. All clauses are born as full and finite CPs. Infinitives are

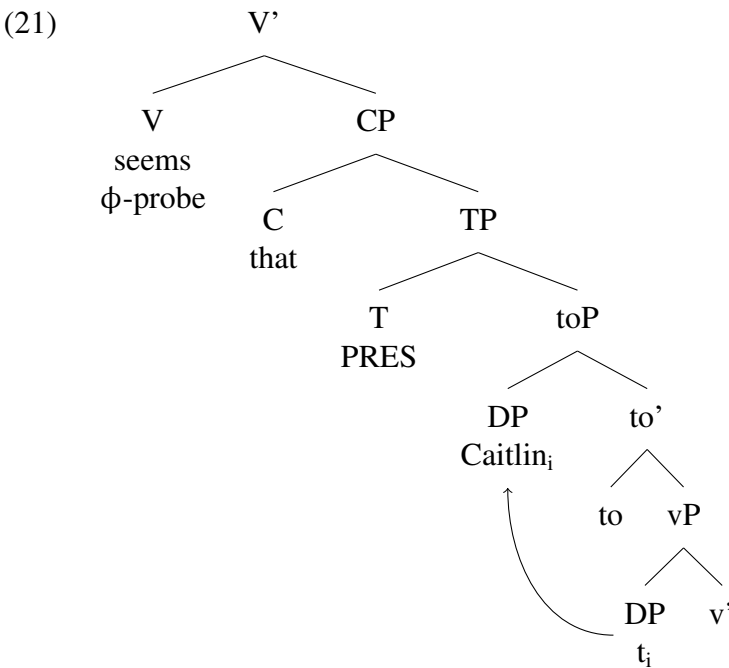
³Pesetsky provides further arguments for Exfoliation, including providing examples beyond the CPs that I have discussed and another argument from unraised nominatives in Icelandic. But for our purposes, this is sufficient, and the reader is referred to Pesetsky (2021) for further discussion.

made, not born, contra selectional accounts in which different predicates, like raising and control predicates, picked the size of their complement.

Under Exfoliation, subject extraction always drives the formation of an infinitive. More specifically, both raising-to-subject and -object constructions involve movement of the embedded subject. This is what drives the formation of the infinitive: a probe has to be able to locate its goal, even across a phase boundary. To get the desired results, some probing across phasal boundaries is required; this is defined as follows:

- (20) a. **Phase Penetrability:** A probe P with an EPP property can locate a goal G across a CP boundary even with G does not occupy the edge of that CP.
 b. **Phase Impenetrability:** But G can move to P only if it occupies the edge of its clause.

Let's see how a derivation of the sentence *Caitlin seems to be happy* would work. First, it is assumed that the embedded clause is born finite, so the embedded clause might look like *seems that Caitlin is happy* at a point in the derivation, as shown in the tree below. Further, all clauses are born with a toP, the relevance of which will be discussed shortly: it can only be pronounced post-Exfoliation. A crucial assumption in the tree below is that the embedded subject does not move to TP immediately; the EPP need not be satisfied immediately:

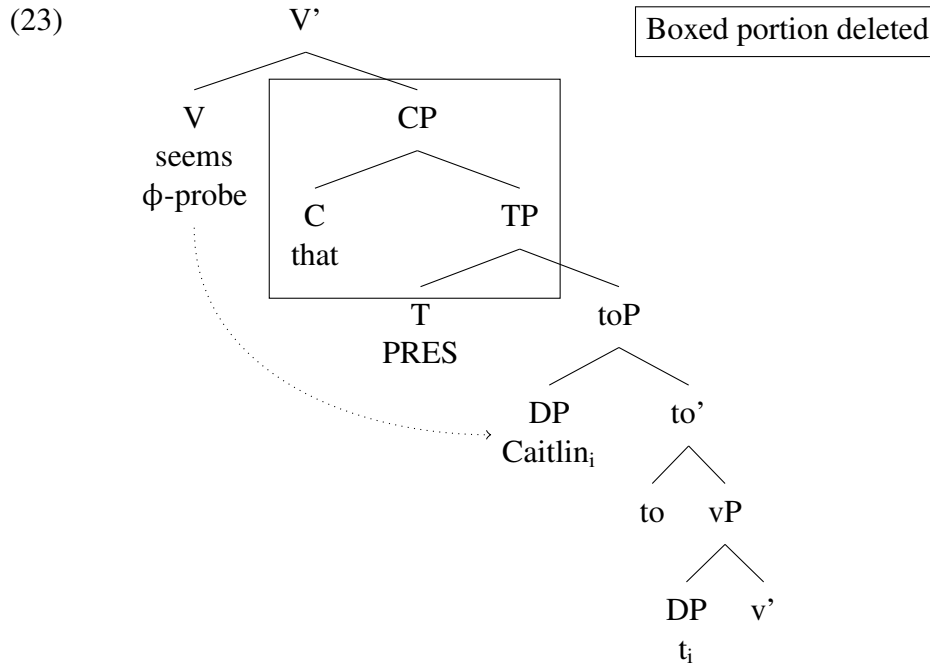


At this point, by Phase Impenetrability above, the embedded subject cannot be extracted because it is not at the edge of the clause. So, the operation Exfoliation comes into play, defined as follows, to ensure that the subject is at the edge:

- (22) **Exfoliation:**
 a. Structural Description: ... A ... [XP (phase) ... [YP (non-phasal) ... B ...]], where:
 i. XP is the phase that dominates B but not A,
 ii. B occupies the edge of YP, and

- iii. a movement triggering probe on A has located B as its goal.
- b. Structural Change: Replace XP with YP, which takes the phasal property of its predecessor.

Exfoliation removes structure to ensure that the embedded subject is now at the phase edge, and the probe on *V* may now extract the subject. Exfoliation removes the CP and TP layers:



The projection toP is present in all finite clauses, as well. Though it is present, to ensure that *to* is pronounced only with infinitives, Pesetsky adds a further condition—dubbed the Exposure Condition—on how certain elements can be pronounced if they head a phase:

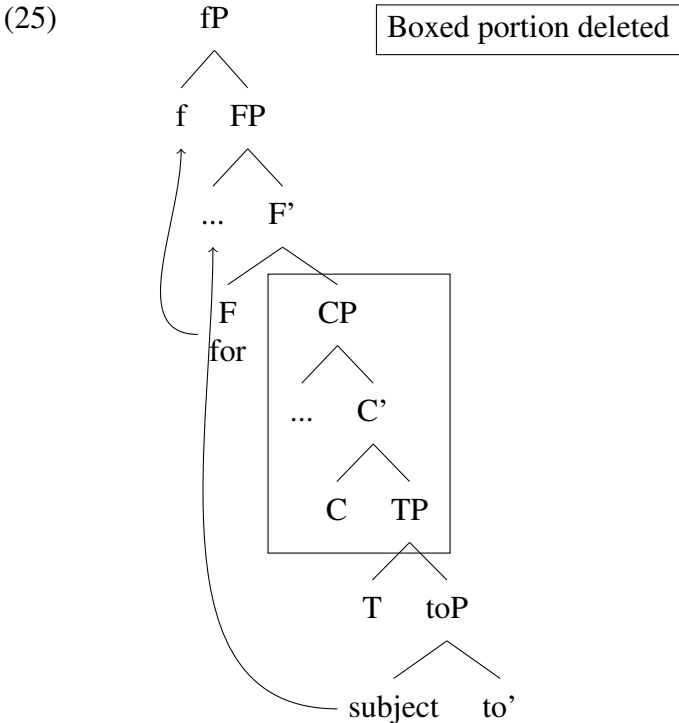
(24) **The Exposure Condition**

- a. A is *exposed* iff it heads a phase and does not retain a specifier. (In other words, if it is the highest element in its phase.)
- b. A functional head is overt iff it is exposed.

It's easy to see how derivation would apply to raising-to-subject and -object constructions. But under Exfoliation, sentences with *for*-infinitives like *Mary is eager for Caitlin to discuss the topic* involves subject extraction, as well. This seems *prima facie* counterintuitive given that *for* only occurs with infinitives to begin with: if infinitives are made and not born, how would *for* even come into play during a derivation? The answer is simple: *for*-infinitives have a similar syntax with raising-to-object constructions.

I will now discuss what I find to be the most controversial notion in this framework: the notion of a *superstructure*. *For* is not a complementizer, but rather an irrealis element that takes a CP as its complement.⁴ This irrealis element is contained in a superstructure that Exfoliates and allows the embedded subject to raise to a position at which *for* can assign it with accusative Case. A simplified illustration of a derivation of a *for*-infinitive is provided below:

⁴The reader is referred to Pesetsky (2021) for empirical evidence for this claim, which I will not be presenting in this paper. Under my account, *for* is a low complementizer in English.



To get a structure for control infinitives, we have two options. First, we can either assume Hornstein (1999)'s movement theory of control, which would have a derivation identical to that of (23), involving subject extraction in a very natural way. But if we don't assume Hornstein's theory, the subject extraction is not obvious. In that case, the derivation of a control infinitive would require a superstructure and an invisible *for*, as in (25).⁵

Before concluding, let me point out that Exfoliation *predicts* the ISG in (14) above, because it entails the deletion of *that*, or the topmost CP₂ layer. If the ISG is true, this is a very strong argument in favor of Exfoliation. But in section 4, I will empirically show that there are at least six different maximal sizes for infinitives crosslinguistically. Pesetsky would have to assume that superstructures themselves have a left periphery, in the style of Rizzi (1997), and they can come in different sizes in different languages. This ultimately means that some selectional aspect is necessary: the entire clause size of the infinitive cannot be derived via Exfoliation.

Putting aside superstructures, we've seen that under Exfoliation, infinitives all come in the same size: toP, which is smaller than CP and TP but larger than vP. This is at odds with Wurmbrand and Lohninger (2019)'s (W&L) recent work which, in my view, conclusively show that infinitives can also come in different sizes. W&L provide empirical data that control complements can in fact have CP and TP layers. They propose that there are three kinds of control complements: propositional, which are CPs; situational, which are TPs; and events, which are vPs.

Propositional complements involve those which can be assigned a truth value, ex. *ESA claimed life to be on Venus, which seems true*. But situational ones cannot, ex. **Mary asked me to buy an apple, which is true*. One empirical test that they provide is given below; propositional infinitives behave like finite clauses in that they cannot occur in the non-progressive form when

⁵Pesetsky assumes further conditions on the pronunciation of *for* and PRO that we need not get into. However, for Pesetsky, PRO is no different than pro. But this is at odds with Landau (2015) and Pearson (2015)'s conclusion, among others, that PRO is a bound minimal pronoun. I leave solving this confound open for future research.

referring to a non-generic episodic event, but situational infinitives can:

- (26) Clara decided to eat salad right now.
(27) Clara claimed to be eating/*eat salad right now.

Under Exfoliation, it is not straightforward to capture such contrasts, given that all infinitives—putting aside superstructures—are only as large as toP. But the most problematic issue is that the Exposure Condition cannot be used together with the arguments that infinitives can come in three different sizes. This would mean that a great deal of the framework would have to be altered.⁶

Before concluding this section, I will note that this paper has much in common with W&L. We both show that infinitives can come in different sizes. For W&L, the maximal size for infinitives crosslinguistically is CP, but this is without splitting the C domain. Once we do so, we observe that infinitives can come in at least **eight** different sizes across languages: CP2 > IntP > FocP > TopP > WhP > CP1 > PropP > TP > vP.⁷ Let us now split up the C domain.

3 Splitting up the C domain

This section will lay the foundation for the theory of finiteness that I propose in this paper: namely that finiteness itself is a property of the C domain. I present Rizzi (1997)'s arguments in favor of splitting up the C domain into many (and potentially ordered, crosslinguistically) different functional projections. I provide evidence for there being high and low complementizers—two separate complementizers—in the C domain. I discuss existing accounts of the truncation of infinitives. At the end, I also provide my update to Rizzi's structure, changing the labels of Rizzi's ForceP and FinP. Rather than having FinP, the low complementizer head determine the finiteness of the clause, I argue that finiteness itself can be derived via truncation, assuming that finiteness is a matter of clause size, as suggested by the Exfoliation framework.

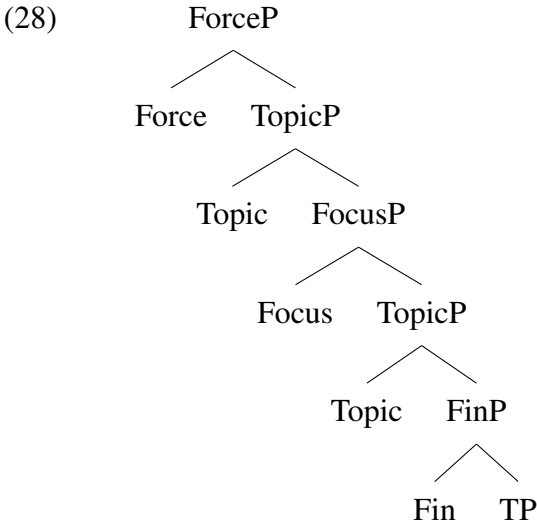
3.1 Rizzi (1997)'s split-CP structure

Rizzi (1997) provides arguments for splitting up the C domain as follows in (28). If we had just one C projection—CP, as is commonly assumed—it would be impossible for a single projection to be responsible for all of these properties that I will discuss in this section.⁸

⁶See Satk (2020) for an attempt to eliminate the Exposure Condition under the Exfoliation framework while getting W&L's empirical observations. Also see Pesetsky (2021)'s Principle of Unambitious Reverse Engineering for a recent attempt at accounting for some of W&L's observations under an Exfoliation framework.

⁷See section 4.1 for the possibility that there might be eight sizes for infinitives crosslinguistically. Although Keine (2020) treats Hindi and German infinitives, there is reason to believe that they are slightly larger than TP but smaller than CP1, given that German allows propositional infinitives but no infinitival complementizers.

⁸This raises the interesting question of what exactly is a phase in this structure. This is at odds with Chomsky (2001) to some degree given that there are many potential phase candidates but it is not obvious which one is the phase head. At the very least, I assume that ForceP—the highest projection of the C domain—is a phase head. Given that wh-movement takes place to a position right above FinP, as I will argue later in this next section, and successive cyclic wh-movement, it might be assumed that FinP is a phase as well. But this is at odds with Carstens and Diercks (2009)'s observations of FinP never being phasal in Lubukusu. Regardless, apart from the phasehood of ForceP, it is out of the scope of this paper to determine what potential phase heads in this structure are.



Rizzi (1997) argues that two complementizers in Italian, *che* and *di*, are realized by Force and Fin respectively. ForceP is the locus of the semantic force of the clause (such as an assertion, a question or an imperative). FinP, on the other hand, simply encodes whether the clause is finite or not. Under Rizzi’s account but not mine, finiteness is to be understood as a very rudimentary specification of mood, tense and agreement in the IP domain. Fin itself does not have a semantics but it is endowed with certain features that allow this aforementioned specification to take place.

Topic and Focus, on the other hand, are projections with an independent semantics of their own, and their specifier position is for topicalized and focalized DPs respectively. There is a difference between focalization and topicalization: they can be teased apart by using different contexts. For our purposes, it is not necessary to discuss this in too much detail, but let us follow Swart and de Hoop (2000) in assuming the following contrast: topic is on expected and uninformative (given) information, while focus is on unexpected (new) information. Focus may also be used contrastively—in fact, Rizzi reports that focus fronting is only available with contrastive focus in Italian. Rizzi contrasts between these two in Italian: while (29a) involves Clitic Left Dislocation (CLLD), (29b) involves focus fronting in a context with contrastive focus:

- (29) a. Il tuo libro, lo ho letto. b. Il tuo libro ho letto.
 the your book, it I have.read the your book I have.read
 ‘Your book, I have read it.’ ‘Your book I have read.’ (but not his)

Furthermore, TopicP in (28) is *recursive*, in that it can appear both before or after FocusP—or before or after other projections between ForceP and FinP; it is commonly assumed that there are. Rizzi provides evidence from this in Italian, which we need not go into; in this paper, I will assume for simplicity that FocusP is always ordered above TopicP.

3.2 What are high and low complementizers?

This sets the stage to allow us to distinguish between *high* and *low* complementizers, which are complementizers realized at Force (my CP2) and Fin (my CP1) respectively. Rizzi was the first to note this contrast, which will be essential for the theory of finiteness in this paper. We see in (30) below that it is impossible to topicalize to a position to the left of the high complementizer *che* (which Rizzi calls a finite complementizer), but it is possible to topicalize to its right.

- (30) a. Credo che, il tuo libro, loro lo apprezzerebbero molto.
I.think that[+fin] the your book them it will.appreciate much
'I think that they will appreciate your book very much.'
- b. *Credo, il tuo libro, che loro lo apprezzerebbero molto. Italian

This contrasts with the behavior of the low complementizer *di* (which Rizzi calls a nonfinite complementizer), which only allows topicalization to its right in (31):

- (31) a. Credo, il tuo libro, di apprezzar-lo molto.
I.think the your book that[-fin] appreciate-it much
'I think that they will appreciate your book very much.'
- b. *Credo di, il tuo libro, apprezzar-lo molto. Italian

This indicates that *di* in Italian cannot be in the same position as *che*: but if *di* is a low complementizer in FinP whereas *che* is a high complementizer in *ForceP*, these facts would immediately be explained. Some languages like Spanish even allow double complementizer constructions:

- (32) Susi dice **que a** los alumnos (**que**) les van a dar regalos
Susi says that DAT the students that cl. go to give presents
'Susi says that they are going to give the students presents.' Spanish

There is a great deal of evidence of high and low complementizers, and even double complementizer constructions even outside of Romance. Even in English, Haegeman (2012) notes two such examples below. Because *that* never behaves as a low complementizer alone, It appears that *that* in FinP can only be licensed if *that* is also realized in *ForceP*:

- (33) a. She maintained **that** when they arrived **that** they would be welcomed.
b. He reminds me **that** in the days of Lloyd George **that** business leaders were frequently buying their way in.

Larsson (2017) provides a survey of double complementizer constructions across the Scandinavian languages, providing an example from Icelandic, from Thráinsson (2007) below. *Sem* is a relative complementizer. The high or low complementizer *að* can follow it. It appears that *sem* is in CP2 while *að* is in CP1 in this case:⁹

- (34) þetta er bokin sem (að) eg keypti
This is book.DEF that that I bought
'This is the book that I bought.' Icelandic

I conclude this subsection with evidence that some Bantu languages distinguish between a high, phasal complementizer and a low, non-phasal complementizer. Carstens and Diercks (2009) shows that in Lubukusu, some clauses are transparent for hyperraising, which is raising out of a finite clause, while others are not transparent for it. Here are some examples from Lubukusu, where what they call hyperraising is possible with the complementizer *mbo*:

⁹Icelandic allows infinitival relatives but they cannot contain *sem*; instead they have the preposition *til*:

- (i) Þetta er bón [til að bóna bíla með _].
this is wax for to polish cars with
'This is wax to polish cars with.'

Hoskuldur Thráinsson (p.c.) has pointed out to me that *til* behaves as a preposition in such constructions rather than a complementizer, based on the fact that the genitive form of *það* 'it,' *þess*, can be inserted between *til* and *að*.

- (35) Mikaeli a-lolekhana **mbo** a-si-kona.
 Michael 1SA-seem that 1SA-PRES-sleep
 ‘Michael seems to still be sleeping.’ Lubukusu

But this raising is not possible with the complementizer *-li* which agrees with the matrix subject:

- (36) * Mikaeli a-lolekhana **a-li** a-si-kona.
 Michael 1SA-seem 1CA-that 1SA-PRES-sleep
 ‘Michael seems to still be sleeping.’ Lubukusu

Under this analysis, *mbo* is the low, non-phasal complementizer, and *-li* is the high, phasal complementizer.¹⁰ We now move onto infinitives.

3.3 Infinitives are truncated in the C domain

Adger (2007) notes a contrast between English and Italian that we will build further upon in section 4.1: topicalization is not allowed at all in English infinitives (Hooper and Thompson (1973)):

- (37) * I decided, [your book]_i, to read t_i.

Adger also notes that the complementizer *for* in English rejects topics. As Adger suggests, I agree with him that this indicates that *for* is a low complementizer in Fin:

- (38) * I propose, [these books]_i, for John to read t_i

Following Adger among others such as Haegeman (2006), Barrie (2007) and Shlonsky and Soare (2011), I also take this to be evidence that infinitives are truncated: as we will see, this truncation can differ between languages like English and Italian.

There is strong reason to believe that there are many more projections than what Rizzi (1997) has initially claimed, and the number of functional projections has indeed increased in works since then such as Haegeman (2012). For our purposes, I will present only the additional projections which are relevant to infinitives–IntP and WhP in particular.

The layer IntP is short for InterrogativeP, which according to Rizzi (2001) is higher than FocusP: Spec,IntP houses *why*. Shlonsky and Soare (2011) provides a convincing argument that *why* is base-generated in position lower than Spec,IntP and moves up to it, in the form of infinitives. Note that the infinitive form is very marginal at best, but the finite form is fine:¹¹

- (39) a. ?? I asked Bill why to serve aubergines.
 b. I asked Bill why I should serve aubergines.

¹⁰As a matter of fact, under my analysis of finiteness, it will turn out that Lubukusu does not have hyperraising at all, because *mbo* is a low complementizer, and all clauses headed by a low complementizer are nonfinite. As such, according to my account, this would in fact be an instance of raising.

¹¹Although it is not relevant for our purposes, Shlonsky and Soare (2011)’s argument that it is base-generated lower is as follows. The following question can be construed in two ways: one in which *why* is construed within the matrix clause, and one in the embedded infinitival clause:

- (i) Why did you ask her to resign?
 a. What is the reason x, such that for x, you asked her to resign?
 b. What is the reason x, such that you asked her to resign for that particular reason x?

Given that we have already seen that TopicP is truncated in English infinitives, it is unsurprising that a functional projection ordered even higher is truncated as well.

Let us move to WhP. The fact that focalization is impossible with English infinitives whereas wh-infinitives in English do exist, ex. *I know what to eat*, is not expected under Rizzi's original account, where all wh-words move to Spec,FocP. As such, Barrie (2007) and Shlonsky and Soare (2011) have assumed the addition of a further functional projection on top of FinP, WhP, which wh-elements first move into prior to moving to Spec,FocP.¹² Even in a language where fronted focus is possible such as Italian, which also has wh-infinitives, Haegeman (2006) and Bocci (2007) note that focalization is very marginal:

- (40) ?? Gli sembra le sedie di aver venduto (, non il tappeto)!
 To him-seems the chairs to have sold (, not the carpet)
 'It seems to him that the chairs have sold! (not the carpet).' Italian

I have shown that infinitives are truncated under a Rizzi-style account of the C domain. If Pesetsky is right in that finiteness is a matter of clause size, then it is difficult to reconcile this with the fact that for Rizzi, finiteness is determined via FinP—because for Rizzi, finiteness is not a matter of clause size. Unlike Rizzi, Pesetsky's derivational account of finiteness is able to make the correct empirical predictions in section 2—whereas Rizzi's account does not make any predictions. Although it is commonly assumed that *that* is realized in ForceP as briefly discussed in section 1, why shouldn't it always be able to be realized at FinP, as long as it is finite?

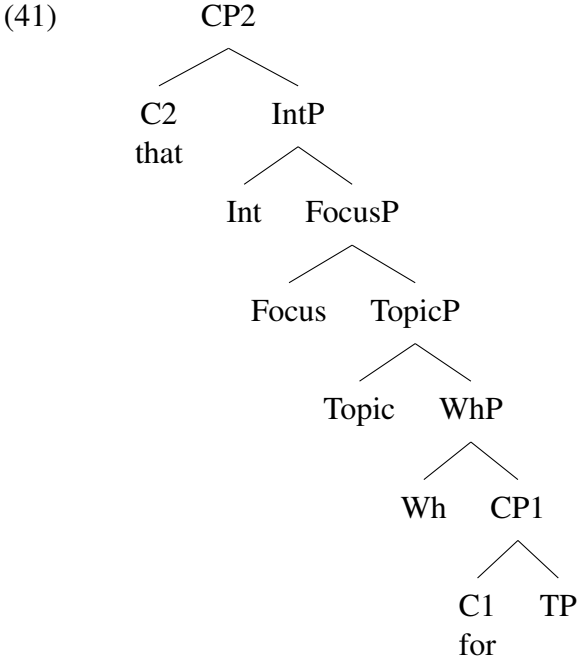
There are more general problems with Rizzi's definition of finiteness, as well. It is circular, in that whether a clause is finite iff its finiteness feature is encoded as + at FinP, following Adger (2007). It may be possible to define finiteness in terms of other features, such as past tense, agreement and indicative mood on FinP. But even then it is circular, because defining FinP itself contains the notion of finiteness. As such, it is not a fully explanatory theory of finiteness.

Furthermore, there is redundancy between those who assume that infinitives are truncated under Rizzi's framework, and the notion of a FinP to begin with. Why do infinitives need to be truncated if finiteness is determined at FinP? What seems more reasonable is that these infinitives are nonfinite *because* they are truncated, and this makes sense if finiteness is a matter of clause size. Thus, I believe that Rizzi's account of finiteness is missing a greater empirical generalization here: namely that *all* infinitives are truncated in some manner. And this what I will argue for in the next section. To start doing so, I propose getting rid of the labels of ForceP and FinP and replacing them simply with CP2 and CP1 respectively:

¹²It seems that there is a WhP on top of ForceP as well. Henry (1995) notes that Belfast English permits indirect questions introduced by a wh-element that isn't a subject, to the left of the high complementizer *that*:

- (i) I wonder which dish that they picked.

This seems to be very common crosslinguistically; Larsson (2017) notes that several Scandinavian languages allow such constructions. At this point, an obvious question to be asking is why there isn't yet another FocusP, TopicP, IntP etc. on top of CP2 as well. But it simply seems to be the case that this is not empirically attested. So this does not put my definition of a high complementizer in jeopardy.



This is what I hope to be the novel idea of the paper. While I am far from the first to assume that infinitives are truncated, I am synthesizing the approach to finiteness as a matter of clause size together with Rizzi’s work on the split C domain. Unlike Rizzi, I am assuming that finiteness is *not* determined by CP1, and I will argue that finiteness is simply determined by whether CP2 is truncated or not. This enables us to eliminate any circularity with Rizzi’s definition of finiteness.

Here is how. The notion of finiteness is not included in the definition of finiteness, eliminating problems with circularity. A high complementizer is defined as the projection that precedes topics and other elements of the C-domain. There is no mention of finiteness in any of the projections of the C domain, not even CP2. It is not logically necessary for high complementizers to have been associated only with finite clauses. We certainly can imagine a language with nonfinite clauses with high complementizers. But for some reason, nonfinite clauses never co-occur with CP2 crosslinguistically. Why so? In the next section, I empirically justify this claim.

4 The size of infinitives

The following hierarchy that was represented in (41) above will be assumed throughout the rest of this section:

(42) CP2 > IntP > FocP > TopP > WhP > CP1 > TP

I present a crosslinguistic survey of infinitive sizes in 4.1. In 4.2, I give reasons to believe from four languages that distinguish between high and low complementizers (or lower clausal heads), very similarly to what Rizzi (1997) noted in Italian above, but these are with elements with the exact same phonetic form. This, I believe, shows a fundamental inability for nonfinite clauses to co-occur with high complementizers. Section 4.3 provides further cartographic generalizations concerning the order in (42). Section 4.4 summarizes the rest of the section.

4.1 Infinitives can differ in size, but are always truncated

We have just seen preliminary evidence that Italian infinitives allow topics while English ones do not. This is the first piece of evidence that infinitives can come in different maximal sizes. In this subsection, I have two goals related to this. First, I provide a survey of 22 languages in which the maximal infinitive size is not CP2—that is, they cannot co-occur with high complementizers. Second, in 19 out of these 22 languages, I also determine the maximal size for infinitives. The final list that we will be left with is as follows in (43). It appears that TP is the *minimal maximal size* for infinitives crosslinguistically, whereas IntP, and crucially not CP2, is the *maximal maximal size* attested. I will argue that this observation is the key to understanding finiteness:

- (43) a. **Hierarchy:** CP2 > IntP > FocP > TopP > WhP > CP1 > PropP > TP
b. **Maximally TP Infinitives:** Turkish, Serbian_{INF}, Hindi, Bangla, Jordanian Arabic
c. **Maximally PropP Infinitives:** German
d. **Maximally CP1 Infinitives:** Icelandic, Swedish, Norwegian
e. **Maximally WhP Infinitives:** English, Spanish, French, European Portuguese, Dutch, Mandarin
f. **Maximally TopP Infinitives:** Italian, Catalan
g. **Maximally IntP Infinitives:** Serbian_{DA}, Hungarian, Hebrew
h. **Maximally CP2 Infinitives:** ∅
i. **Unspecified but not CP2:** Middle English, Old Norse, Old Swedish

Let us see how the empirical tests in this section will work. I now provide a quick summary of the properties of the C domain of English infinitives:¹³

- (44) a. *Infinitival complementizers:* I am eager for Caitlin to please.
b. *Wh-infinitives:* I know what to eat.
c. *No topicalization within infinitives:* *I wanted this book, to read.
d. *No focalization within infinitives:* *I wanted THIS BOOK to read (not that one).
e. *No why-infinitives:* ??I asked Caitlin why to eat salad.
f. *No high complementizer:* I seem (*that) to be happy.

¹³One puzzle is the difference *whether* and *if* in infinitives. These words are often interchangeable, ex. *I asked my mom whether/if I should take out the trash*. But only *whether* is permitted in infinitives:

- (i) a. Caitlin asked whether to take out the trash.
b. *Caitlin asked if to take out the trash.

Following Shlonsky and Soare (2011), one explanation is to suppose that *whether* can be Merged in either Spec,WhP or Spec,IntP, whereas *if* must be Merged in Spec,IntP. It then follows that only *whether* can be licensed in infinitives. It does not appear to be possible to assume that *whether* is always Merged in Spec,WhP, as Jonathan Bobaljik (p.c.) has pointed out to me with the following contrast. Adjuncts which precede the embedded subject must follow *whether* in finite embedded clauses. Adjuncts in the C domain will be discussed further in 5.1.

- (ii) Caitlin asked whether under any circumstances she should leave.
(iii) *Caitlin asked whether under any circumstances to leave.

This indicates that English infinitives are *maximally* as large as WhP. The maximal size of an infinitive is the most crucial notion of this paper: Languages appear to vary as to the maximal size of their infinitive, and there are at least six different maximal sizes which are attested. I will start from the languages that can have infinitives as large as IntP.

4.1.1 IntP Infinitives

Shlonsky (2014) notes that Hebrew infinitives appear to be almost untruncated in the C-domain, allowing focalization and even *why*-infinitives, as shown in (45a)-(45b) below:

- (45) a. ani roce [et ugat ha pereg]_i lenasot t_i (lo et ugat ha tapuxim).
 I want DOM cake the poppyseed to.try (not DOM cake the apples)
 ‘I want to try the poppyseed cake (not the apple cake).’
- b. ani lo mevin lama la’avor dira.
 I not understand why to move apartment
 ‘I don’t understand why to move apartments.’ Hebrew

But there seems to be at least one property which its C domain lacks: the ability to co-occur with the high complementizer *še*. I conclude that Hebrew infinitives may be as large as IntP.¹⁴

- (46) ani roce (*še) lenasot et ugat ha tapuxim.
 I want (*that) to.try DOM cake the apples
 ‘I want to try the apple cake.’ Hebrew

In Hebrew, given the presence of *why*-infinitives, we would predict all of the properties below IntP to be present as well, if there is indeed a cartographic hierarchy. Shlonsky shows almost all of these in Hebrew, apart from the presence of infinitival complementizers in Hebrew. As we will see in section 4.3, Sabel (2006) predicts that they ought to be present in Hebrew as well. This is borne out: according to Landau (2013), Hebrew has the dedicated complementizer *me-*, appearing with control infinitives but not raising ones.

Hebrew is not alone in having infinitives with a nearly complete left periphery. Szécsényi (2009) reports that this is also the case in Hungarian, allowing infinitives with topicalization and focalization. I have verified independently that Hungarian also allows *why*-infinitives below, like Hebrew. Hungarian infinitives do not allow the presence of finite, or high, complementizer *hogy*.

- (47) John meg kérdezte, minek en-ni.
 John VM asked eat-INF
 ‘John asked why to eat.’ Hungarian

Moving onto Serbian, whether its *da*-constructions are nonfinite is controversial. But in section 4.2 I will discuss the differences between Serbian infinitives and *da*-constructions in greater detail, arguing that the latter are in fact nonfinite. I have verified that Serbian *da*-constructions also allow an embedded *why* preceding *da*. If the clause is finite, *why* must follow *da*.

¹⁴In section 5.1, I will give more evidence, apart from CP2, that Hebrew infinitives are truncated. Furthermore, Edit Doron, in an unpublished handout, has claimed that Hebrew infinitives lack semantic force. Given the lack of clarity as to the presence of force within the narrow syntax, I have chosen not to present her arguments. But considering that my CP2 is similar to Rizzi’s ForceP, this is worth pointing out.

- (48) Pitao sam zašto da jedem.
 asked.SG.M AUX.1SG why DA eat.1SG
 ‘I asked why to eat.’ Serbian

On the other hand, I have verified Serbian infinitives appear to lack the C domain entirely, not allowing *wh*-elements inside them, or topicalized elements, or *why*.

- (49) *Ne znam šta jesti.
 NEG eat.1SG what eat.INF
 ‘I don’t know what to eat.’ Serbian

I will hence classify Serbian infinitives as maximally TP.¹⁵

4.1.2 TopP Infinitives

Some languages, such as Italian and Catalan, have infinitives that are larger than WhP but smaller than ones of Hebrew and Hungarian. Although fronted focus in Italian infinitives is borderline unacceptable, repeated in (50a), CLLD is in fact completely acceptable in (50b):

- (50) a. ?? Gli sembra le sedie di aver venduto (, non il tappeto)!
 To him-seems the chairs to have sold (, not the carpet)
 ‘It seems to him that the chairs have sold! (not the carpet).’
 b. Gli sembra, il tappeto, di averlo venduto.
 To him-seems, the carpet, to have-it sold
 ‘It seems to him that the carpet has sold.’ Italian

We have already seen in (31) that Italian has a low complementizer *di*. Given the ordering TopP > WhP, *wh*-infinitives should exist in Italian. According to Kayne (1981), they do, as seen below. I conclude that Italian infinitives can be slightly larger than English ones, or as large as TopP:

- (51) Gli ho detto [dove andare].
 Him I told [where go.INF]
 ‘I told him where to go.’ Italian

Catalan also allows CLLD and *wh*-elements inside infinitives, according to Villalba (2009). But as we will shortly see, most Romance languages—Spanish, French and European Portuguese—actually don’t allow CLLD inside infinitives.

Further, although it appears to no longer be attested, at least in my sample, in the Germanic languages today, Faarlund (2015) and Kalm (2016) arguments can precede both the infinitival marker and the verb in Old Norse and Old Swedish respectively. I provide an illustrative example from Faarlund (2015) below. As in languages like Icelandic, *at* appears to be the phonetic form for both the infinitival marker and finite complementizer in Old Norse and Old Swedish.

- (52) ek hafða nú ætlat [sex skip ór landi]; at hafa t;
 I had now intended six ships from country to have
 ‘I had now intended to take six ships out of the country.’ Old Norse

¹⁵Languages with focalized elements inside infinitives but not *why* does not seem to be attested crosslinguistically. I am unable to answer why this is the case at this time: perhaps the sample in the paper is not large enough, or alternatively, *why* actually involves obligatory movement to FocP rather than IntP, which does not exist.

This leads these authors to reject that *at* in these contexts is a complementizer. But if we treat *at* as a low complementizer inside infinitives rather than a high one, as I will argue in section 4.2, this conclusion will not be necessary. However, given the lack of solid data—such as the possibility of *wh*-infinitives—and impossibility of further investigation, I have not classified the languages together with Italian and Catalan. But given that the infinitives show the inability for *at* to be a high complementizer—that is, obligatorily precede topicalized elements—I have classified Old Norse and Old Swedish as languages without CP2 infinitives.

4.1.3 WhP Infinitives

The most commonly attested maximal infinitive size in my sample is WhP. This is the case in English, as seen in (44) prior, which has the infinitival complementizer *for* and *wh*-infinitives, but not *why*-infinitives, topicalization or focalization in infinitives. Dutch is the only Germanic companion to English among my sample, also allowing *wh*-infinitives (from Wheelock (2015)):

- (53) Ik weet niet [wie te bezoeken].
 I know not [who to visit.INF]
 ‘I do not know who to visit.’ Dutch

This is also the case in other Romance languages with CLLD, in fact: Rizzi (1997) reports that the Italian facts in (50a)-(50b) are not acceptable in French. Villalba (2009) and Barbosa (2001) also report that Spanish and European Portuguese respectively pattern with French, rather than Italian or Catalan, in not allowing CLLD within infinitives.

Although I will save further discussion of finiteness in Mandarin—which is controversial—to section 4.2, Ussery et al. (2016) note that both control and raising complements in Mandarin allow *wh*-elements, while only control complements in English allow *wh*-elements. This leads me to classify Mandarin as a WhP language as well, even though it does not have infinitives, because it appears to have a finite-nonfinite distinction, as mentioned in section 1 above.¹⁶ In (54a)-(54b) below, *how* inside a raising complement is not acceptable in English, but it is in Mandarin:

- (54) a. * Alex seems how to have gotten fat.
 b. Xixi kanqilai zenme zhangpang le.
 Xixi seem how grow fat ASP
 (lit.) ‘Xixi seems how to have gotten fat.’ Mandarin

It appears that, following a derivational approach like Pesetsky (2021), this indicates that raising and control complements have slightly different operations in English, given that the former can only be as large as TP while the latter can be as large as WhP. Though I leave open the details of this solution, this concludes our survey of WhP languages.

4.1.4 TP Infinitives

With the size of English infinitives established, let us move onto TP languages prior to considering CP1 languages. TP, as far as I am aware of, represents the minimal maximal size for infinitives based on my crosslinguistic survey.

¹⁶We will see in section 4.2 that Mandarin does not allow external topicalization to the left periphery in infinitives, like languages like Italian or Hebrew, but merely topicalization to a verb-medial topic or focus position. This is why I have chosen to classify Mandarin as a WhP language rather than, say, a TopP or IntP language.

Keine (2020), based on tests from Wurmbrand (2001) among others, provides convincing arguments that Hindi nonfinite complements are smaller than English infinitives. For example, the *wh*-element *kyaa* ‘what’ can take scope within the finite embedded clause, as in (55a). But it cannot take embedded scope inside the infinitive, as shown in (55b). Keine reports that the sentence is acceptable as long as the *wh*-element takes matrix scope:

- (55) a. tum jaan-te ho [(ki) us-ne kyaa ki-yaa]
 you know-IPFV.M.PL be.PRES.2PL that he-ERG what do-PFV.M.SG
 ‘You know what he did.’
- b. * tumhe [kyaa kar-naa] aa-taa hai
 you.DAT what do-INF.M.SG come-IPFV.M.SG be.PRES.3SG
 ‘(Intended) You know what to do.’ Hindi

But it is not only WhP that is missing. Keine also makes a stronger claim: that the C domain is entirely missing in Hindi infinitives. (56) below involves illicit A’-movement from a doubly embedded finite clause into the embedded infinitive. A’-movement is allowed from finite clauses, so the reason why (56) is ruled out, for Keine, is because the infinitive lacks all A’-positions:

- (56) * [mai caah-taa huu [kitaab-ko kah-naa [ki mai-ne parh-aa
 I want-IPFV.MSG be.PRES.1SG book-ACC say-INF.M.SG that I-ERG read-PFV.M.SG
 hai]]]
 be.PRES.3SG
 ‘(Intended) I want the book, to say that I read.’ Hindi

Like English, Hindi does not allow a high complementizer to co-occur with the infinitive:

- (57) siitaa [(*)ki] prataap-ko dekh-naa] caah-tii thii
 Sita that Pratap-ACC see-INF.M.SG want-IPFV.F.SG be.PST.F.SG
 ‘Sita wanted to see Pratap.’ Hindi

Based on these data, Keine concludes, as I do, that Hindi infinitives are only as large as TP. Dasgupta (1982) reports that Bangla, another Indo-Aryan language, patterns with Hindi in that it lacks *wh*-infinitives and the complementizer *je* cannot co-occur with infinitives. This is despite the fact that according to Hsu (2015), *je* in Bangla can occur in either ForceP (my CP2) or FinP (my CP1), indicating further that it is genuinely a TP language.

Kornfilt (1996) also reports that although Turkish has infinitives, it lacks infinitival *wh*-questions. My own judgment is that infinitival complementizers are not present in the language, either. An illustrative example is given below:

- (58) * Ahmet Ayşe-ye [PRO ne oku-mak] söyle-di.
 Ahmet Ayşe-DAT what read-INF say-PST
 ‘Ahmet told Ayşe what to read.’ Turkish

A particularly interesting case of a language that appears to have a finiteness contrast despite not having infinitives, according to Al-Aqarbeh (2011), is in Jordanian Arabic. Al-Aqarbeh (2011) argues that in Jordanian Arabic, finite complements are those which project a C domain, and non-finite complements are those which do not project a C domain. Two illustrative examples are given below, in which the complement clause cannot have a complementizer or an embedded topicalized element:

- (59) a. 9ali bid-uh (*innu) il-banaat yi-l9ab-an.
 Ali want-3SG.M (*that) the-girls 3-play-PLF
 ‘Ali wants the girls to play.’
- b. *9ali bid-uh il-ghurfah il-banaat yi-naththif-an-ha.
 Ali want-3SG.M the-room the-girls 3-clean-PLF-it
 ‘Ali wants the girls to clean the room.’ Jordanian Arabic

On the other hand, a complementizer and embedded topicalized elements may appear with propositional complement clauses. Although it would be out of the scope of this paper to discuss in detail the semantics of the complements are nonfinite and those which are finite, nonfinite complements cannot have propositional semantics, at least in Arabic. Hence, Al-Aqarbeh (2011) relates the presence of the C domain to finiteness in Jordanian Arabic.

4.1.5 PropP languages

The languages that I have classified as TP languages are those which appear to lack propositional semantics in their infinitives entirely, along with all properties of the C domain, as in Jordanian Arabic. For example, according to Bhatt (2006), Hindi lacks raising predicates like that of English entirely. German appears to be the only language in my sample which appears to lack an infinitival complementizer—following Sabel (2006) among others—but has raising constructions with propositional semantics, like the one below.

- (60) Er scheint [intelligent zu sein].
 he seems intelligent to be
 ‘He seems to be intelligent.’ German

Wh-infinitives are impossible in German, as shown in Wheelock (2015)’s example:

- (61) *Ich weiß nicht [was zu kaufen].
 I know not [what to buy.INF]
 ‘I do not know what to buy.’ German

For Keine, German infinitives are the same size as Hindi’s, but he does not discuss the lack of propositional infinitives in Hindi. I believe that this indicates the need to distinguish between German on one hand, and languages like Hindi on the other. I follow Wurmbrand and Lohninger (2019) in assuming that the propositional semantics is a part of the C domain, but belonging to a head lower than CP1, which I call PropP. Although I am unable to improve on this stipulation at this time, this allows us to straightforwardly capture the difference between German on one hand and Hindi on the other.

4.1.6 CP1 Infinitives

There appears to be a final size in between WhP and PropP/TP, which I have saved for last given that it is likely to be the most controversial. The most difficult maximal size to determine is that of languages like Icelandic, Swedish and Norwegian, because their finite complementizers share the same phonetic form as the so-called infinitival marker. This is unlike that of German: *zu* is not the phonetic form of the finite complementizer *dass*. But I would like to provide an analysis of languages like Icelandic in which this phonetic similarity is not a mere coincidence.

As mentioned in section 1, *að* seems to come in two different varieties: finite clauses allow embedded topicalization to the right of *að*, whereas the *að* found in control infinitives does not allow topicalization. It appears, then, that TopP in Icelandic infinitives is always truncated. Data from Thraínsson (1993) is repeated in (62a)-(62b) below.

- (62) a. *Risarnir lofa [að [á morgun]_i éta ríkisstjórnina t_i].
the-giants promise to to-morrow eat the-government
‘The giants promised to eat the government tomorrow.’
b. Risarnir segja [að [á morgun]_i éti þeir ríkisstjórnina t_i].
the-giants say that tomorrow eat they the-government
‘The giants said that they will eat the government tomorrow.’ Icelandic

Faarlund (2015) points out that argument preposing of this kind is not possible in Norwegian embedded infinitives, either, while it is in Old Norse. Furthermore, Icelandic lacks *wh*-infinitives, as Sabel (2006) points out, indicating the absence of *WhP*.

Thraínsson (1993) took (62a)-(62b) as evidence that *að* in AgrSP–right above TP but below CP. Further evidence of this is the fact that Icelandic has V-to-T (V-to-I in older frameworks) movement in infinitives, unlike English, and the verb still occurs after *að*. I believe the first to note this was Sigurðsson (1989). Notice that in (63a), the movement of the auxiliary to T precludes the movement of the embedded verb to T, but this is not the case in (63b), and it does move to T. In control infinitives, V to T is still possible and it is to the right of *að*, as in (63c).

- (63) a. Risarnir segja [að þeir hafi stundum [_{VP} étið ríkisstjórnir].
the-giants say that they have sometimes eaten governments
‘The giants say that they have sometimes eaten governments.’
b. Risarnir segja [að þeir éti_i stundum [_{VP} t_i ríkisstjórnir]].
the-giants say that they eat sometimes governments
‘The giants say that they sometimes eat governments.’
c. Risarnir lofa [að éta_i oft [_{VP} t_i ríkisstjórnir]].
the-giants promise to eat frequently governments
‘The giants promised to eat governments frequently.’ Icelandic

The evidence that *að* is above TP seems to be strong. Assuming that it is in AgrSP would not contradict anything in this paper. However, with Rizzi’s split-CP structure, we do not need to give up the idea that *að* in Icelandic is *always* a complementizer—it could simply be a low complementizer realized in CP1 if it is not first realized in CP2. This has an advantage over Thraínsson (1993)’s account of infinitival *að* in AgrSP, given that it would be mysterious as to why the two *að* with different properties have the same phonetic form.

Another CP1 language appears to be Swedish. Platzack (1986) notes that the Swedish complementizer *att* is similar to that of Icelandic. We’ve already seen that Swedish lacks *wh*-infinitives; Engdahl (1986) also provides evidence that Swedish infinitives are not full CPs either, based on the inability of pied-piped material to appear in infinitival relative clauses:

- (64) a. ett rum att arbata i _ b. *ett rum i vilket att arbata [_{PP} _]
a room to work in a room in which to work
‘a room to work in.’ Swedish ‘a room in which to work.’

Finally, Wheelock (2015) notes that both Swedish and Norwegian lack *wh*-infinitives. I believe that the data given above imply the need to stipulate a different projection, WhP, above CP.¹⁷

4.1.7 Conclusion

I understand that some of the classifications may be controversial—for example, assuming a separate projection WhP for *wh*-movement and PropP purely for propositional semantics, without a complementizer. But ultimately, none of this contradicts the primary goal of this paper, which is to show that infinitives cannot co-occur with high complementizers.

Regardless, our survey is almost complete. But it has been claimed that Middle English infinitives project ForceP. According to van Gelderen (1998), it is possible for *ai* in (65) below to be a focus marker; in which case, *til* would be in ForceP (my CP2), flatly falsifying my upcoming generalization: no infinitive projects CP2. My attempt at glossing her ideas is below:

- (65) Til [all oure bale] ai for to bete
COMP all our sorrow FOC COMP to heal
'For all our sorrow to heal...' Middle English

However, according to Jay Jasanoff (p.c.), it appears that this is not a double complementizer construction. *Til* plays the role of complementizer *for* in this construction, making it as large as CP1. *Ai* is not a focus marker but rather a word that means *forever*, whereas "for to" in Middle English is itself the infinitive marker, (cf. *to* in English). When this sentence is translated with modern lexical substitutions into its syntactic structure, we obtain *for all our sorrow forever to amend*, which is not so exotic after all.

Infinitives appear to never project the full C domain; in particular, Rizzi's ForceP, or my CP2. I have shown that even in Hebrew and Hungarian, with the largest attested infinitives, infinitives cannot co-occur with the so-called high complementizer *še* in Hebrew or *hogy* in Hungarian.

Of course, one might be allege that this might simply be because finite complementizers don't select nonfinite clauses. But I believe this simply begs the question of *why* finite complementizers (in our terminology, high) do not select nonfinite clauses, and does not lead to a greater understanding of this fact. To explain this, I present a potential finiteness generalization in (66):

- (66) **Infinitive Size Generalization (ISG):** No infinitive projects CP2.
No infinitive can co-occur with a high complementizer.

But we do not yet have enough evidence to conclude that the ISG is true, of course. Absence of evidence is not evidence of absence: the fact that there does not seem to be a language reported in the literature with a high or double complementizer construction—with the exception of Middle English, which we have rejected—does not mean that we have a universal.

In other words, the pattern seen in (43) is not enough to conclude that infinitives are *always* truncated, and that finiteness can be defined in terms of the presence or lack of the CP2 layer. But in the next subsection, I attempt to present evidence of absence in favor of the ISG, in which I argue that nonfinite clauses are fundamentally unable to co-occur with a high complementizer.

¹⁷We might try to avoid not splitting up the C domain into WhP and CP by assuming that C has a *wh*-feature in English but not in, for example, Icelandic. But I believe that this would miss the upcoming cartographic generalizations in section 4.3.

4.2 Languages with the same phonetic form for high complementizers and other clausal heads

This subsection presents further evidence for the generalization in (66) above. We will be investigating a specific pattern in several languages. In particular, all of these languages have an element which is uncontroversially high complementizer, corresponding to *that* in English. This element can also appear in nonfinite clauses, which might seem as a genuine counterexample to the ISG. But this element, it turns out, has very different properties when it heads a clause we would consider nonfinite clause: in other words, it is not a high complementizer in these contexts.

Italian is not one of these languages. But recall the data from (30)-(31) above, repeated in (67) below. It is possible to topicalize to the right of the high complementizer *che* in Italian but not to its left; it is also possible to topicalize to the left of the low complementizer *di* but not to its right:

- (67) a. Credo che, il tuo libro, loro lo apprezzerebbero molto.
I.think that[+fin] the your book them it will.appreciate much
'I think that they will appreciate your book very much.'
- b. *Credo, il tuo libro, che loro lo apprezzerebbero molto.
- c. Credo, il tuo libro, di apprezzar-lo molto.
I.think the your book that[-fin] appreciate-it much
'I think that they will appreciate your book very much.'
- d. *Credo di, il tuo libro, apprezzar-lo molto. Italian

It turns out that similar contrasts are seen crosslinguistically, even with elements that share the same phonetic form. The first of which is, of course, Icelandic, which we have already discussed in subsection 4.1 prior: it allows topicalization to its right in finite contexts as in (62a) above, but not at all in control infinitives, as in (62b). This, in my view, is because *að* cannot behave as a high complementizer in control infinitives, because CP2 is truncated.

Similar facts are seen in Norwegian and Swedish, according to Faarlund (2015) and Kalm (2016) respectively. In Old Norse, we saw in (52) prior that unlike Icelandic, Swedish and Norwegian, argument preposing inside infinitives is allowed. Although it is not clear whether this involves topicalization, it does at the very least show that *at* does not behave as a high complementizer in this context. Here is a similar example from Kalm (2016) illustrating this in Old Swedish:

- (68) þa ær han skyldugher han at ola
then is he obliged him to oil
'Then he is obliged to oil him.' Old Swedish

A language similar to Icelandic in some respects is Spanish, according to Villa-Garcia (2012), for which (69) is repeated below. Villa-Garcia (2012) refers to the first bolded *que* as a high complementizer, just like *that*, whereas the lower *que* he refers to as a "jussive/optative" complementizer, which is characteristic of subjunctives. (69) shows that topicalization occurs to the right of the high variety of *que*.

- (69) Susi dice **que** a los alumnos (**que**) les van a dar regalos
Susi says that DAT the students that cl. go to give presents
'Susi says that they are going to give the students presents.' Spanish

It seems that for independent reasons, the complementizer *que* cannot occur in Spanish infinitives; according to Lujan (1980) a separate complementizer *de* is used instead, so the facts would not be very different from Italian. But there are other nonfinite contexts outside of infinitives in which low *que* can be used, such as imperatives. In (70) below, Demonte and Fernández-Soriano (2009) point out that the topic *a ese alumno* ‘to that student’ moves to the left of *que*. They analyze the two *que* precisely as I and Villa-Garcia do: *que* comes as both a high and a low complementizer, and the low variety is present in nonfinite contexts like (70).

- (70) A ese alumno, que los profesores no lo dejen salir hasta las 6.
to that student that the teachers not CL.3SG allow leave until the 6
‘Let the teachers not allow that student to leave before 6.’ Spanish

We are now moving onto Serbian, for which Wurmbrand et al. (2020) has already provided us with a well-developed analysis of complementation that will lay the foundation for the arguments in this subsection—although I will disagree with their conclusion on what finiteness in Serbian is. Wurmbrand et al. (2020) notes that Serbian allows both “finite” and nonfinite complements of verbs like *try*. We see two forms that can be the complement of *try* in (71a): the bare infinitive form without *da*, and *da* together with agreement on the embedded verb. But the infinitive is impossible with the propositional complement of *claim*, as in (71b) below:

- (71) a. Pokušala sam {da čitam / čitati} ovu knjigu.
tried.SG.F AUX.1SG DA read.1SG / read.INF.IPFV this book
‘I tried to read this book.’
b. Tvrdim {da čitam / *čitati} ovu knjigu.
claim.1SG DA read.1SG / *read.INF.IPFV this book
‘I claimed to be reading this book.’ Serbian

Our objection of investigation is this *da*. For Wurmbrand et al. (2020), assuming the framework of Wurmbrand and Lohninger (2019), the complement of *try* is an event complement, and only as large as a vP–no TP or CP layers. On the other hand, the complement of *claim* is a full CP. Under this account, *da* itself is not a complementizer, but rather a lower clausal head that can mark vPs, TPs or CPs. I will adopt this analysis for Serbian and Mandarin, but not Icelandic or Spanish. Given the presence of verb-medial focus and topic positions in Serbian and Mandarin, but not in Icelandic or Spanish, it is difficult to determine whether focalization or topicalization would take place within the V or C domain, as Jim Huang (p.c.) points out.

What Wurmbrand et al. and I diverge on is the nature of finiteness. For them, finiteness is a language specific property, and it is agreement in Serbian. So, the complement of *try* may be finite. By contrast, I claim that finiteness is in fact not a language specific property, and it is merely the presence of an untruncated C domain. As such, under my account, the complement of *try* is never finite, as it is only as large as vP, but rather something akin to an inflected infinitive.

Todorović and Wurmbrand (2016) notes that tenseless complements of predicates such as *try* and propositional complements of predicates like *claim* allow topicalization and focalization, but with different word order. This is possible given that Serbian has verb-medial topic and focus positions. Topicalization in the embedded complement of *try* must precede *da*, but follow *da* with the complement of *claim*. I present my own illustrative examples below:¹⁸

¹⁸Zeljko Bošković (p.c.) has suggested to me that these examples are marginal with topicalization, but better with

- (72) a. Pokušala sam [ovu knjigu]_i da čitam t_i.
 tried.SG.F AUX.1SG this book DA read.1SG
 ‘I tried to read this book.’
- b. *Pokušala sam da [ovu knjigu]_i čitam t_i.
- c. Tvrdim da [ovu knjigu]_i čitam t_i.
 claim.1SG DA this book read.1SG
 ‘I claimed to be reading this book.’
- d. *Tvrdim [ovu knjigu]_i da čitam t_i. Serbian

This looks like Italian. On one hand, we see *da* behave as a high complementizer in the complement of *claim*, as evidenced by (72d). On the other, *da* must behave as a lower clausal head, as shown in (72b), in which *this book* moves to a verb-medial focus or topic position. Once again, I believe that this is evidence of a fundamental inability of nonfinite clauses to co-occur with high complementizers, which language specific accounts of finiteness do not predict.

Before moving to Mandarin, given that I will claim later in this paper that control is fundamentally a property of clauses which are truncated in the C domain, it would be important to determine whether subjects can be licensed in the complement of *claim* but not *try*. This is precisely what is the case; the complement of *try* requires OC PRO but that of *claim* can license subjects:

- (73) a. Pokušala sam da (*Mari) čitam ovu knjigu.
 tried.SG.F AUX.1SG DA read.1SG this book
 ‘I tried (*for Mary) to read this book.’
- b. Tvrdim da Mari voli John.
 claim.1SG.F DA Mary loves John
 ‘I claimed that Mary loves John.’ Serbian

Mandarin has a similar pattern to Serbian. Huang (2018) makes precisely the same argument that I made for Serbian, but in Mandarin instead—his analysis can be straightforwardly translated to mine. As Huang (2018) convincingly shows, *shuo* behaves as a finite complementizer (in our terminology *high*) when it heads a finite embedded clause. In (74), topicalization is only allowed within the embedded clause, because the complement of *believe* must be finite.

- (74) a. Wo xiangxin [shuo Lisi [zhe-pian baogao]_i xie-wan-le t_i].
 I believe SHUO Lisi this-CL report write-finish-PFV
 ‘I believe that Lisi has written this report.’
- b. *Wo [zhe-pian baogao]_i xiangxin [shuo Lisi xie-wan-le t_i].

But *shuo* behaves as a lower clausal head when it heads a nonfinite embedded clause, such as the complement of *try*, with which the pattern in (74b) is possible. The complement of *try* in (75), which appears to be nonfinite—as evidenced by the requirement of a controlled PRO—involves restructuring, as it allows the embedded object to move up and precede the verb:

- (75) Wo [zhe-pian baogao]_i hui shefa [shuo jinkuai xie-wan t_i].
 I this-CL report will try SHUO as-soon-as-possible write-finish

contrastive focus. Furthermore, I have verified that with a control predicate like *decide* which takes situation complements, the complement allows topicalization both before and after *da*, as predicted by Wurmbrand and Lohninger (2019)’s ICH, which Wurmbrand et al. (2020) assumes and is based on. That predicates like *decide* can take both finite and nonfinite complements is true in English, as well.

‘I will try to finish this report as soon as possible.’

Mandarin

Once again, we see the fundamental inability of a high complementizer to co-occur with nonfinite contexts. The untruncated CP2 layer blocks topicalization to a matrix verb-medial topic or focus position, as in (74b). But restructuring, and removal of the CP2 layer, allows for this movement to take place, as in (75).

Concerning subject licensing, the complement of *like*—a predicate that takes vP complements similar to *try*—requires an OC PRO but that of *hope* does not, which according to Grano (2017) takes a CP, as predicted:

- (76) a. Xiaoming_i xihuan (*ta_{i/j}) chi shousi. b. Xiaoming_i xiwang (ta_j) chi shousi.
Xiaoming like he eat sushi Xiaoming hope he eat sushi
‘Xiaoming likes to eat sushi.’ ‘Xiaoming hopes to eat sushi.’

This section, in my view, shows that complementizers, when put into nonfinite clauses, cease to behave as high complementizers: depending on the language they must either behave as low complementizers or as lower clausal heads. This is further evidence that the CP2/ForceP layer of nonfinite clauses is truncated.

4.3 Cartographic predictions

Recall the order of the projections of the C domain from Rizzi, and the following empirical pattern from 4.1. If this ordering is correct, we would be able to make further cartographic predictions on the nature of infinitives crosslinguistically—although this is not the primary goal of my paper, it appears to be an interesting corollary. For example, we would expect WhP languages to have infinitival complementizers, TopP languages to have wh-infinitives and infinitival complementizers, and IntP languages to have all of that.

- (77) a. **Hierarchy:** CP2 > IntP > FocP > TopP > WhP > CP1 > PropP > TP
b. **Maximally TP Infinitives:** Turkish, Serbian_{INF}, Hindi, Bangla, Jordanian Arabic
c. **Maximally PropP Infinitives:** German
d. **Maximally CP1 Infinitives:** Icelandic, Swedish, Norwegian
e. **Maximally WhP Infinitives:** English, Spanish, French, European Portuguese, Dutch, Mandarin
f. **Maximally TopP Infinitives:** Italian, Catalan
g. **Maximally IntP Infinitives:** Serbian_{DA}, Hungarian, Hebrew
h. **Maximally CP2 Infinitives:** ∅
i. **Unspecified but not CP2:** Middle English, Old Norse, Old Swedish

In fact, Sabel (2006) was the first to do a survey of infinitives, concluding that if a language has wh-infinitives, then it also has infinitival complementizers. This is evidence for the ordering WhP > CP1 under a Rizzi framework, though Sabel does not assume it. For Sabel, wh-movement simply takes place to Spec,CP, so the presence of wh-movement necessitates the presence of a C head, but not vice versa. The presence of an infinitival complementizer does not mean wh-movement is possible. But the Rizzi framework might allow for us to build on Sabel’s work.

For example, if a language allows topicalization, such as Italian and Hebrew, then we would also predict that it has wh-infinitives and infinitival complementizers. This is already borne out in

Italian and Catalan according to Sabel and Villalba (2009) respectively. Furthermore, we would predict that Hungarian and Hebrew infinitives should allow topics, focalized elements, and *wh*-elements given that they allow *why*-infinitives. As we saw, this prediction was also borne out. Though this is a good starting point, given the limited size of my survey, more evidence would be helpful in order to determine whether these predictions.

Building on this, I provide a survey of *tough*-constructions crosslinguistically. I will argue the pattern in (77) is tightly connected to their distribution: in particular, I will show that what we call *tough*-constructions in TP languages like German, CP1 languages like Swedish and WhP languages like English all have different properties.

But first, let us discuss Chomsky (1977b)'s arguments in favor of *tough*-movement involving a step of *wh*-movement. Here is an example of such a construction from English:

- (78) a. It is easy to play sonatas on the violin. (without *tough*-movement)
 b. The violin is easy to play sonatas on. (with *tough*-movement)

Whether inside the C-domain of the infinitive in (79a)-(79b) blocks the embedded *where* from moving to the matrix Spec,CP position:¹⁹

- (79) a. I am wondering whether to eat lunch at Chipotle.
 b. *Where am I wondering whether to eat lunch?

Similarly, extraction out of the infinitive yields this same kind of ungrammaticality, as seen in (80a)-(80d). The middle Spec,CP position was occupied by a Copy of *what sonatas* prevents *this violin* from moving up in (80d).

- (80) a. It is easy to play these sonatas on this violin.
 b. These sonatas are easy to play on this violin.
 c. What sonatas are easy to play on this violin?
 d. *What sonatas is this violin easy to play on?

Under a more modern understanding of the C domain, *tough*-movement takes place to Spec,WhP in English infinitives. But according to Chomsky, it is not because the embedded object moves to Spec,WhP, an A'-position, and then to matrix Spec,TP, as this would be a violation of Chomsky (1977a)'s Improper Movement constraint. Instead, the embedded object is a null operator that moves to Spec,WhP while the coreferring matrix subject is base-generated:

- (81) Caitlin_i is [_{WhP} Op_i [_{TP} PRO_{arb} tough to please t_i.]]

What about TP languages? We have seen that the maximally TP-infinitive languages do not allow *wh*-infinitives at all, so they should not have *tough*-movement. And yet, according to Comrie (1997) among others, German, a maximally TP language might *prima facie* appear to have *tough*-constructions, along with both of the CP1 languages.²⁰

¹⁹This position for *whether* is justified as follows. As this example shows, *whether*-infinitives exist, so they must be in either WhP or CP1. On the other hand, the element *if* is commonly assumed to be in Spec,IntP as Shlonsky and Soare (2011) claim. We predict correctly that infinitives like **I know if to eat salad* are ruled out in English because *if* occurs in a higher position than WhP.

²⁰Stefan Keine (p.c.) has pointed out to me that Hindi does not have *tough*-constructions, which would be expected given that its infinitives do not have a CP layer.

This is contradictory, given Chomsky (1977b)'s observation that *tough*-movement involves *wh*-movement. How is this possible if maximally TP languages lack a WhP layer? I propose that in fact, the maximally TP languages do not have *tough*-movement after all, allowing us to make significant empirical generalizations concerning languages which do have *tough*-movement, which has consequences on the Exfoliation framework.

Wurmbrand (1994) argues that German does not in fact have *tough*-constructions because it has different properties from *tough*-constructions that we see in English. Out of four of her tests, I will include two. For example, they do not allow arguments intervening between the embedded object and matrix subject (82a) and do not license parasitic gaps (82b):

- (82) a. * Dieses Buch ist schwer Hans zu überzeugen zu lesen.
 this book is hard John to convince to read
 ‘This book is hard to convince John to read.’ German
- b. * weil das Buch_i [ohne vorher *pg*_i zu kaufen] schwer *t*_i zu lesen ist
 because the book [without before to buy] hard to read is
 (Intended?) ‘Because the book is hard to read without having bought beforehand.’

Following Wurmbrand, I propose that we call this kind of long A-movement in German *leicht*-movement, with the resultant construction a *leicht*-construction. By contrast, genuine *tough*-movement involves a step of A'-movement to Spec,WhP prior to A-movement to the matrix subject position, as Chomsky proposes.

What about CPI languages like Swedish and Icelandic? Surprisingly, according to Klingvall (2018), Swedish *tough*-movement patterns somewhere in between English and German. Klingvall argues that there is a step of A'-movement in Swedish *tough*-constructions and they don't just involve long A-movement like in German. For example, they pattern with English rather than German in licensing parasitic gaps (83a) and are not sensitive to arguments intervening between the embedded object and matrix subject (83b).²¹

- (83) a. [Den artikel-n]_i är svår att övertala Lisa att be Johanna att läsa *t*_i.
 that paper-CMN.DEF is hard.CMN to convince Lisa to ask Johanna to read
 ‘That paper is hard to convince Lisa to ask Johanna to read.’
- b. Bok-en_i är lätt att kritisera *t*_i utan att ha läst *pg*_i
 book-CMN.DEF is easy to criticize without to have read
 ‘The book is easy to criticize without having read.’ Swedish

But Swedish infinitives are unlike English ones. Klingvall notes the data we have above concerning the lack of *wh*-infinitives and pied-piped material in infinitival relative clauses, indicating the lack of a full C domain. This indicates that WhP is not present in Swedish, the lack of which seems to lead to further differences. For example, Klingvall notes that although English TCs do not allow a pronoun inside a subject to be bound by something inside the embedded clause, Swedish ones do (though this is subject to dialectical variation):

- (84) a. * [The bad news about her_i goat]_k was hard for John to tell every farmer_i *t*_k.

²¹For space reasons I've trimmed Klingvall's example. Also, Klingvall distinguishes between verbal TCs and adjectival TCs, to be more specific, but this distinction is immaterial for this paper.

- b. % Nog var [sin_i (rättmätiga) lön]_k svår (för oss) att ge varje anställd_i
 surely was REFL rightful salary difficult.CMN for us to give every employee
 t_k igår eftermiddag.
 yesterday afternoon
 ‘His/her rightful salary was surely difficult (for us) to give every employee yester-
 day afternoon.’

Klingvall suggests that although Chomsky’s original approach is right for English, it is not in Swedish. In Swedish, the null operator moves to an A’-position in the T domain—not Spec,TP, which is an A-position, but higher than that. This explains why *tough*-constructions in Swedish pattern with English.

This allows us to make the following generalization concerning languages with genuine *tough*-movement—movement of a null operator to the C domain—like that of English:²²

- (85) If a language has *tough*-movement, then it has wh-infinitives.

I have two remarks concerning this generalization. First, the implication is in one direction: there are many languages such as Turkish which have wh-infinitives but do not have *tough*-constructions. Second, Sabel (2006) and Gärtner (2009) have argued for the generalizations in (86a) and (86b) respectively. A robust indefinite/interrogative ambiguity refers to languages like English which use different words for *who* vs. *someone* whereas German does not.

- (86) a. If a language has wh-infinitives, then it has infinitival complementizers.
 b. If a language has wh-infinitives, then its pronominal system does not have a robust indefinite/interrogative ambiguity.

Elementary logic allows us to extend Sabel (2006) and Gärtner (2009)’s generalizations to mine:

- (87) a. If a language has *tough*-movement, then it has infinitival complementizers.
 b. If a language has *tough*-movement, then its pronominal system does not have a robust indefinite/interrogative ambiguity.

These generalizations are interesting in their own right, but as we will see in section 5.3, they, along with the other cartographic predictions that I have discussed in this subsection, have consequences on Pesetsky (2021)’s theory of Exfoliation.

4.4 Summary

This section has primarily been concerned with crosslinguistic generalizations on the size of infinitives. I have argued for the following empirical generalization: a high complementizer cannot co-occur with a nonfinite clause. I had a two-pronged approach: I first presented a survey on the maximal size of infinitives in several different languages that have been discussed in the literature, noting that none of them co-occur with a high complementizer.

²²I suspect that Norwegian and Danish may pattern similarly to Swedish. According to Hartman (2011), Italian, French and Spanish all have *tough*-constructions, and according to Sabel (2006) all of these languages have wh-infinitives, so no problem arises. According to Selvanathan (2017), Tamil has *tough*-movement like English, whereas Selvanathan (2018) claims that Malay has *leicht*-movement. The predictions that I am making that Tamil would have wh-infinitives whereas Malay would not. I believe this covers most, if not all, of the languages which have reported to have *tough*-constructions in the literature.

Yet, absence of evidence is not evidence of absence. It could be that such a language simply has yet to be reported. Therefore, I attempted to provide evidence of absence by presenting four different languages—Icelandic, Spanish, Serbian and Mandarin—in which an element with a certain phonetic form behaves as a high complementizer in contexts we would consider finite, but never as a high complementizer in contexts we would consider nonfinite. A plausible explanation for this fact is that nonfinite clauses necessarily cannot co-occur with a high complementizer.

This, I believe, gives us a foundation to create a theory of finiteness in terms of clause size. It allows us to make precise and falsifiable definitions for a clause which is finite and nonfinite.

- (88) a. A clause is finite iff it is untruncated in the C domain.
b. A clause is nonfinite iff its CP2 layer is truncated.

Notice that properties that have often been associated to finiteness in the literature such as tense, subject licensing and agreement are not a part of my definition. Such properties merely *correlate* with the presence of CP2 under my account. Indeed, we have seen examples of the complement of *try*—as small as vP as Wurmbrand et al. (2020) argues—bearing agreement in Serbian, and nonfinite clauses in Tamil licensing subjects and bearing even tense. None of this is contradictory under my theory, as it should be. In the next section, let us see whether this definition of finiteness still holds once we consider a range of facts crosslinguistically.

5 Implications

Many questions remain at the end of section 4, but the three that I focus on are the following:

- (89) a. Do truncated finite clauses exist?
b. Is opacity a problem for clause size theories of finiteness?
c. Do clauses with C-domains truncated in the middle exist?
d. What consequences does this theory have on Exfoliation?

The first and second question instantiate potential counterexamples of my theory of finiteness, which I dedicate section 5.1 to discussing. Section 5.2 discusses the third question, in which I argue that clauses which are truncated in the middle do exist, and are instantiated in the form of subjunctive clauses. I answer the final question in section 5.3: although I believe that my empirical generalization concerning nonfinite clauses is evidence in favor of an Exfoliation-style framework, the fact that infinitives can come in many different sizes is troubling for Pesetsky (2021).

5.1 Potential counterexamples

5.1.1 *That*-less embedded clauses

The central empirical claim of this paper is that infinitives necessarily lack the ability to co-occur with high complementizers. But there is a great deal of controversy in the literature as to whether *that*-less embedded clauses have a CP2 layer or not, which could lead to a confound.²³ For ex-

²³Of course, in the literature previous authors did not refer to CP2; they referred to CP. But to be in line with the rest of this paper I will refer to CP2 rather than CP. For accounts in which CP2 is present but null, the reader is referred to Pesetsky (1992), Pesetsky and Torrego (2001), Pesetsky and Torrego (2007) and Bošković and Lasnik

ample, Bošković and Lasnik (2003) notes the following contrast, in which (90d) cannot occur without the high complementizer but (90b) can:

- (90) a. It was widely believed [that he liked linguistics].
- b. (?) It was widely believed [he liked linguistics].
- c. [That he liked linguistics] was widely believed.
- d. * [He liked linguistics] was widely believed.

Here is the problem. (90b) is uncontroversially finite, but if it truly lacks CP2, this is a counterexample to the definition of finiteness presented in 4.4.

Wurmbrand (2017) provides an interesting discussion of stripping phenomena—the elision of declarative TPs—that may be problematic for my theory of finiteness. Based on the contrast between (91a)-(91b) on one hand and (91c)-(91d) on the other, Wurmbrand (2017) claims that stripping of embedded clauses is only possible when the embedded clause lacks a CP2.

- (91) a. * Abby claimed (that) Ben would ask her out, but she didn't think that Bill (too).
- b. Abby claimed (that) Ben would ask her out, but she didn't think Bill (too).
- c. * Jane loves to study rocks, and John says that geography too.
- d. Jane loves to study rocks, and John says geography too.

For Wurmbrand, ellipsis is the option of not realizing a Spell-Out domain. To get the contrasts in (91a)-(91b) and (91c)-(91d), Wurmbrand assumes a hierarchy CP2 > FocusP > TP. If CP2 is present, CP2 is phasal but not FocusP, and when CP2 is not present FocusP is phasal. The Spell-Out domain of CP2 is FocusP, not TP, so it cannot be elided, because stripping is just the elision of TP. But if CP2 is not present, then TP can be elided, because FocusP is phasal. This allows for a natural explanation of her Embedded Stripping Generalization: that stripping of embedded clauses is only possible if the embedded clause lacks TP. This might imply that CP2 really is missing, and not merely null, in instances of embedded stripping.

It is out of the scope of this paper to contribute to this debate. But it is essential to note that whether or not *that*-less embedded clauses have CP2 or not does not have any bearing on whether the ISG is true or not. If the generalization is true, it has to be explained. But here are two potential strategies to deal with Wurmbrand's generalization.

I could take for granted approaches in which CP2 is present but null in *that*-less embedded clauses, and no problem would arise. Alternatively, the simple definition of finiteness that I present in this paper can be changed slightly to accommodate approaches where CP2 is not present in *that*-less embedded clauses. Recall that the language with the largest attested infinitives from 4.1 is Hebrew. There is at least one more independent reason to believe that Hebrew infinitives are truncated, and that this is not due to the truncation of the CP2 layer.

This data involves negative polarity item (NPI) licensing. Matrix negation can license NPI licensing inside infinitive or subjunctive complements but not indicative ones, as first noted by Landau (2004). This is shown in (92a)-(92c) below; we see that the subjunctive is headed by the high complementizer *še* and still allows NPI licensing, so this restructuring property may be due

(2003). For accounts in which CP2 is truncated, see Hegarty (1991), Webelhuth (1992), Doherty (2000), Svenonius (1994), Bošković (1997) and Wurmbrand (2014).

to the truncation of some other functional projection in the C domain.²⁴

- (92) a. Lo darašti me-Gil ledaber im af-exad.
not demanded.1SG from-Gil to-speak with anybody
'I didn't demand of Gil to speak to anybody.' Infinitive
- b. Lo darašti me-Gil_i še-pro_i yedaber im af-exad.
not demanded.1SG from-Gil that-pro will-speak-3SG.M with anybody
'I didn't demand of Gil that he speak to anybody.' Subjunctive
- c. *Lo he'emanti še-Gil yedaber im af-exad.
not believed.1SG that-Gil will-speak.3SG.M with anybody
'I didn't believe that Gil would speak to anybody.' Indicative, Hebrew

It is likely that there is (at least) one other functional projection together with CP2 that is truncated when a nonfinite clause is made. Let us call this layer XP. One possibility is that CP2 is necessarily deleted whenever XP is deleted. In other words, one could define nonfinite clauses as lacking both CP2 and XP, rather than just CP2 as I have done in this paper, to get the right results with *that*-less embedded clauses. But I must leave open to future research as to what XP is.

5.1.2 Selective opacity

One aspect of my theory that may seem counterintuitive is the fact that different structures vary crosslinguistically in terms of their opacity. According to Keine (2020), nonfinite clauses in Russian are transparent to A'-movement such as topicalization but opaque to A-movement such as subject-to-subject raising, as shown by the contrast in (93a)-(93b) below:

- (93) a. Kažetsja [čto èti studenty znajut tri jazyka].
seem.3SG that these students know.3PL three languages
'It seems that these students know three languages.'
- b. Èti studenty_i kažutsja [t_i učit' tri jazyka].
these students seem.3PL learn.INF three languages
(Intended) 'These students seem to be learning three languages.' Russian

It may seem *prima facie* puzzling that a Russian nonfinite clause with a truncated CP2 layer does not allow raising, unlike English. Equally puzzling is the operation of *hyperraising*—that is, raising from a finite clause—which does not seem to require structure removal to take place. As Wurmbrand (2019) notes, it is a common phenomenon crosslinguistically.

An illustrative example of hyperraising in Greek, which was first noted by Felix (1989), from Bird (1999) is given below. Greek systematically lacks infinitives and allows hyperraising from subjunctive complements co-occurring with an overt complementizer:

- (94) Ta pedhia arxisan na trexoun.
the children.NOM started.3PL COMP.SBJV run.3PL
'The children started to run.' Greek

²⁴I have been unable to verify whether NPI licensing is possible across propositional infinitives in Hebrew. Subjunctives do not seem to have a propositional semantics. If it is not possible, that would indicate that the functional projection of the C domain responsible for this blocking is PropP. If it is possible, then this layer is something else. I have to leave it open to future research as to what the functional projection between CP2 and IntP is that allows NPI licensing.

On one hand, we see that Russian nonfinite clauses do not allow raising. On the other, what seem to be finite clauses in Greek allow it. But I do not think these facts are problematic. I do not know whether Keine (2020)'s theoretical tool of probes having different search domains is the right notion to capture selective opacity effects, due to its lack of independent predictions.

But Keine convincingly shows that selective opacity is a pervasive phenomenon, which is all this theory needs. The lack of raising from Russian nonfinite clauses and possibility of hyperraising in languages like Greek, in my view, are merely an instance of selective opacity effects, and it is not the case that Russian nonfinite clauses are larger than Greek finite clauses.

5.1.3 Factives are not truncated

As has been noted extensively in the literature thus far, factives do not allow many of the properties of the C domain such as topicalization or focalization, as Hooper and Thompson (1973), Haegeman (2012) and others point out. An example of topicalization with the complement of *regret* is given below:

- (95) * John regrets that this book Mary read.

This has led Miyagawa (2017) to claim that factives are in fact truncated in the C domain. This is at odds with my definition of a finite clause, which is fully untruncated in the C domain. As such, I adopt and defend Haegeman (2012)'s analysis of null operator movement in complements of factive predicates, rather than truncation. I present some corroborating evidence for her account in the form of infinitives, as well.

Haegeman is not the first to suggest null operator movement in factives. Hegarty (1992) points out that the complement clauses of factives are weak islands for extraction, whereas those of non-factives are not, as seen below.

- (96) a. How do you suppose that Maria_i fixed the car t_i?
 b. * How did you notice that Maria_i fixed the car t_i?
 c. Why does Mary_i think that Bill left the company t_i?
 d. * Why does Mary_i regret that Bill left the company t_i?

As Haegeman (2012) points out, almost every property of the C domain that we have discussed thus far involves a step of A'-movement. Both null operators and a truncation analysis would get the desired result as both disallow movement. If it ever were possible to base-generate elements into a Spec position in the C-domain, for example Spec,TopP, then it would be possible to distinguish between the accounts, as they make different predictions.

Temporal adjuncts, in fact, seem to be base-generated into a Spec position of the articulated left periphery. Rizzi (1997) assumes they are Merged to Spec,TopP, although Rizzi (2001) distinguishes the position of topics from modifiers, positing a dedicated projection, ModP. However, for simplicity, I will continue assume that it is Merged onto Spec,TopP:

- (97) [TopP Last week, [TP I was in Tokyo.]]

If temporal adjuncts are base-generated, then we would predict that they should be acceptable with factives. This prediction is borne out:

- (98) John regrets that during dinner Mary read this book.

As mentioned in 3.2, colloquial English appears to have double complementizer constructions:

- (99) She maintained **that** when they arrived **that** they would be welcomed.

According to my consultants, this sentence is equally acceptable with the factive *regret*, indicating the presence of CP2, TopP and CP1 layers and therefore a highly articulated structure:

- (100) She regretted **that** when they arrived **that** they weren't welcomed.

On the other hand, we would also predict that, as English infinitives are quite truncated, that they cannot take temporal adjuncts. This prediction is borne out, according to data from Shlonsky and Soare (2011). In the contrast below, the adjunct *at 5* cannot refer to the cooking of dinner; it must refer to the time of the promise—that is, it must be an adjunct to the matrix sentence rather than the infinitive. However, this is possible with the finite version of the sentence:

- (101) a. *John promised us at 5 to cook dinner for his children.
b. John promised us that at 5 he would cook dinner for his children.

I conclude that factives are not truncated in the C domain.

5.2 Subjunctives and "finite" control

We now move to subjunctive clauses. The status of the finiteness of subjunctives has been perplexing for decades: they seem to both have finite (for example agreement, high complementizers) and nonfinite properties (OC PRO); see, for example, Landau (2004). I believe that the novel approach to finiteness in the paper provides a new angle for understanding the finiteness of subjunctives. I will provide novel evidence from the C domain from subjunctives to show that they are, as has been claimed in the literature, borderline between finite and nonfinite.

However, there does not appear to be one unified structure for all subjunctive clauses crosslinguistically. That is, in languages like English and Hebrew, they must be headed by a high complementizer, indicating the presence of CP2, but they also appear to have some truncation of the C domain in the middle. On the other hand, in languages like Spanish, subjunctives *must* be headed by a low complementizer *que*. Though I must leave a complete account of subjunctives open for future research, here is how such an investigation might take place under this account.

Recall the following paradigm from 4.1:

- (102) a. *No topicalization within infinitives*: *I wanted this book, to read.
b. *No focalization within infinitives*: *I wanted THIS BOOK to read.
c. *No why-infinitives*: ??I asked Caitlin why to eat salad.
d. *No if*: *I asked Caitlin if to eat salad.
e. *No temporal adjunct*: *I asked Caitlin during dinner to eat salad.

Surprisingly, even though subjunctives are headed by a CP2 projection, most of these tests fail.

- (103) a. *No topicalization within infinitives*: *I suggested that this book he read.
b. *No focalization within infinitives*: *I suggested that THIS BOOK he read.
c. *No why-infinitives*: *I suggested why she eat salad.
d. *No if*: *I suggested that if he eat ice cream, then he exercise.

- e. *No temporal adjunct*: ??I suggested that during dinner she eat salad.

This indicates that subject licensing in English is somehow tied to the presence of a CP2 projection: PRO can be licensed with complements as small as TP or even vP (following W&L), but a full subject which is not merely a minimal bound pronoun requires CP2.

Under accounts of finiteness like Bouchard (1984), Koster (1984) and Hornstein and Lightfoot (1987), and Pesetsky (2021), obligatory control (OC) is possible into clauses which are as large as IP/TP, whereas CPs block OC—the latter of which are seen as phases in today’s minimalist framework. Landau (2013) considers clause size a "bogus" criterion for OC, because there seem to be cases of so-called "finite" control in languages like Hebrew and the Balkan languages. In Landau (2004)’s example (104) from Hebrew below, Landau argues that the embedded clause is in the subjunctive mood, and headed by the high complementizer *še*.²⁵ The null subject of the embedded clause must refer to *Gil*.

- (104) himlacti le-Gil_i še-ec_{i/*k} yearšem la-xug le-balšanut.
 I-recommended to-Gil that-ec will-register.3SG.M to-the-department to-linguistics
 ‘I recommended to Gil to register to the linguistics department.’ Hebrew

At the time, this was a very strong argument that control complements can be as large as CP. Indeed, it’s also unclear how an Exfoliation framework could derive "finite" control constructions. But with the articulated left periphery that I have assumed in this paper, we need to reanalyze what we mean by finite control. It appears that finite control complements, such as in Hebrew, can be as large as CP2. But it is possible that these "finite" control complements are in fact truncated in the middle, and hence, not finite but not nonfinite either under my account.

I will claim that control is an operation which can only take place in complements that are truncated in the C domain by discussing restructuring phenomena in "finite" control complements. Ultimately, both defenders of clause size theories of finiteness on one side such as myself and Pesetsky, and Landau on the other side end up both being right: "finite" control complements do project a CP2 layer and contain a high complementizer.

It will not be easy to determine whether subjunctives in Hebrew are truncated at all, given that Hebrew infinitives are the largest on record. I suspect that Hebrew subjunctives might also be equally large, with an additional CP2 layer on top. But there is still independent evidence that subjunctives are truncated, as well. Recall from section 5.1 above the pattern with NPI licensing across clause boundaries in Hebrew; matrix negation can license NPIs across infinitive and subjunctive complement clauses but not indicative ones:

²⁵That *še* is a high complementizer in indicative clauses can be verified with the following example from Shlonsky (2014), in which the topicalized or focalized constituent *Dani* follows *še*.

- (i) ani xošev še et Dani_i pitru t_i.
 I think that DOM Dani, (they)-fired
 ‘I think that Dani, they fired.’

Furthermore, this complementizer behaves as such in subjunctive clauses as well (Ur Shlonsky, p.c.):

- (ii) Hem_i kivu še ha-bayta hem_{i/k} yelxu t mukdam.
 they hoped that home they will-go.3PL early
 ‘They hoped that they would go home early.’

For independent reasons, a null subject in a subjunctive with topicalized/focalized elements is ruled out.

- (105) a. Lo darašti me-Gil ledaber im af-exad.
not demanded.1SG from-Gil to-speak with anybody
'I didn't demand of Gil to speak to anybody.' Infinitive
- b. Lo darašti me-Gil_i še-pro_i yedaber im af-exad.
not demanded.1SG from-Gil that-pro will-speak-3SG.M with anybody
'I didn't demand of Gil that he speak to anybody.' Subjunctive
- c. *Lo he'emanti še-Gil yedaber im af-exad.
not believed.1SG that-Gil will-speak.3SG.M with anybody
'I didn't believe that Gil would speak to anybody.' Indicative

I assumed in 5.1 that this restructuring phenomenon was possible *because* some functional projection common to both the Hebrew infinitive and subjunctive was truncated, calling it XP.²⁶

As it turns out, restructuring phenomena seems to be common with subjunctive control complements crosslinguistically. Ewe subjunctive control complements patterns with Hebrew in terms of NPI-licensing; it is possible across subjunctive clauses headed by complementers, which have an overt PRO, as Satik (2019) argues, but not ones in the aorist mood:

- (106) a. *Kofi_i me-be yè_i dzo o. b. Kofi_i me-be yè_i-a dzo o.
Kofi NEG1-COMP YÈ leave NEG2 Kofi NEG1-COMP YÈ-POT leave NEG2
'Kofi_i didn't say that he_i left.' Ewe 'Kofi_i didn't say that he_i could leave.'

This pattern isn't limited to NPI licensing across subjunctive clauses; in line with Keine's selective opacity effects, we find that different subjunctive complements are transparent to different operations crosslinguistically. Felix (1989), for example, points out subjunctive complements in Greek are transparent to A-movement, allowing raising in addition to control; Watanabe (1993) notes the same for Romanian—in both languages, indicatives are opaque to A-movement.²⁷ Landau describes both of these languages as exhibiting finite control.

We've so far seen languages—English, Hebrew and potentially Ewe—in which the complementizer appears to be located in CP2. But there is reason to believe that not all subjunctives are truncated in the middle; sometimes, in Spanish, they may be truncated at the top like infinitives, as well. According to Villa-Garcia (2012), *que* in Spanish is a low complementizer in CP1, exclusively associated with the jussive or optative mood. In such a context, the topic must precede *que*:

- (107) *A la fiesta* *(que) vayan
to the party that go.3PL.SUBJ
'I demand that they go to the party.' Spanish

²⁶It appears that the only way to handle such a case is by assuming Keine (2020)'s horizon framework for syntactic operations. That is, the probe that is responsible for NPI licensing in Hebrew is sensitive to the projection XP of the C-domain—XP is only found in indicative embedded clauses and not in subjunctives or infinitives—rather than the topmost, and likely phasal, CP2.

²⁷Alexiadou et al. (2010) argue that Greek subjunctive complements cannot be analyzed as instances of restructuring. Their evidence is based on two facts: first, event modifiers can modify the event of both the matrix and embedded clause. Second, they also claim that NPI licensing can take place across the subjunctive clause boundary, but it can also be in the matrix clause, as well. I do not find these arguments convincing: even in languages like English, the infinitive complement of *try* can be modified by an event modifier, so even if restructuring was present we would predict this to be possible. NPI licensing itself might merely be an instance of selective opacity: for example, Hebrew bans NPI licensing across indicative clause boundaries while English allows it, so this is again not surprising. Felix's observation and the fact that indicative complements do not allow raising is itself evidence for restructuring.

According to Villa-Garcia (2012), Spanish subjunctives can optionally have a high complementizer and an overt realization of the Top⁰ as well—all of the form *que*. Villa-Garcia (2012) provides an example with two topics, indicating the presence of an articulated left periphery:

- (108) a. Que *a tu hijo*, (que) como *va a suspender*, *(que) lo castiguen
 that your son that since goes to fail that cl. punish.3PL.SUBJ
 ‘I/somebody ordered that they punish your son, since he’s going to fail (the course).’ Spanish
- b. [_{CP2} [_{C2} que [_{TopP} Topic 1 [_{Top} que [_{TopP} Topic 2 [_{Top} ∅ [_{CP1} [_{C1} que ...]]]]]]]]]

Although this subjunctive is highly truncated, it’s not in principle impossible for it to be truncated in the middle as well. One would have to verify, for example, whether contrastive focus is possible in these constructions. But there seem to be at least two strategies that are attested for subjunctives crosslinguistically.

It is not clear which group Japanese subjunctives belong to, but there is evidence that they are truncated, as well. Uchibori (2000) extensively notes selective opacity effects in Japanese subjunctive complements, which also have been noted to exhibit finite control with some, but not all predicates. Here I will focus on the subjunctives that allow control, though the transparency effects obtain for the ones that do not as well.

Crucially, though, Uchibori’s solution is to posit that although subjunctives in Japanese are CPs, the C head is not a strong phase. It would be possible to get Uchibori’s result, however, under a Rizzi-style framework, if we assume that the subjunctive complementizers are low complementizers in CP1 and the phasal CP2 has been truncated—or perhaps some truncation in the middle, as in Hebrew. Here is the data Uchibori discusses to come to this conclusion. Uchibori notes that scrambling out of subjunctive complements can remedy WCO violations (109a), but not out of an indicative complement (109b):

- (109) a. Daremo_i-o [[soitsu_i-no hahayoya]-ga [iinkai_k-ni e_k t_i
 everyone-ACC guy-GEN mother-NOM committee-DAT
 suisensu-ru-yoo(-ni(-to))] tanon-da].
 recommend-NONPAST-SUBJ-COMP ask-PAST
 (lit.) ‘Everyone_i, his_i mother asked the committee to recommend.’
- b. *Daremo_i-o [[soitsu_i-no hahayoya]-ga [iinkai-ga t_i suisensi-ta
 everyone-ACC guy-GEN mother-NOM committee-NOM recommend-PAST
 to] omot-ta].
 COMP think-PAST
 (lit.) ‘Everyone_i, his_i mother thought that the committee recommended.’ Japanese

Furthermore, a quantifier scrambled out of subjunctive complements may have wide scope over other quantifiers (110a), but not out of indicative clauses (110b):

- (110) a. Daremo-o_i [dareka-ga iinkai_j-ni [e_j t_i
 Everyone-ACC someone-NOM committee-DAT
 suisenru-ru-yoo(ni(-to))] meiji-ta].
 recommend-NONPAST-SUBJ-COMP order-PAST
 ‘Everyone, someone ordered the committee to recommend.’ (∀ > ∃)

- b. Daremo_i-o [dareka-ga [John-ga t_i hihansi-ta to] it-ta]
 Everyone-ACC someone-NOM John-NOM criticize-PAST COMP say-PAST
 (lit.) ‘Everyone, someone said that John criticized.’ (*∀ > ∃)

Finally, Uchibori notes that the reciprocal anaphor *otagai* ‘each other’ must be locally A-bound. Scrambling out of a subjunctive complement can license the anaphor (111a), but not out of an indicative one (111b):

- (111) a. ? Karera_i-o [otagai_i-no sensei-ga [John_j-ni [e_j t_i
 them-ACC each.other-GEN teacher-NOM John-DAT
 hihansu-ru-yoo(-ni(-to))] it-ta.
 criticize-NONPAST-SUBJ-COMP tell-PAST
 (lit.) ‘Them_i, each other_i’s teacher told John to criticize.’
- b. * Karera_i-o [otagai_i-no sensei-ga [John-ga t_i hihans-ita to]
 teacher-ACC each.other-GEN teacher-NOM John-NOM criticize-PAST COMP
 it-ta.
 say-PAST
 (lit.) ‘Them_i, each other’s teachers said that John criticized.’ Japanese

I believe that these observations from Hebrew, Ewe, Spanish and Japanese and to a lesser extent Greek and Romanian show that all control complements may be truncated in some regard. To conclude, my theory of finiteness in terms of clause size comes with the welcome advantage of eliminating the surprising phenomenon of "finite" control, which it turns out, based on my theory, is not so surprising after all. This is a new line of research worth exploring.

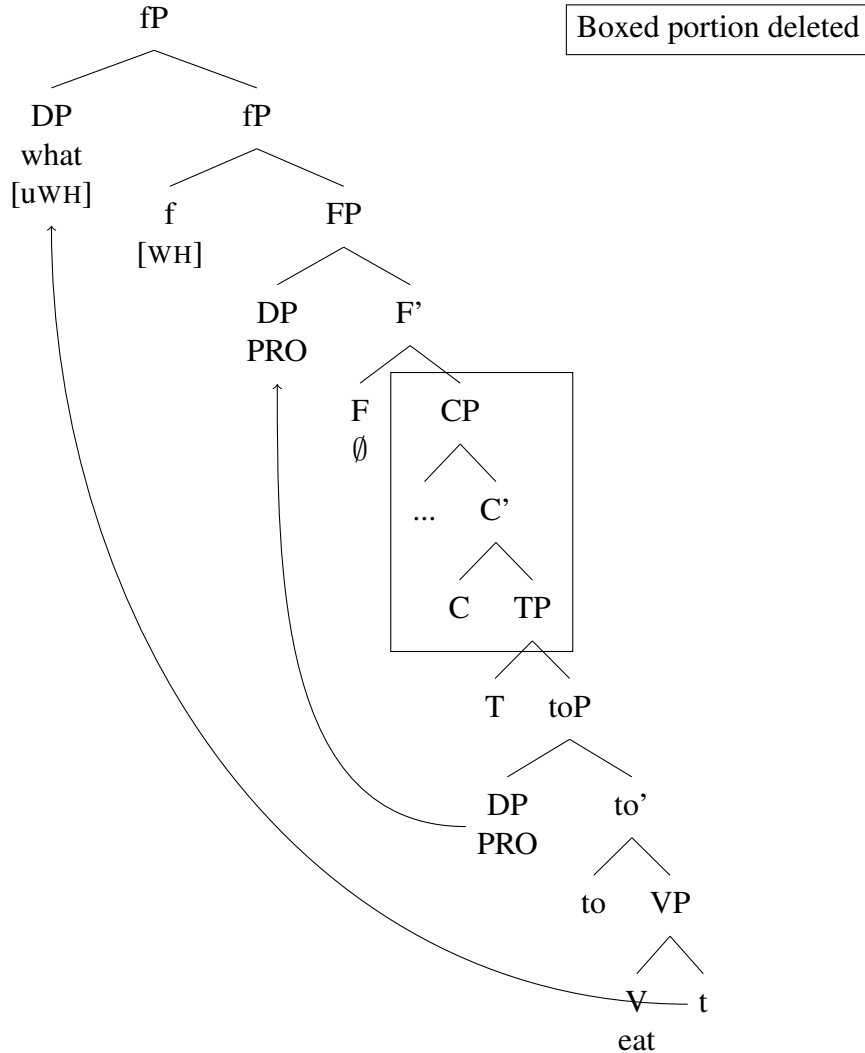
5.3 Consequences on Exfoliation

On one hand, I believe that my empirical generalization—that nonfinite clauses by definition lack a CP2 layer, in which high complementizers are realized—constitutes strong evidence, in my view, that some kind of derivational process, perhaps Müller (2020)’s operation Remove or Pesetsky (2021)’s Exfoliation, is responsible for the truncated size of infinitives. These operations take place because Rizzi’s ForceP, or my CP2, is a phase head, and it serves as a barrier for syntactic operations like subject extraction in the case of raising. Under a non-derivational theory of finiteness, this would be coincidental: why should it be the case that nonfinite clauses cannot co-occur with a high complementizer? Derivational frameworks *predict* my generalization.

On the other, the evidence from section 4 is at odds with Pesetsky’s "one-size-fits-all" approach, where all infinitives have the same size: toP, apart from the superstructure that is sometimes added. To see where this goes wrong, let us see an attempt, under the Exfoliation framework, to derive a wh-infinitive such as *I know what to eat*. In this tree, f⁰ has a WH-feature allowing the wh-infinitive to be formed.²⁸

²⁸I am omitting the movement of F⁰ to f⁰ for simplicity. One might object that this tree violates minimality conditions on movement. See, for example, Preminger (2014) on why it does not: the probe on f⁰ looks specifically for WH-features even if PRO is a more local DP. It can skip past PRO because it does not have WH-features.

(112)



This sets the stage to present the first problem with the Exfoliation framework: it misses generalizations concerning the size of infinitives cross-linguistically. That is, it is not obvious under Exfoliation why *wh*-infinitives do not exist in languages like Hindi, German, Swedish and Icelandic, or why infinitival complementizers do not exist in Hindi and German, given that *all* of these languages have control constructions and hence, superstructures.

Under my account, the presence of infinitival complementizers in English is predicted from the presence of *wh*-infinitives. The differing properties of *tough*-movement in German, Swedish and English is also predicted: the infinitives of these languages come in three different maximal sizes, which are TP, CP1 and WhP. Superstructures do not allow such predictions to be made, because superstructures across languages are the same size: it is mysterious why English has infinitival complementizers, why maximal infinitive size correlates with the kind of *tough*-movement that is present, and why TP and CP1 languages do not have *wh*-infinitives, among other facts.²⁹

²⁹In addition, although I did not discuss this in detail in section 2 due to space constraints, Pesetsky assumes that *for* is not a complementizer but rather an irrealis marker. I believe that section 4.1 strongly implies that this analysis is incorrect, given that we can predict whether a language has infinitival complementizers or not if it has *wh*-infinitives. There seems to be no language in my survey which has *wh*-infinitives but does not have infinitival

To start accounting for the lack of wh-infinitives in TP and CP1-languages, it is possible for David Pesetsky (p.c.) to claim these languages do not license WH-features on their superstructures whereas English does. But to see where this goes wrong, recall that there are even larger infinitives than those of English: Italian's infinitives go up to TopP, Hebrew's go up to IntP. Once again, the presence of WhP and CP1 in Italian is predicted from the presence of TopP, which the Exfoliation framework does not predict; the presence of FocP, TopP, WhP and CP1 is predicted in Hebrew from the presence of IntP, none of which the Exfoliation framework predicts.

To account for these predictions, it seems impossible to not assume a Rizzi-style left periphery for superstructures. But at this point his superstructure has become indistinguishable from Rizzi's articulated C domain, and ultimately, he has to end up admitting that infinitives come in different sizes, as well. That is, the Exfoliation framework must concede that languages can select the sizes of their superstructures. The sizes of the superstructures themselves *cannot* be derived; without further stipulations, this could lead to a problem of infinite regress, with superstructures Exfoliating themselves, without end.

Ultimately, I believe that a derivational theory of finiteness is on the right track. However, a derivational theory must admit that there are empirical generalizations concerning the maximal size of infinitives crosslinguistically, and this must be accounted for. I believe that each language must *select* the maximal size of their infinitives; perhaps such a process ends up being completely random. But, crucially, the fact that the maximal size cannot be CP2 is derived via something an operation like Exfoliation or the Müller (2020)'s operation Remove.

6 Conclusion

This paper has been an investigation on the size of infinitives. After laying the groundwork for this endeavor in sections 1-3, section 4 presented evidence that the size of infinitives can vary crosslinguistically. I showed that generalizations concerning the maximal size of infinitives crosslinguistically are difficult to for under an Exfoliation-style framework with superstructures in section 5. Languages must be able to select the maximal size of their infinitive. And yet, I also gave what I believe to be a very strong argument in favor for a derivational theory of finiteness like Exfoliation: the Infinitive Size Generalization—that infinitives can never co-occur with high complementizers. This is surprising under selectional theories, and requires at least some derivational element to finiteness.

I have argued that finiteness really is a matter of clause size, and defined finite clauses as those which are untruncated in the C domain, whereas nonfinite clauses are those which lack a CP2 layer. This paper has only investigated the size of finite and nonfinite *embedded* clauses, not root ones. As such, there are many open questions left at the conclusion of this paper. The most obvious one is that *that* can only appear with embedded clauses:

(113) (*That) Caitlin likes chocolate.

This is a question for all theories, and not mine specifically—perhaps there is just an independent requirement for *that* to be pronounced only with embedded clauses, but it is always present. But this is by no means a universal; as we will see, it seems to be attested in the Romance languages.

complementizers as well. This implies that *for* is a low complementizer.

Furthermore, one prediction that my account makes involves projections above CP2, which appear to be attested in at least Romance. Cruschina and Remberger (2018) discusses constructions in Romance in which a complementizer is present in root clauses, and is preceded by an adjective or an adverb. It appears to be preceded by an adjective in (114):

- (114) Certo che la capito!
 certain that have.3SG understand.PST.PTCP
 ‘Of course she understood!’ Italian

For Cruschina and Remberger (2018), this indicates that there are a set of projections above CP2 which encode speaker-oriented and pragmatic features such as evaluative, evidential or epistemic values. The prediction is that no projection above CP2 should be present not just with infinitives, but with other nonfinite clauses like imperatives as well.

Of course, I have not discussed the nature of imperatives like "*Catch her!*" and how they come into being. They are puzzling for an Exfoliation-style framework as well, given that all embedded clauses are born finite, yet it's not clear if the technology could be extended to root nonfinite clauses. I leave this open for future research to look into. But it is natural to suppose that they are missing many functional projections, leading to a truncated, nonfinite structure.

It also remains to be seen how this account can be extended to gerunds, which have a nominal nature, and structures like nominalized infinitives in Turkish—for which I presented evidence in section 4.1 that it is truncated. But at the very least—no matter what one thinks of the analysis of finiteness in this paper—the goal of this paper has been to introduce the reader to novel empirical generalizations concerning nonfinite clauses. The inability for nonfinite clauses to appear with high complementizers under the articulated C domain is a mystery worth investigating.

References

- Adger, David. 2007. *Three domains of finiteness: a minimalist perspective*. OUP Oxford.
- Al-Aqarbeh, Rania. 2011. Finiteness in Jordanian Arabic: a semantic and morphosyntactic approach. Doctoral Dissertation, University of Kansas.
- Alexiadou, A., E. Anagnostopoulou, Gianina Iordachioaia, and M. Marchis. 2010. No objections to Backward Control?
- Barbosa, Pilar. 2001. On inversion in wh-questions in Romance. In *Subject inversion in Romance and the theory of universal grammar*, ed. Aafke C. Hulk and Jean-Yves Pollock, 20–59. Oxford University Press.
- Barrie, Michael. 2007. Control and wh-infinitivals. In *Studies in natural language and linguistic theory*, 263–279. Springer Netherlands.
- Bhatt, Rajesh. 2006. *Covert modality in non-finite contexts*. Walter de Gruyter.
- Bird, Steven. 1999. Multidimensional exploration of online linguistic field data. In *North East Linguistics Society*, ed. Tamanji Pius, Masako Hirotani, and Nancy Hall. University of Delaware: GLSA.
- Bocci, Giuliano. 2007. Criterial positions and left periphery in Italian: Evidence for the syntactic encoding of contrastive focus. In *Papers from the consortium Workshops on Linguistic Theory*, volume 3.

- Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, Massachusetts: MIT Press.
- Bošković, Željko, and Howard Lasnik. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34:527–546.
- Bouchard, Denis. 1984. *On the content of empty categories*. Dordrecht.
- Carstens, Vicki, and Michael Diercks. 2009. Parameterizing case and activity: Hyper-raising in Bantu. In *Proceedings of NELS 40*.
- Chomsky, Noam. 1977a. Conditions on transformations. In *Essays on form and interpretation*, 81–162. New York, New York: Elsevier North-Holland, Inc.
- Chomsky, Noam. 1977b. On wh-movement. In *Formal Syntax*, ed. P. Culicover, T. Wasow, and A. Akmajian. New York, New York: Academic Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in linguistics*, ed. Michael Kenstowicz, 1–52. Cambridge, Massachusetts: MIT Press.
- Comrie, Bernard. 1997. Tough movement and its analogs in germanic languages. In *Language and its ecology*. DE GRUYTER MOUTON.
- Cruschina, Silvio, and Eva-Maria Remberger. 2018. Speaker-oriented syntax and root clause complementizers 18:336–358.
- Dasgupta, Probal. 1982. Infinitives in Bangla and English: phrases or clauses. *Bulletin of the Deccan College Research Institute* 41:176–209.
- Demonte, Violeta, and Olga Fernández-Soriano. 2009. Force and finiteness in the spanish complementizer system 21.
- Doherty, Cathal. 2000. *Clauses without that: The case for bare sentential complementation in English*. Garland.
- Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht, The Netherlands: D. Reidel Publishing Company.
- Faarlund, Jan Terje. 2015. The Norwegian infinitive marker 1–10.
- Felix, Sascha. 1989. Finite infinitives in modern greek. In *Linguistik aktuell/linguistics today*, 113–132. John Benjamins Publishing Company.
- Gärtner, Hans-Martin. 2009. More on the indefinite-interrogative affinity: The view from embedded non-finite interrogatives 13:1–37.
- van Gelderen, Elly. 1998. The future of for to 10:45–71.
- Grano, Thomas. 2017. Finiteness contrasts without Tense? A view from Mandarin Chinese. *Journal of East Asian Linguistics* 26:259–299.
- Haegeman, Liliane. 2006. Argument fronting in English, Romance CLLD, and the left periphery. In *Crosslinguistic research in syntax and semantics. negation, tense and clausal architecture*, ed. Rafaella Zanuttini, Héctor Campos, Elena Herburger, and Paul H. Portner, 27–52. Georgetown University Press.
- Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and the composition of the left periphery*. Oxford University Press.
- Hartman, Jeremy. 2011. Intervention in tough-constructions. In *Proceedings of the 39th Meeting of the North East Linguistic Society (NELS 39)*, ed. Suzi Lima, Kevin Mullin, and Brian Smith, 387–398. GLSA.
- Hegarty, Michael. 1991. Adjunct extraction and chain configurations. Doctoral Dissertation, MIT.

- Hegarty, Michael. 1992. Adjunct extraction without traces. In *The Proceedings of the Tenth West Coast Conference on Formal Linguistics*, ed. Dawn Bates, 209–222. Stanford, California: Center for the Study of Language and Information.
- Henry, Alison. 1995. *Belfast English and standard English*. Oxford: Oxford University Press.
- Hooper, Joan B., and Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic inquiry* 4:465–499.
- Hornstein, Norbert. 1999. Movement and Control. *Linguistic Inquiry* 30:69–96.
- Hornstein, Norbert, and David Lightfoot. 1987. Predication and PRO 63:23–52.
- Hsu, Brian. 2015. Variation in Bangla complementizer order at the syntax-prosody interface. In *Proceedings of NELS 45*. GLSA.
- Huang, Nick. 2018. Control complements in mandarin chinese: implications for restructuring and the chinese finiteness debate 27:347–376.
- Kalm, Mikael. 2016. Control infinitives and ECM-infinitives in Swedish: A synchronic and diachronic investigation. Doctoral Dissertation, Uppsala University.
- Kayne, Richard S. 1981. ECP extensions. *Linguistic Inquiry* 12:93–133.
- Keine, S. 2020. *Probes and their horizons*. Linguistic Inquiry Monographs. MIT Press.
- Klingvall, Eva. 2018. Agreement and reconstruction correlate in Swedish 36:1165–1205.
- Kornfilt, Jaklin. 1996. On some infinitival wh-constructions in Turkish.
- Koster, Jan. 1984. On binding and control. *Linguistic Inquiry* 15:417–459.
- Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22:811–877.
- Landau, Idan. 2013. *Control in generative grammar: A research companion*. Cambridge, England: Cambridge University Press.
- Landau, Idan. 2015. *A two-tiered theory of control*. MIT Press.
- Larsson, Ida. 2017. Double complementizers. *Nordic Atlas of Language Structures Journal* 1.
- Lujan, Marta. 1980. Clitic promotion and mood in Spanish verbal complements 381–484.
- McFadden, Thomas, and Sandhya Sundaresan. 2014. Finiteness in south asian languages: an introduction. *Natural Language & Linguistic Theory* 32:1–27. URL <http://www.jstor.org/stable/43697712>.
- Miyagawa, Shigeru. 2017. Topicalization 152:1–29.
- Müller, Gereon. 2020. Rethinking restructuring. In *Syntactic architecture and its consequences II: Between syntax and morphology*, 149–190. Language Science Press.
- Nikolaeva, Irina. 2007. *Finiteness: Theoretical and empirical foundations*. OUP Oxford.
- Pearson, Hazel. 2015. The interpretation of the logophoric pronoun in Ewe. *Natural Language Semantics* 23:77–118.
- Pesetsky, David. 1992. *Zero syntax*, volume 2.
- Pesetsky, David. 2021. Exfoliation: towards a derivational theory of clause size. URL <https://ling.auf.net/lingbuzz/004440>.
- Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 355–426. MIT Press.
- Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: Syntactic derivation and interpretation*, ed. Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 262–294. Amsterdam: John Benjamins Publishing Company.

- Platzack, Christer. 1986. The structure of infinitive clauses in Danish and Swedish. In *Scandinavian syntax*, ed. Dahl sten and Anders Holmberg, 123–137. Stockholm Sweden: University of Stockholm.
- Preminger, Omer. 2014. *Agreement and its failures*. Number 68 in Linguistic Inquiry Monographs. MIT Press.
- Raposo, Eduardo. 1987. Case theory and Infl-to-Comp: the inflected infinitive in European Portuguese. *Linguistic Inquiry* 18:85–109.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.
- Rizzi, Luigi. 2001. Reconstruction, weak island sensitivity and agreement. In *Semantic interfaces: reference, anaphora and aspect*, ed. Carlo Cecchetto, Gennaro Chierchia, and Maria Teresa Guasti, 145–176. Stanford, California: CSLI.
- Sabel, J. 2006. Impossible infinitival interrogatives and relatives. In *Form, structure, and grammar*.
- Satık, Deniz. 2019. Control is not raising: evidence from overt split control in Ewe. In *GLOW 42: Workshop on anaphora at the syntax-semantics-pragmatics interface*. Oslo, Norway.
- Satık, Deniz. 2020. Exfoliating the implicational universal in complementation.
- Selvanathan, Nagarajan. 2017. Subject theta roles and Tamil tough movement. URL <https://ling.auf.net/lingbuzz/003767>.
- Selvanathan, Nagarajan. 2018. Malay tough movement, restructuring and the theory of phases. URL <https://ling.auf.net/lingbuzz/004029>.
- Shlonsky, Ur. 2014. Topicalization and focalization: a preliminary exploration of the hebrew left periphery. In *Peripheries*.
- Shlonsky, Ur, and Gabriela Soare. 2011. Where’s ‘why’? *Linguistic Inquiry* 42:651–669.
- Sigurðsson, Halldór Ármann. 1989. Verbal syntax and Case in Icelandic. Doctoral Dissertation, Department of Scandinavian Languages, University of Lund.
- Svenonius, Peter. 1994. Dependent nexus: Subordinate predication structures in English and the Scandinavian language. Doctoral Dissertation, University of California, Santa Cruz.
- Swart, Henriette de, and Helen de Hoop. 2000. Topic and focus. In *The first glot international state-of-the-article book*, ed. Lisa L.-S. Cheng and Rint Sybesma, 105–130. New York: Mouton de Gruyter.
- Szabolcsi, Anna. 2009. Overt nominative subjects in infinitival complements: Data, diagnostics, and preliminary analyses .
- Szécsényi, Krisztina. 2009. On the double nature of Hungarian infinitival constructions 119:592–624.
- Thráinsson, H. 2007. *The Syntax of Icelandic*. Cambridge Syntax Guides. Cambridge University Press. URL <https://books.google.com/books?id=70dbeoPHgzgC>.
- Thráinsson, Höskuldur. 1993. On the structure of infinitival complements 3.
- Todorović, Neda, and Susi Wurmbrand. 2016. *Finiteness across domains*.
- Uchibori, Asako. 2000. The syntax of subjunctive complements: evidence from japanese. Doctoral Dissertation, University of Connecticut.
- Ussery, Cherlon, Lydia Ding, and Yining Rebecca Liu. 2016. The typology of mandarin infinitives. *Proceedings of the Linguistic Society of America* 1:28.

- Villa-Garcia, Julio. 2012. The Spanish complementizer system: Consequences for the syntax of dislocations and subjects, locality of movement, and clausal structure. Doctoral Dissertation, University of Connecticut.
- Villalba, Xavier. 2009. *The syntax and semantics of dislocations in catalan. a study on asymmetric syntax at the peripheries of sentence*. Lambert.
- Watanabe, Akira. 1993. Agr-based case theory and its interaction with the A-bar system. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge.
- Webelhuth, Gert. 1992. *Principles and parameters of syntactic saturation*. Oxford University Press.
- Wheelock, Kane. 2015. German wh-infinitives: restructuring and cyclic movement. Master's thesis, New York University.
- Wurmbrand, Susanne, Magdalena Lohninger, Neda Todorovic, Iva Kovac, and Caroline Pajancic. 2020. Finiteness in south slavic complement clauses: Evidence for an implicational finiteness universal 119–137.
- Wurmbrand, Susi. 1994. The German leicht-construction. In *Wiener Linguistische Gazette*.
- Wurmbrand, Susi. 2001. *Infinitives: Restructuring and clause structure*. Mouton de Gruyter.
- Wurmbrand, Susi. 2014. The merge condition. In *Language faculty and beyond*, 130–166. John Benjamins Publishing Company.
- Wurmbrand, Susi. 2017. Stripping and topless complements 48:341–366.
- Wurmbrand, Susi. 2019. Cross-clausal A-dependencies. In *Proceedings of CLS 54*.
- Wurmbrand, Susi, and Magdalena Lohninger. 2019. An implicational universal in complementation: Theoretical insights and empirical progress. In *Propositional arguments in cross-linguistic research: Theoretical and empirical issues*, ed. Jutta M. Hartmann and Angelika Wollstein. Mouton de Gruyter.