Building the meaning of a long-distance question via prolepsis: the case of Georgian

Tatiana Bondarenko
Harvard University

1 Introduction

The Georgian language lacks long-distance *wh*-movement (e.g., see Harris 1981): it is not possible to move *wh*-words into the matrix clause regardless of the word order, (1)-(2), and it is also not possible to leave them in situ, (3).

(1) *$\textbf{ra-s}_1$ pikrob-s mariami [rom šota $\textbf{t}_1$ č'am-s]?
what-ACC think-PRS.3SG Mariam COMP Shota eat-PRS.3SG
'What does Mariam think that Shota is eating?'[1]

(2) *mariami $\textbf{ra-s}_1$ pikrob-s [rom šota $\textbf{t}_1$ č'am-s]?
Mariam **what-ACC** think-PRS.3SG COMP Shota eat-PRS.3SG
'What does Mariam think that Shota is eating?'

(3) *pikrob-s mariami [rom šota **ra-s** č'am-s]?
think-PRS.3SG Mariam COMP Shota **what-ACC** eat-PRS.3SG
'What does Mariam think that Shota is eating?'

In order to convey the meaning of a long-distance *wh*-question, a construction involving prolepsis can be used, (4):[2] there is a proleptic *wh*-phrase in the matrix clause (with the postposition *ze* 'on') and a corresponding gap in the embedded CP.

[1]Dative and Accusative cases are syncretic in Georgian. I gloss the form as accusative if the DP is a direct object and as dative if it is an indirect object/applicative argument.

[2]There are two other strategies that can also be used to express the meaning of a long-distance *wh*-question. One is a *wh*-scope marking construction, (1), where an invariant *wh*-expression *ra* 'what' in the matrix clause co-occurs with an embedded clause that contains a *wh*-item. Another strategy involves a proleptic *wh*-phrase in the matrix clause co-occurring with another *wh*-expression in the embedded clause, (2). I leave investigation of these strategies for future research.

(1) *Wh Scope Marking*
ra-s pikrobs mariami, [rom šota **vis** xedavs]?
what-ACC thinks Mariam.NOM COMP Shota.NOM **who.ACC** sees
'Who does Mariam think that Shota sees?'
(lit. 'What does Mariam think that Shota sees who?')

(2) *Prolepsis + wh-doubling construction*
vis-ze pikrobs mariami, [rom xink'ali **vin** šeč'ama]?
who-on thinks Mariam.NOM COMP khinkali.NOM **who.ERG** ate
'Who does Mariam think that ate khinkali?'
(lit. 'About who does Mariam think that who ate khinkali?)

(4) *Prolepsis + gap construction*
vis-ze pikrobs mariami, [rom xink'ali __ šeč'ama]?
who-on thinks Mariam.NOM COMP khinkali.NOM ate
'Who does Mariam think that ate khinkali?'
(lit. 'About who does Mariam think that ate khinkali?)

One question that arises is whether sentences like (4) instantiate long-distance *wh*-movement despite the non-canonical form of the *wh*-phrase: is there a movement chain connecting the *wh*-phrase to the gap? If these elements are connected by movement, what explains the morphological form of the *wh*-phrase? If they are not connected by movement, what is the syntax of sentences like (4) and how does it result in the meaning of a long-distance question?

In this paper I argue that there is no long-distance movement in sentences like (4), but that there are movements happening *within* each clause: the proleptic *wh*-phrase moves within the matrix CP, and there is a null operator that moves within the embedded CP. The resulting structure is similar to the structure that has been proposed for parasitic gaps (Nissenbaum 2000). I make a proposal about how it arrives at the meaning of a long-distance question without cross-clausal *wh*-movement.

2 Background on Georgian

Georgian (South Caucasian family) is a language with generally very flexible word order, which is underlyingly SOV (Pochkhua 1962; Aronson 1990; Nash 1995; Harris 2000, a.o.). Borise (2023) has argued that while the verbal phrase is head-final, higher projections in the clause are head-initial, and I will adopt this view in this paper. It has been argued that DPs do not leave the verbal phrase for the purposes of case assignment in Georgian: they are case-licensed in situ (Legate 2008; Nash 2017), so I will be positing no case-driven movement of subjects to Spec,TP or other similar position. The case alignment in a Georgian clause depends on the tense/aspect/mood of the clause, and there are three possible alignments: nominative-accusative, "active" alignment with Agents getting ergative case and Patients getting nominative case, and dative-nominative alignment (Aronson 1990).

The key observation about distribution of *wh*-phrases in simple questions in Georgian is that all *wh*-words must occur immediately before the verb (Harris 1981; Borise 2023), (5). For example, consider the sentences in (6)-(9). While all of these word orders would have been possible if the direct object was a non-*wh* DP, having a *wh*-DP as a direct object places restrictions on possible word orders. $SO_{wh}V$ and $O_{wh}VS$ are acceptable orders, because in these cases the *wh*-direct object immediately precedes the verb. SVO_{wh} and $O_{wh}SV$ are ungrammatical orders, as in these cases the *wh*-phrase does not immediately precede the verb.

(5) **Generalization: word order in *wh*-questions** (Harris 1981; Borise 2023)
All *wh*-phrases must immediately precede the verb. Other phrases of the sentence can either precede or follow the [*wh*+verb] complex.

The question that arises is why would (5) hold: why do *wh*-phrases have to precede the verb, and furthermore be immediately adjacent to it? Is there *wh*-movement in Georgian, and if yes, how does it result in the observed word orders?

(6) ✓SO$_{wh}$V
bebia ras
grandma.NOM what.ACC
alagebda?
clean.IPFT.3SG
'What did grandma clean?'
(Borise 2023:184)

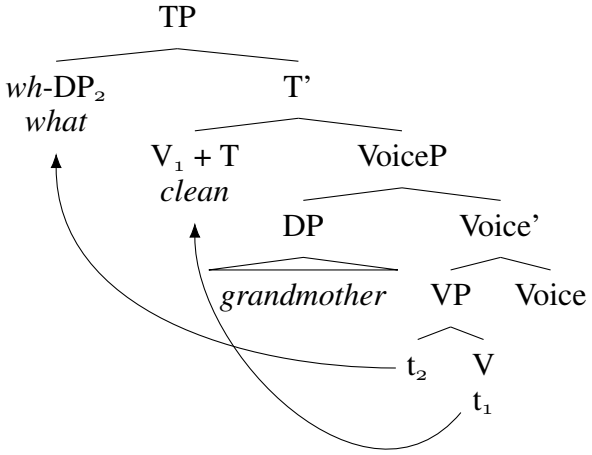
(7) ✗SVO$_{wh}$
*bebia alagebda
grandma.NOM clean.IPFT.3SG
ras?
what.ACC
'What did grandma clean?'
(Borise 2023:185)

(8) ✗O$_{wh}$SV
*ras bebia
what.ACC grandma.NOM
alagebda?
clean.IPFT.3SG
'What did grandma clean?'
(Borise 2023:184)

(9) ✓O$_{wh}$VS
ras alagebda
what.ACC clean.IPFT.3SG
bebia?
grandma.NOM
'What did grandma clean?'

Borise (2023) proposed answers to these questions which I will adopt: she argued that Georgian has *wh*-movement, but it targets a lower position in the clausal spine compared to languages like English and is accompanied by head movement of the verb to the head of the same projection. She argued that the relevant projection is lower than CP and higher than VoiceP. I will assume that this projection is T, though nothing will hinge on this assumption.[3] Thus, (9) will have the syntax in (10).

(10) *Wh-movement in Georgian*



Because the *wh*-phrase moves to the specifier of the same head that the verb movement targets, nothing can intervene between the *wh*-phrase and the verb. The subject, as well as other clausal material, can stay in situ within the VoiceP, as in (10), but can also move outside of it to the left periphery of the clause, giving rise to the generalization in (5). For example, the subject moving to some position above TP creates the SOV order in (6). Borise provides evidence from the order of the [*wh* + verb] complex with respect to adverbs, neg-words and the complementizer to

[3]Borise posits a projection distinct from T that is dedicated to *wh*-movement: WP.

argue for the syntax in (10). Sentences with *wh*-movement in Georgian also observe island effects, illustrated in (11) and (12) with Coordinate Structure Constraint and Relative Clause Island respectively. This supports existence of *wh*-movement.

(11) *šotam [**ra** č'ama] da [ğvino dalia]?
Shota.ERG **what.NOM** ate and wine.NOM drank
'What is the *x* such that Shota ate *x* and drank wine?'

(12) *natiam **<vin>** č'ama xink'ali [romelic **<vin>**
Natia.ERG **<who.ERG>** ate khinkali.NOM REL **<who.ERG>**
gaak'eta]?
made
'Who is *x* such that Natia ate khinkali that *x* made?'

Another phenomenon that Georgian has is *prolepsis*. It is not limited to the *prolepsis + gap* construction that is the focus of this paper, (4), but also occurs outside of sentences with *wh*-dependencies. In (13) we see that there is a proleptic DP with the post-position *-ze* in the matrix clause[4], and the embedded clause can contain either a gap or a pronoun co-referential with it. In most cases a gap or a pronoun is needed for the sentence to be grammatical, but there are certain cases with inanimate proleptic objects when it is possible to not have any obvious gap in the embedded CP. (14) illustrate this with the proleptic DP *amindi* 'weather' and the embedded verb *cvims* 'rain' which takes no arguments.

(13) mariami **givi-ze** pikrobs, [rom am sakmem ___**/is**
Mariam.NOM **Givi-on** thinks COMP this task.ERG **/3SG.NOM**
gaağiziana].
upset
'Mariam thinks about Givi$_1$, that this task upset him$_1$.'
Consultant's Comment: Better with a gap, but the pronoun is acceptable.

(14) keti **amind-ze** pikrobs, [rom ic'vimebs].
Keti.NOM **weather-on** thinks that will.rain
'Keti thinks about the weather that it will rain.'

3 Against direct dependency

I would like to argue that the *prolepsis + gap* construction in (4) does not involve a direct dependency: the proleptic *wh*-DP is base generated in the matrix clause and not in the position where we see the gap. First, consider (15)-(16).

(15) visi tma dadga qalq-ze?
who.GEN hair.NOM stood back.legs-on
'Who got scared?' (lit. 'Whose hair stood up on its back legs?')

[4]Some speakers can also use an ACC-marked DP instead of the one marked with the postposition *-ze*, reporting that these are two interchangeable variants, with the one with *-ze* being the "more correct" one. I will make a simplifying assumption that these are the same: a DP with special case marking assigned by Θ_{About}.

(16) ***No idiomatic readings***
mariami [vis tma-ze] pikrobs,
Mariam.NOM who.GEN hair-on thinks
[rom ___ qalq-ze dadga]?
COMP back.legs-on stood.up
*Idiomatic: 'Who does Mariam think that got scared/angry?'
lit.: 'About whose hair does Mariam think that it stood up on its back legs?'

(15) illustrates that the idiomatic expression *tma dadga qalq-ze* 'got scared' (lit. '(s.o.'s) hair stood up on its back legs') can retain its idiomatic meaning in questions: the possessor of *tma* 'hair' can be a *wh*-phrase, resulting in the question *'Who got scared?'*. (16) shows that *vis tma* 'whose hair' cannot surface as the proleptic object and form the idiomatic expression with the embedded predicate in the *prolepsis + gap* construction. This suggests that *vis tma* 'whose hair' has never been inside of the embedded CP and is base-generated inside of the matrix clause.

(17) and (18) show the absence of reconstruction effects for the proleptic *wh*-DP. In (17) we see that an anaphor inside of the proleptic object can be bound by the matrix subject but not by the embedded subject. In (18) we see that the possessor *mis* cannot be bound by the embedded subject *titoeuli gogo* 'every girl'. These data support the idea that the proleptic DP does not originate in the position of the gap.

(17) ***No reconstruction for anaphor binding***
[romel [tav-is-i tav-is]$_1$/*$_2$ naxat'-ze] pikrobs mariami$_1$,
which self-GEN-NOM self-GEN picture-on thinks Mariam.NOM
[rom givi-m$_2$ ___ šeako]?
COMP Givi-ERG praised
'Which picture of herself$_1$/*himself$_2$ does Mariam$_1$ think that Givi$_2$ praised?'

(18) ***No reconstruction for variable binding***
mariami$_1$ [mis$_{1/*2}$ romel masc'avlebel-ze] pikrobs
Mariam.NOM 3SG.GEN which teacher-on thinks
[rom [titoeuli gogo]$_2$ ___ ec'via]?
COMP each girl visited
'Which teacher of hers$_{1/*2}$ does Mariam$_1$ think that [every girl]$_2$ visited?'

Finally, the proleptic object can contain R-expressions that should induce principle C violation had they been inside of the embedded clause, (19). The grammaticality of (19) suggests that the proleptic *wh*-phrase was never c-commanded by the embedded subject *man*, and is thus not related to the gap via movement.

(19) ***No principle C violations***
mariami [šotas$_1$ romel natesav-eb-ze] pikrobs
Mariam.NOM Shota.GEN which relative-PL-on thinks
[rom man$_1$ ___ šeuracxqopa miağena]?
COMP 3SG.ERG offense make.receive
'Which relatives of Shota$_1$ does Mariam think that he$_1$ offended?'

To sum up, there is no evidence that the proleptic DP has ever been inside of the embedded CP. Lack of idiomatic readings, reconstruction for binding and principle C effects all point to the conclusion that this DP is generated in the matrix CP.

4 Island-sensitivity

Given the evidence against cross-clausal movement we have seen in the previous section, we might expect the gap to be able to occur in an island. However, that is not what we find: embedding the gap in any kind of island leads to ungrammaticality, as is illustrated with three different islands in (20)-(22).

(20) ***Coordinate Structure Constraint***
*__ra-ze__ pikrobs mariami
__what-on__ thinks Mariam.NOM
[rom šota-m ___ č'ama da ǧvino dalia]?
COMP Shota-ERG ate and wine drank
'What is the thing *x* about which Mariam thinks that Shota ate *x* and drank wine?'

(21) ***Relative Clause Island***
*__vis-ze__ pikrobs mariami, [rom natia-m gaak'eta xink'ali,
__who-on__ thinks Mariam.NOM COMP Natia-ERG made khinkali
[romelic ___ č'ama]]?
REL ___ ate
'Who$_1$ does Mariam think that Natia made the khinkali that (s)he$_1$ ate?'

(22) ***Adjunct Island***
*__vis-ze__ pikrobs mariami, [rom bebia-m inerviula
__who-on__ thinks Mariam.NOM COMP grandmother-ERG was.worried
[imitom, rom am sakme-m ___ gaağiziana]]?
because COMP this task-ERG upset
'Who$_1$ does Mariam think that the grandmother was worried because this task upset him/her$_1$?'

Interestingly, it turns out that this is not a property specific to the long-distance *wh*-dependency: structures with non-*wh* proleptic objects also show island-sensitivity. For example, in (23) we see that the gap corresponding to the proleptic DP, which is not a *wh* item, cannot occur inside of an adjunct.

(23) *__givi-ze__$_1$ pikrobs mariam-i, [rom bebia-m inerviula
__givi-on__ thinks Mariam-NOM COMP grandmother-ERG was.worried
[imitom, rom am sakme-m ___$_1$ gaağiziana]].
because COMP this task-ERG upset
'Mariam thinks about Givi$_1$ that the grandmother was worried because this task upset him$_1$.'

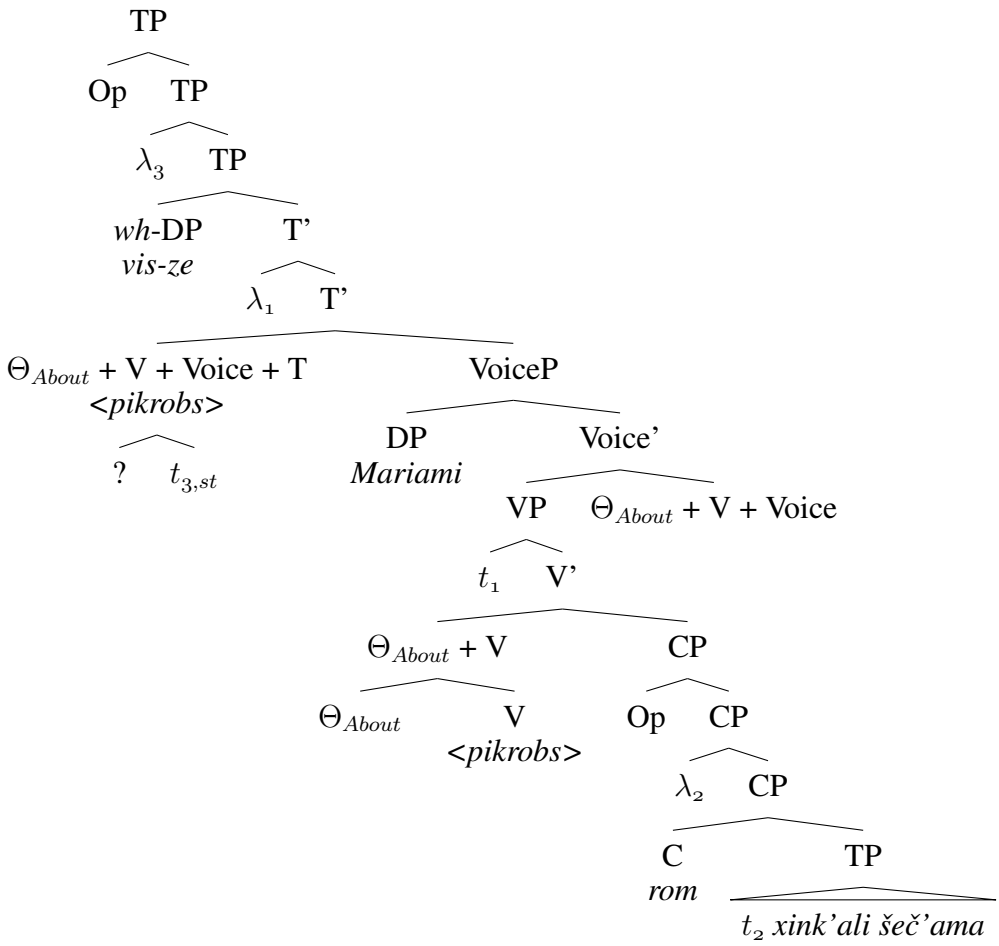
The data above raise a question: how do we reconcile the evidence against base-generation in the embedded clause with the observed locality constraints? Moreover, why is prolepsis island-sensitive? What is the semantics of prolepsis, and how does Georgian create the meanings of long-distance questions with its help? In the next section I suggest answers to these questions.

5 Proposal

I propose that the *prolepsis* + *gap* construction instantiates a structure that is similar to a parasitic gap configuration: the embedded clause is an adjunct to the matrix verb inside of which operator movement takes place. Analyzing the structure in this way will explain both why we see no evidence of the proleptic DP ever being inside of the embedded CP, and also why we observe island effects for the embedded gap. Let us consider the sentence in (24), for which I propose the LF in (25).

(24) **vis-ze** pikrobs mariami, [rom xink'ali __ šeč'ama]?
who-on thinks Mariam.NOM COMP khinkali.NOM ate
'Who does Mariam think that ate khinkali?'
(lit. 'About who does Mariam think that ate khinkali?)

(25)



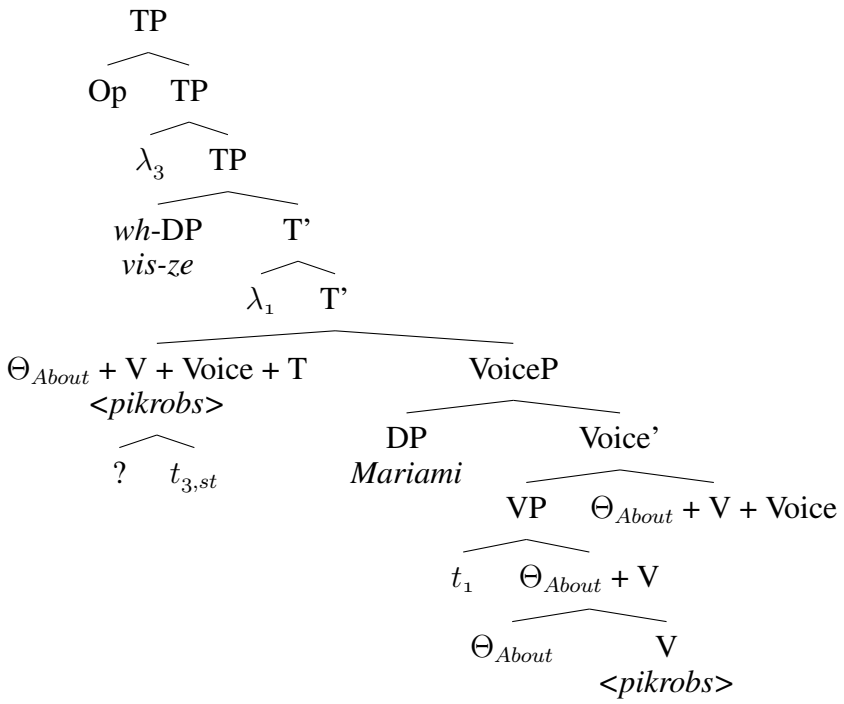
I assume the neo-Davidsonian approach to argument structure (Castañeda 1967, a.o.) according to which all arguments are severed from the verb and are introduced by functional projections. So the verb *pikrobs* 'think' combines with the theta-head Θ_{About}, which introduces the proleptic argument—the individual that the thinking is *about*. I propose that the embedded CP is a modifier of the verb, and it combines

after the Θ_{About} head has combined with the verb, but *before* the ABOUT-argument has been saturated by the proleptic DP (trace t_1 in the tree in (25)).

Let us first consider the syntax and semantics of the matrix clause. I propose that the structure of the matrix clause in the prolepsis + gap construction in (24) is identical to the structure of simple sentences with *wh*-ABOUT-DPs like in (26)-(27). As proposed by Borise (2023), the verb undergoes movement to the head of the same projection that *wh*-movement targets. Here I assume that this projection is TP.

(26) **vis-ze** pikrobs mariami?
who-on thinks Mariam.NOM
'Who does Mariam think about?'

(27)



I assume that the question meaning is created in a standard way (e.g., von Fintel & Heim 1997-2020): the T head hosts a question operator, which combines with a null <s,t>-type operator that undergoes movement and creates abstraction.

Given this syntax, the semantic derivation of (26), and thus also of the matrix clause in (24), will proceed in the following way. The verb, which denotes a predicate of thinking events, (28), will first combine with the Θ_{About} head, (29), resulting in the function that takes an individual as its argument and returns a predicate of thinking events that are about this individual, (30). This individual will be saturated by the trace t_1 of the proleptic *wh*-DP that is undergoing movement.

(28) $[\![\textit{pikrobs}]\!] = \lambda e_v.\text{think}(e)$

(29) $[\![\Theta_{About}]\!] = \lambda f_{vt}.\lambda x_e.\lambda e_v.f(e) \wedge \text{ABOUT}(e)=x$

(30) $[\![\Theta_{About} + V]\!] = \lambda x_e.\lambda e_v.\text{think}(e) \wedge \text{ABOUT}(e)=x$

The question operator in T has the meaning in (31): it takes two propositions and equates them. It first combines with the trace of the null <s,t>-type operator, and then with the VoiceP, equating the two propositions that they denote. The abstraction over the proleptic argument of *think* then gives us the meaning of T' in (32): a predicate of individuals x such that the proposition $g(3, st)$ (the meaning of the trace) equals the proposition *"Mariam thinks about x"*. The *wh*-phrase *vis-ze* 'about who' denotes a generalized existential quantifier over individuals, (33), and takes the meaning of T' as its argument.[5] The abstraction contributed by the null operator movement results in the denotation of the sentence in (34): it is a set of propositions of the form *"Mariam thinks about X"*, where *X* is some human individual.

(31) $[\![?]\!] = \lambda p_{st}.\lambda q_{st}.p=q$

(32) $[\![T']\!] = \lambda x.\ g(3, st) = \lambda w'$.Mariam thinks about x in w'

(33) $[\![\textit{vis-ze}]\!] = \lambda k_{et}.\exists x[human(x) \wedge k(x)]$

(34) $[\![TP]\!] = \lambda p.\ \exists x[human(x) \wedge p = \lambda w'$.Mariam thinks about x in w']

Now let us consider the syntax and semantics of the embedded clause. I adopt Kratzer's semantics for clausal embedding (Kratzer 2006; Kratzer 2013; Kratzer 2016), according to which embedded CPs denote predicates of contentful entities. In particular, I will assume that CPs under consideration are predicates of events (see Bogal-Allbritten 2016; Bogal-Allbritten 2017; Elliott 2020; Bondarenko 2020; Bondarenko 2022, a.o.). I will also assume *equality semantics* of complementation, according to which the relationship established between the embedded proposition and the contentful entity is equality (Moulton 2009; Elliott 2020; Bassi & Bondarenko 2021; Bondarenko 2022; Elliott & Bondarenko 2024). Thus, the complementizer will have the meaning in (35): a function that combines with a proposition and returns the set of eventualities whose content is that proposition.

(35) $[\![\textit{rom}]\!] = \lambda p_{st}.\lambda e_{v}.\textsc{Cont}(e)=p$

I propose that embedded CP in the *prolepsis + gap* construction contains a null operator that is base-generated in the position of the gap and raises to the left periphery of the clause, contributing abstraction over the relevant argument. In our example in (24)-(25), this will be abstraction over the individual who is the Agent of the event of eating khinkali. Thus, the meaning of the CP in this example will be the function of the type <e,<v,t>> in (36): it takes an individual x and an event e as its arguments, and returns true iff the propositional content associated with the event e is the proposition *"x ate khinkali"*.

(36) $[\![CP]\!] = \lambda x_{e}.\lambda e_{v}.\textsc{Cont}(e)=\lambda w'.$ x $ate_{w'}$ khinkali

[5] There is a question of what is the role of the postposition *-ze* 'about' in the syntax and semantics of proleptic DPs. Here I am assuming that its contribution is vacuous in both regards: it does not contribute structure or meaning, but can be thought of as a case assigned by the Θ_{About} + V head to the proleptic argument. But there is a plausible alternative: it could be that *-ze* is actually the exponent of the Θ_{About} head, and this head cliticizes onto the noun and is pronounced together with it. I see no issues with this alternative, and it is compatible with my proposal: as long as it is just the DP that undergoes the movement, my analysis of the construction will remain unaffected.

The key feature of the *prolepsis + gap* construction is how the clause is integrated with the verb. When the verb combines with the Θ_{About} head, the individual argument that the resulting function wants to combine with is not saturated immediately. Instead, the complex Θ_{About} + V head combines first with a CP via Generalized Predicate Modification. The Θ_{About} + V head and the CP are both functions of the type <e,<v,t>>, and so they can compose intersectively by identifying their individual and event arguments. Thus, the denotation of the V' node in (25) is in (37): this is a function that takes an individual x and returns a predicate of events of thinking about x where the content of the thinking is *"x ate khinkali"*.

(37) $[\![\text{V'}]\!] = \lambda\mathbf{x}_e.\lambda e_v.\text{think(e)} \wedge \text{ABOUT(e)=}\mathbf{x} \wedge \text{CONT(e)=}\lambda\text{w'. }\mathbf{x}\text{ ate}_{w'}\text{ khinkali}$

The individual that will saturate this function will bear two theta-roles at once: it will be interpreted both as the ABOUT-argument of the verb and as the AGENT of eating khinkali according to the attitude holder of thinking. In the construction under consideration the proleptic DP is a *wh*-phrase, and so it is its trace that will saturate the function in (37). The *wh*-phrase will move to matrix Spec,TP, and the rest of the derivation will proceed just like in any matrix question, (26)-(27). The whole sentence will have the meaning in (38): it is a set of propositions of the form *"Mariam thinks about X that X ate khinkali"*, where X is some human individual.

(38) $[\![\text{TP}]\!] = \lambda\text{p. }\exists\mathbf{x}[\text{human(x)} \wedge \text{p} = \lambda\text{w'.}\exists\text{e [think(e)}_{w'} \wedge \text{HOLDER(e)=Mariam} \wedge \text{ABOUT(e)=}\mathbf{x} \wedge \text{CONT(e)=}\lambda\text{w''. }\mathbf{x}\text{ ate}_{w''}\text{ khinkali]]}$

Here is how the proposed syntax and semantics account for the properties of the *prolepsis + gap* construction. First, note that even though the individual corresponding to the embedded gap is the same as the individual that the thinking is about, the proleptic DP has never been inside the embedded CP according to my analysis. The coreference between the two individuals is achieved in a different way: by having null operator movement within the embedded clause, together with composing the verb with the CP by Generalized Predicate Modification. The fact that under my analysis the proleptic DP is base-generated in the matrix clause explains why it cannot form an idiom together with the embedded predicate, contain anaphors and pronouns bound by the embedded subject, and also why it can contain R-expressions that are co-referential with the embedded subject.

Second, the locality restrictions that we observed in section 4 follow from the proposed syntax. The gap cannot be embedded in an island because it corresponds to the null operator that undergoes movement to the left periphery of the embedded clause. The proleptic *wh*-DP undergoes movement too, to matrix Spec,TP. Thus, in the proposed structure there are two instances of A-bar movement—within matrix and embedded clauses, but no instance of long-distance *wh*-movement.

The meaning of the *prolepsis + gap* construction that we arrived at is very close to the meaning of a sentence with the long-distance *wh*-movement: the only difference is the presence of the ABOUT-argument. Instead of the set of propositions of the form *"Mariam thinks that X ate khinkali"* we have the set *"Mariam thinks <u>about X</u> that X ate khinkali."*. We have arrived at this set by taking a simple matrix question *"About which individual x does Mariam think?"*, and modifying it by the embedded clause—*"...such that the content of her thoughts is 'x ate khinkali'"*.

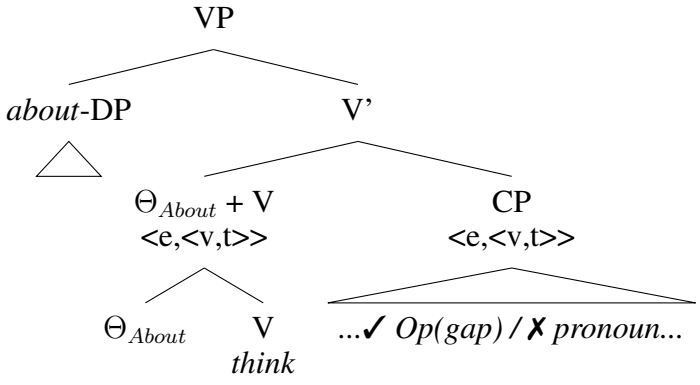
Note that without the verb having an individual argument we could not have created a long-distance dependency. Both the introduction of the individual argument and the embedded clause denoting a predicate of contentful events have been crucial ingredients for creating the indirect *wh*-dependency: they resulted in the complex verbal head having the same semantic type (<e,<v,t>>) as the embedded CP with the abstraction contributed by the operator movement inside of it. Thus, in order for a language to have the *prolepsis + gap* construction like in Georgian, the following conditions need to hold: (i) the matrix verb has to be compatible with introduction of an individual argument (like an ABOUT-argument); (ii) the embedded CP should be able to denote a predicate of contentful eventualities; (iii) the language should have a null operator that can move to the left periphery of the embedded CP.

6 Predictions

My proposal made an assumption that embedded CPs are predicates of contentful events (Bogal-Allbritten 2016; Bogal-Allbritten 2017; Elliott 2020; Bondarenko 2020; Bondarenko 2022, a.o.), and I suggested that being a modifier allows the embedded clause to have more flexibility with respect to when it is introduced into the derivation—e.g., the CP can be merged not with the verb directly, but with the complex Θ_{About} + V head. This raises two questions. First, what happens if the CP modifies the VP instead of the complex head? The semantic type of the clause should allow for such composition where the CP is merged *after* the proleptic argument has been introduced. Second, how does the possibility of a gap vs. a pronoun in structures with prolepsis depend on the attachment site of the embedded clause? When does the CP have to contain a gap? When can it contain a pronoun? Are there cases in which neither a gap nor a pronoun are possible?

Let us first review the case where the CP combines with the Θ_{About} + V head:

(39) *CP modifies the Θ_{About} + V head*



If the CP contained a pronoun co-indexed with the proleptic argument, it would have denoted a predicate of events (<v,t>), as it would lack the abstraction that movement would create. Composing such a CP with the Θ_{About} + V head would lead to a type mismatch.[6] Thus, the prediction is that when the CP attaches to the

[6]Here I am assuming that a principle like Event Identification (Kratzer 1996) does not exist.

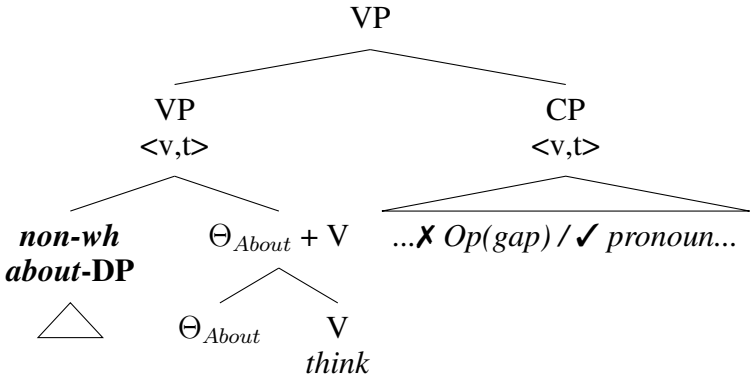
Θ_{About} + V head, the gap in the CP must be obligatory—a co-referential pronoun should not be able to be used.

Now let us see what the predictions are for when the CP attaches to the VP with the proleptic object inside of it. In this case the individual argument introduced by Θ_{About} has been saturated, and the VP thus is of type <v,t>. To successfully combine with it by Predicate Modification, the embedded clause has to also be of type <v,t>. This means that it cannot contain a null operator undergoing movement, and the co-referential gap should be ungrammatical.

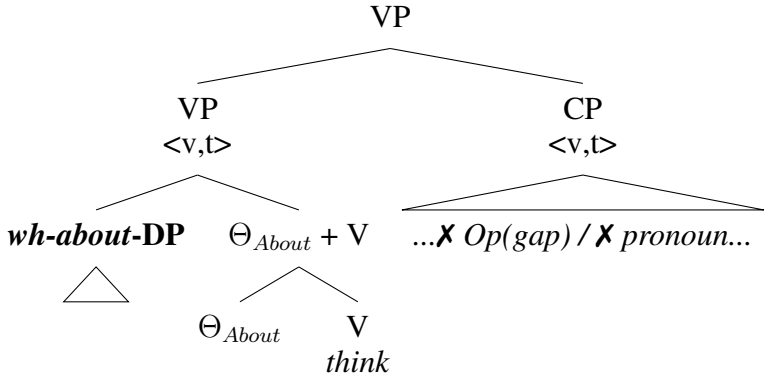
The properties of the proleptic DP seem important for evaluating whether a co-referential pronoun should be grammatical in this structure. I would like to suggest that we expect a co-referential pronoun to be possible if the proleptic DP is not a *wh*-phrase, (41), and to be impossible with a proleptic *wh*-DP, (42). The reason for this distinction has to do with constraints on the positions of traces and pronouns bound by the same quantifier: *wh*-phrases have to move and are subject to the Weak Crossover (WCO) effect, whereas non-*wh*-DPs do not have to move and are not subject to WCO. I will take the descriptive generalization in (40) as the definition of the WCO effect: the trace must c-command the pronoun if they are bound by the same quantifier. Now consider the structures in (41) and (42).

(40) *Weak Crossover (descriptive generalization)*
In a configuration where a pronoun P and a trace T are both bound by a quantifier Q, T must c-command P. (Lasnik & Stowell 1991:690)

(41) *CP modifies VP with non-wh proleptic object*



(42) *CP modifies VP with wh proleptic object*
*



Since (41) contains no quantifier, using inside the embedded clause a pronoun co-referential with the proleptic DP should be possible. Note that the proleptic DP in this case will not be binding the pronoun, as it does not c-command it: this will be just a case of co-reference. In (42) on the other hand the proleptic DP is quantificational, and so it will undergo movement and be subject to the WCO constraint. Because the position of the trace does not c-command into the embedded CP, the prediction is that the use of a co-referential pronoun would violate the WCO constraint. A sentence with the structure in (42) should thus be ungrammatical: neither a pronoun nor a gap would lead to a well-formed syntactic derivation.

As it stands, my proposal also predicts that the embedded clause that modifies the VP could also in principle contain neither a gap nor a pronoun—nothing in the syntax or semantics I proposed requires the presence of a co-referential item. Of course, there might be additional semantic and/or pragmatic constraints on what it means to be an ABOUT-argument, and how such an argument must be related to the content of thinking (see, e.g., Rawlins 2013). But according to my proposal, such restrictions would not be stemming from the details of semantic composition.

To sum-up, here are the predictions that my proposal makes. If the proleptic DP is a *wh*-phrase, then the embedded CP must combine with the Θ_{About} + V head before the proleptic DP saturates the individual argument, and the CP must contain a gap that corresponds to a null operator. If the proleptic DP is not a *wh*-phrase, then the CP could combine either before or after the proleptic DP is merged. In the former case, the CP will have to contain a gap. In the latter case, it will have to contain a pronoun. In other words, we predict that a co-referential pronoun will be optional with proleptic non-*wh*-DPs. Finally, my proposal also in principle allows a CP without any co-referential material to modify a VP that contains a proleptic object. These predictions are borne out, and are illustrated in (43)-(44).

(43) mariami **vis-ze**$_i$ pikrobs, [rom am sakmem ___$_i$**/*is**$_i$
Mariam.NOM **who-on** thinks COMP this task.ERG **/3SG.NOM**
gaağiziana]?
upset
'About who$_i$ does Mariam think that this task upset them$_i$?'

(44) mariami **givi-ze**$_i$ pikrobs, [rom am sakmem ___$_i$**/is**$_i$
Mariam.NOM **Givi-on** thinks COMP this task.ERG **/3SG.NOM**
gaağiziana].
upset
'Mariam thinks about Givi$_i$, that this task upset him$_i$.'

(45) keti amind-ze pikrobs, [rom ic'vimebs].
Keti.NOM weather-on thinks COMP will.rain
'Keti thinks about the weather that it will rain.'

In (43) we see that sentences with the proleptic *wh*-DP require a gap: the use of a co-referential pronoun is ungrammatical. (44) and (45) are repeated examples in (13)-(14) from section 2. They show that proleptic DP that is not a *wh*-phrase can

occur with either a gap or a pronoun in the embedded clause, and that sometimes the embedded clause lacks a co-referential DP altogether. I would note though that sentences like (45) are quite rare: in most cases we do see a pronoun or a gap, and animate proleptic objects for example seem to require the presence of a co-referential expression. I have to leave a more detailed investigation of conditions under which CPs like (45) are acceptable for future research.

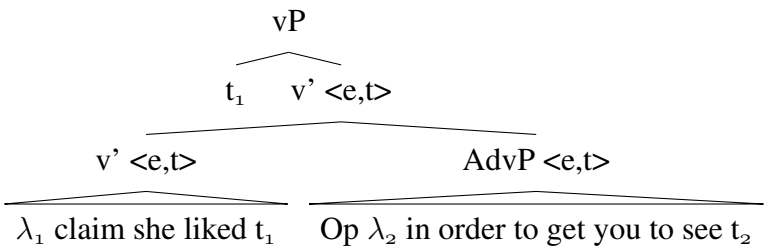
7 Concluding remarks

In his book on reconstruction and resumption in indirect A-bar dependencies, Salzmann (2017:314) writes: *"One...in principle expects instances of prolepsis that occur with a gap and display island sensitivity. However, such cases are not attested...Whether this points towards a fundamental property of the construction in need of explanation or just an accidental lexical gap is unclear because sufficient information about prolepsis is currently only available for rather few languages."*

In this paper I argued that Georgian fills this gap: it has island-sensitive structures with a proleptic object and a gap in the embedded clause, which corresponds to a null operator undergoing movement. I proposed that Georgian, which lacks true long-distance *wh*-movement, uses this kind of structure for creating long-distance *wh*-dependencies: the proleptic *wh*-phrase in the matrix clause binds the argument that the null operator abstracts over, creating meanings like *"About which X does Mariam think that X ate khinkali?"* My analysis of the Georgian construction is thus very similar to Nissenbaum's (2000) proposal for parasitic gaps, (46)-(47).

(46) [What movies]$_1$ did Mary [claim she liked t_1 [in order to get you to see PG$_1$]]?

(47) *vP in structure with a PG*



The view that embedded CPs can be predicates of events allows us to unify the two phenomena: both constructions instantiate a case where a modifier (embedded CP/adverbial CP) successfully intervenes between a function (Θ_{About} + V, v' node after abstraction) and the argument that saturates it (the trace of the moving DP).

Acknowledgments

I thank the members of CLS 60, Syntax Square at MIT, The South Caucasian Chalk Circle-3, and the WOLF-lab at Harvard for valuable feedback. I am extremely grateful to my Georgian consultants for their intuitions and insightful comments: Liza Davitadze, Luka Edzgveradze, Tamar Korkotashvili, Anastasia Leladze, Saba Lepsveridze, Otar Tchitchinadze, and Nene Zhvania. All errors are my own.

References

Aronson, H. I. 1990. *Georgian: A Reading Grammar*. Slavica.

Bassi, I., & T. Bondarenko. 2021. Composing CPs: Evidence from disjunction and conjunction. In *Semantics and Linguistic Theory*, volume 30, 583–602.

Bogal-Allbritten, E. 2016. *Building meaning in Navajo*. University of Massachusetts Amherst dissertation.

Bogal-Allbritten, E. 2017. Basic pieces, complex meanings: Building attitudes in Navajo and beyond. Invited talk at the Workshop on polysemy and coercion of clause embedding predicates. 39th Annual Meeting of DFfS. Saarbrucken, Germany.

Bondarenko, T. 2020. Factivity from pre-existence: Evidence from Barguzin Buryat. *Glossa: a journal of general linguistics* 5. 1–35.

Bondarenko, T. 2022. *Anatomy of an Attitude*. Massachusetts Institute of Technology dissertation.

Borise, L. 2023. The syntax of wh-phrases, narrow foci, and neg-words in Georgian. *The Linguistic Review* 40. 173–215.

Castañeda, H.-N. 1967. Comments on Donald Davidson's "The logical form of action sentences". In *The logic of decision and action*, ed. by N. Resher, 104–112. Pittsburgh, PA: University of Pittsburgh Press.

Elliott, P., & T. Bondarenko, 2024. On the monotonicity of attitudes: NPIs and clausal embedding.

Elliott, P. D. 2020. *Elements of Clausal Embedding*. UCL (University College London) dissertation.

von Fintel, K., & I. Heim, 1997-2020. Intensional Semantics. MIT.

Harris, A. 1981. *Georgian syntax: A study in relational grammar (Cambridge Studies in Linguistics)*, volume 33. Cambridge: Cambridge University Press.

Harris, A. 2000. Word order harmonies and word order change in Georgian. *Amsterdam Studies in the Theory and History of Linguistics Science Series* 4. 133–164.

Kratzer, A. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, 109–137. Springer.

Kratzer, A., 2006. Decomposing Attitude verbs. Talk given at the Hebrew University of Jerusalem.

Kratzer, A., 2013. Modality and the semantics of embedding. Talk given at the Amsterdam Colloquium.

Kratzer, A., 2016. Evidential moods in Attitude and Speech reports. Talk given at the 1st Syncart Workshop.

Lasnik, H., & T. Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22. 687–720.

Legate, J. A. 2008. Morphological and abstract case. *Linguistic Inquiry* 39. 55–101.

Moulton, K. 2009. *Natural Selection and the Syntax of Clausal Complementation*. UMass Amherst dissertation.

Nash, L. 1995. *Portée argumentale et marquage casuel dans les langues SOV et dans les langues ergatives: l'exemple du géorgien*. Paris 8 dissertation.

Nash, L. 2017. The structural source of split ergativity and ergative case in Georgian. In *The Oxford Handbook of Ergativity*, ed. by J. Coon, D. Massam, & L. Travis, 175–200. Oxford: Oxford University Press.

Nissenbaum, J. 2000. *Investigations of covert phrase movement*. Massachusetts Institute of Technology dissertation.

Pochkhua, B. 1962. Sit'q'vatganlagebisatvis kartulši ("about the word order in georgian"). *Iberiulk'avk'asiuri enatmecniereba* 13. 119–121.

Rawlins, K. 2013. About 'about'. In *Semantics and Linguistic Theory 23*, ed. by N. Snider, 336–357.

Salzmann, M. 2017. *Reconstruction and Resumption in Indirect A'-Dependencies: On the syntax of prolepsis and relativization in (Swiss) German and beyond*, volume 117. Walter de Gruyter GmbH & Co KG.