# Chapter 1: Introduction: Rare Wings are Universal Forelimbs

## 1. The challenge of rare constructions

As typology and generative linguistics have advanced, it has become clear that there exist crosslinguistically rare constructions. Some grammatical features are salient aspects of one language but have no obvious analog in most other languages of the world. This poses a challenge for those of us who think that universal grammar (UG) is something real and important. The key question is whether or not UG specifies the possibility and core properties of the rare construction. On the one hand, if one thinks that it does not, then a significant aspect of the language under study is unconstrained by UG, which is not a very satisfying view within the paradigm. In the limit, one could start to wonder what UG is for if it does not guide the acquisition of important aspects of particular languages. On the other hand, if one thinks that UG does specify the possibility of the rare construction, then one needs to face the possibility that there is a chapter of UG that most languages do not read at all. Moreover, if there are many rare constructions attested in the human languages taken as a whole, then the book of UG contains many chapters that most languages read only one or two of. Although I admit that there are long books on my shelves which I have only read a chapter or two of (some of which I am glad to have), this is also not very satisfying as a view about UG.

This issue can be illustrated with a series of examples, which introduce the specific topics that this book is concerned with. Consider switch-reference (SR). SR can be defined as a construction in which an affix or particle at the periphery of an embedded clause signals whether the subject of the embedded clause refers to the same entity as the subject of the main clause (same subject, SS) or to a different entity (different subject, DS). (1) is a pair of examples from Shipibo, a Panoan language spoken in the Peruvian Amazonian. The suffix *-ax* on the embedded verb signals SS, whereas *-tian* signals DS (Valenzuela 2003, Baker and Camargo Souza 2020).

(1) Shipibo (Baker and Camargo Sousa 2020).

*a. José=ra [<pro> Rosa oin-**ax**] xobo-n ka-ke.*
José=EV he Rosa see-SS.ABS house-LOC go-PFV
"He$_i$ seeing Rosa$_j$, José$_i$ went home."

*b. [José-kan Rosa oin-ke-**tian**]=ra, (ja) xobo-n ka-ke.*
José-ERG Rosa see-PFV-DS=EV 3SG home-LOC go-PFV
"When José$_i$ saw Rosa$_j$, he/she$_k$ went home."

The term "switch-reference" was coined by Jacobsen (1967). It evoked a flurry of typological interest in the 1970s and 1980s, and the first major generative analysis was Finer's (1984, 1985). It is not a super rare phenomenon. It is found in quite a few languages of Western North America (about 70 in McKenzie's (2015) survey), in a substantial region of Australia (Austin 1981), and in New Guinea (Foley and Van Valin 1984). It is found in a good number of South American languages as well (see Baker & Camargo Souza 2020 for a partial list). But even if there are hundreds of languages that have SR, there are more languages that do not have it than that do have it. Indeed, the phenomenon is absent from large areas of the world, including Europe, Africa, mainland Asia, and Eastern North America. SR is particularly interesting in that it is often a big deal in languages that have it. An average page of text in Shipibo may have a dozen or more examples of SS/DS marking, and it is used in complex and expressive ways, which are seen only in sentences consisting of more than one clause. A theory of UG that says nothing about SR seems to be missing a key part of the genius of Shipibo. Furthermore, the literature on SR strongly suggests that its essentials are quite stable in the languages that have it. SR in Shipibo and Quechua is not that different from SR in Washo and Mojave in North America, or from SR in Diyari and Jiwarli in Australia. If we looked only at those languages that have SR, it would seem to provide a great case for UG. Nevertheless, the majority of languages do not have an SR system, so that it does not feel at all "universal" in the ordinary sense.[1]

Another example of a crosslinguistically rare phenomenon is

---

[1] There is, however, no contradiction between this and what Chomsky has always meant by Universal Grammar, which is the innate knowledge that a child brings to bear on learning their native language. Chomskyan universal grammar never refers to surface features that are claimed to be present in all languages.

logophoric pronouns. These are special pronouns that are used in an embedded clause to refer to a designated argument of the matrix clause—usually its subject. (2a) is an example from Ibibio, a Niger-Congo language spoken in Nigeria. The pronoun *ímọ̀* cannot be used in a root clause, and in (2a) it must refer to the main clause subject *Okon*. In contrast, the ordinary pronoun *anye* in (2b) can be used freely in matrix clauses, and in embedded clauses it can refer to the object of the matrix verb 'tell' or to someone else salient in the context, as well as to the matrix subject, just like pronouns in English.

(2) Ibibio (fieldwork, Willie Willie)

*a. Okon á-ké-dọ̀kḳ́ Edem ké Emem í-maá-ghá* ***ímọ̀***.
Okon 3SG-PST-tell Edem that Emem 3.LOG-like-NEG LOG
"Okon$_i$ told Edem$_k$ that Emem does not like him$_{i,*k,*n}$."

*b. Okon á-ké-dọ̀kḳ́ Edem ké Emem í-maá-ghá* ***anye***.
Okon 3SG-PST-tell Edem that Emem 3.SG-like-NEG 3SG
"Okon$_i$ told Edem$_k$ that Emem does not like him$_{i,k,n}$."

This term "logophoric pronoun" was coined by Hagège (1974). The first landmark generative study was Clements (1975). Cully (1994) provides a well-known typological overview. Such pronouns are a reasonably salient feature in the languages that have them. There will not be a dozen examples on most pages of a text, but there will be examples in an average short story. Crosslinguistically, this phenomenon has an even narrower distribution than SR. As far as is known, using a relatively narrow definition, it may be found only in West Africa (Cully 1994). Given this, one might not be so tempted to use UG for a phenomenon like this. However, it is robust in the sense that it found in many languages in this region, and it is attested in languages from different families (e.g., languages from the Chadic branch of Afroasiatic as well as Niger Congo languages) and with different typological characters (Yoruba is isolating, whereas Ibibio is more agglutinative). So it is not very satisfying just to say that logophoric pronouns are an idiosyncratic ornament to core language that happened to develop only once or twice because of special historical circumstances. It has also long been thought that logophoric pronouns are similar to the more widespread phenomenon of long distance anaphors in the languages of Europe and East Asia (Clements 1975, (Sells 1987), etc.), a hypothesis that I consider in Chapter 5

A third rare construction is upward complementizer agreement. This can be defined as a construction in which the complementizer of the

clausal complement of a verb agrees in person/number/gender features with another argument of that verb—almost always the subject. (3) exemplifies this with a pair from the Bantu language Kinande spoken in the Eastern Congo.

(3) Kinande (fieldwork, Philip Mutaka)

*a. Kámbale mw-a-kabw-ir-a abá-kalí a-ti Maryá mw-á-gúl-ir-é ehí-lole.*
CL1.Kambale AFF-CL1.TNS-told-APPL-FV CL2-women CL1-that CL1.Mary AFF-CL1.TNS-buy-ASP-FV CL19-bananas
"Kambale told the women that Mary bought bananas."

*b. Aba-kali mo-ba-kabw-ir-a Kambale b**a-ti** Maryá mw-á-gúl-ir-é ehí-lole.*
CL2-women AFF-CL2.TNS-told-APPL-FV CL1.Kambale CL2-that CL1.Mary AFF-CL1.TNS-buy-ASP-FV CL19-bananas
"The women told Kambale that Mary bought bananas."

This is mentioned in Baker (2008), among other places. The first full-fledged generative study is Diercks (2010, 2013). This construction is not as prominent an overall design feature for languages that have it as SR is for Shipibo. However, it is arguably a special case of what is a general design feature for the Bantu languages: the fact that they are particularly rich in agreement, extending it to a wide variety of functional heads. Like logophoricity, upward C agreement is known primarily from one area of the world—a smattering of languages from Nigeria and Angola in the west to Kenya in the east. There are languages that are very rich in agreement in other parts of the world, but most do not have this kind of agreement.[2] Furthermore, several of the languages seem to have developed C agreement independently: the agreeing C comes from 'say' in Kinande but from 'be' in Lubukusu; Kipsigis is a Nilo-Saharan language, not a Bantu language (Diercks and Rao 2019); Angolan languages have developed their agreeing Cs from pronouns rather than verbs (Kawasha 2007).[3] But despite these

---

[2] The Amerindian language Kanienkeha (Mohawk) is a good example of a language from another part of the world in which agreement is ubiquitous on all types of lexical categories, but Cs do not bear agreement.

[3] It has been debated for particular languages (Lubukusu, Kipsigis) whether a given element is synchronically a complementizer or an inflected form of the verb 'say'. I enter into this discussion this in Chapter 2.

historical differences, this range of languages seem to have developed essentially the same construction. That is the sort of convergence that could be explained by having UG shape the space of what an agreeing C-like particle can be like despite superficial differences in the input. However, reports of this phenomenon in languages from outside this region are very sporadic. The one reasonably well-documented case is the Teiwa language of Indonesia (Sauerland, Hollebrandse et al. 2020); other possible candidates are the New Guinean language Arapesh (Baker 2008: 182-183) and the Peruvian language Nanti (Michael 2008: 111-112). But should UG cover something that is so sparsely attested around the globe?

Other constructions that I consider in this work have other kinds of distributions across languages. A relatively well-discussed one is indexical shift. This can be characterized as a construction in which first person pronouns inside the CP complements of certain verbs can be interpreted as referring to an argument of the selecting verb—usually its subject. (4) is an example from Magahi, an Indo-Aryan language spoken in Eastern India (Alok and Baker 2018, Alok 2020).[4]

(4) Magahi (fieldwork, Deepak Alok)

*a. Santee-aa soch h-ai ki **(ham)** tej h-i.*
Santee-FM think be-3.NH.S that (I) smart be-1.S
"Santee$_i$ thinks that he/I$_{i,sp*}$ am smart."

*b. Santee-aa Bantee-aa-ke kahl-ai ki **(ham)** tej h-i.*
Santee-FM Bantee-FM-DAT tell:PFV-3.NH.S that (I) smart be-1.S
"Santee$_i$ told Bantee$_k$ that he/I$_{i,sp*,*k}$ am smart.'

Indexical shift was brought to the attention of generative linguists as something distinct from direct quotation by Schlenker (1999, 2003); see Deal (2020) for a thorough recent overview. Its crosslinguistic distribution is not yet clear, as the tests that distinguish it from direct quotation need to be more widely applied. It is turning up in more and

---

[4] In addition to showing possible coreference relationships by subscripts (i, k, n…) on pronouns and noun phrases in the usual way, I use the special subscript sp* to show possible reference to the speaker of the sentence as a whole and the special subscript ad* to show possible reference to the addressee as a whole. For participant pronouns that can get shifted readings, I use glosses like "she/I" and "him/you" to give a sense of the two possible meanings that the pronoun has.

more languages, including a wide range of Turkic languages, Zazaki and Kurmanji, Amharic, Nez Perce, and Matses, as well as Magahi. Still there are probably fewer languages that have this phenomenon than that do not. It is not known in the languages of Europe, and it is not possible in Niger-Congo languages like Ibibio, Kinande, and Lubukusu. This does not seem to be as strongly an areal phenomenon as logophoric pronouns and upward C agreement are. Rather indexical shift languages seem to be sprinkled throughout the world—although this picture could change some once we have thoroughly distinguished indexical shift from direct quotation.

The last of the five core constructions that I focus on in this work is allocutive agreement. This can be characterized as a verb bearing some kind of agreement, not (only) with its subject or object, but with features of the person that the sentence is addressed to. The paradigm case is Basque, with Oyharçabal (1993) an early generative landmark. (5) is an example, where the auxiliary verb is different depending on whether the sentence is addressed to someone with whom one has a formal relationship ((5a)) or a male person with whom one has a close relationship ((5b)), or a female person with whom one has a close relationship ((5c)).

(5) Basque (Oyharçabal 1993)

*a. Pette-k lan egin di-**zü**.*
Peter-ERG work do AUX.3.ERG-2SG.H.AL
"Peter worked." (to a person with a distant relationship)

*b. Pette-k lan egin di-**k**.*
Peter-ERG work do AUX.3.ERG-2SG.M.AL
"Peter worked." (to a close male)

*c. Pette-k lan egin di-**n**.*
Peter-ERG work do AUX.3.ERG-2SG.F.AL
"Peter worked." (to a close female)

This is a relatively isolated phenomenon in Europe (although see Alok & Haddican (2022) on possible allocutivity in Gallician). More recently, it has been turning up in a range of South Asian languages as well, both Dravidian (e.g., Tamil) and Indo-Aryan (e.g., Magahi). Miyagawa (2012, 2017) argues that it is present in Japanese, although it is less clear that the relevant verbal morphology qualifies as agreement in this language. Antonov's (2015) typological study also lists Pumé (Venezuelan), Nambikwara (Brazilian), Mandan (Siouan),

and Beja (Cushitic). This gives the impression of a phenomenon very thinly sprinkled through the languages of the world. As always, it makes a difference here how one defines the phenomenon, since a larger set of languages have politeness markings of various kinds (e.g., Korean) that could be akin to allocutive marking. But as a coherent phenomenon of allocutive *agreement*, it seems very likely that only a small percentage of languages of the world have it.

An admittedly strong/provocative view that helps underline what is at stake here is the so-called (strong) uniformity hypothesis, sometimes taken to be part of Chomsky's Minimalist Program (Chomsky 2001: 2, Sigurdsson 2004, Miyagawa 2012). This is the hypothesis that there are essentially no syntactic parameters, but only a universal grammar plus morphological variation at PF. According to this view, all languages have essentially the same syntax. In some important domains, that may be true, for example when it comes to the basic stock of categories (especially lexical categories; see Baker 2003) and the principles by which they are merged together to make larger phrases. Similarly, it may not be implausible to think that all languages have *wh*-movement of one form or another (Huang 1982, many others). One can also make a case for simple case marking and agreement being universal, covert if not overt, as in Vergnaud's famous GB-era hypothesis (Chomsky 1981, Vergnaud 2008, although I personally have been led away from this; see, for example, Diercks 2012, Baker 2015). But whatever one thinks of these classic cases, the strong uniformity hypothesis begins to strain the imagination when it comes to the rarer constructions that I have sampled from here. Taken literally, it would amount to saying that all languages have a rich SR system like Shipibo, along with a system of logophoric pronouns like Ibibio, along with upward C-agreement like Lubukusu—not to mention things like noun incorporation and serial verb constructions. This makes one's head spin to think about seriously, and all the more so as one exposes oneself to more of the rare constructions that are attested in languages of the world.

# 2. An analogy from anatomy

To shed some light on how we might think about the paradoxes that rare constructions create for notions of UG, consider an analogy taken from biology. The analogy focuses on the comparative anatomy of mammals. I believe that most mammals do not have wings. I, for example, do not (sadly). In fact, only the various species of bats have

them. Furthermore, most mammals do not have flippers for swimming. But whales and seals do. Arms are also a rare thing for mammals to have. But we humans have them (compensating somewhat for the lack of wings). Finally, horses have front legs, whereas these other types of mammals do not. So there are rare anatomical features of mammals, as there are rare syntactic constructions in natural language.

However, forelimbs are not at all a rare thing for mammals to have. On the contrary, *all* mammals have them: bats, seals, whales, humans, horses, mice, and so on. Furthermore, we have learned that mammal forearms all have essentially the same syntactic structure. By this I mean that they have the same number of bones, which are connected to one another in the same ways. This common structure then takes on strikingly different functions in different animals. There is thus a theory of "universal mammal" (UM) skeletons that applies just as well to bats and seals as to humans and horses. The wings of a bat are not counterexamples to this theory of UM, nor do they fall outside its scope. On the contrary, they are good exemplars of the theory, when they are properly understood.

It could be that something similar holds for certain kinds of rare syntactic constructions. Suppose that the same syntactic "skeleton" underlies all five of the phenomena I inventoried in §1.1, as well as a few more. Upward complementizer ageement is undeniably a rare feature of language, as is switch-reference. But suppose that they are both manifestations of a structure in which a complementizer licenses a null nominal which is controlled by the matrix subject and with which the complementizer enters into an Agree relationship. (This is unpacked below.) Then that underlying structure is not as rare or geographically restricted as the two surface constructions are when considered separately. Indeed, if there are other surface realizations of this underlying structure, it might even be universal in a meaningful sense. That is the top-level thesis of this book.[5]

# 3. Some motivating similarities

[5] Note however that my claim is that the underlying structure is relatively common and widespread, not that it is clearly attested in every language.

Of course, a unifying view such as this one is more plausible if the phenomena being unified have something readily observable in common, even before we start dissecting them in earnest. Mammalian forelimbs do have such similarities. Whether wings or flippers or front feet or arms, there are two of them, and they are symmetrically placed protruding from the upper/front torso of the animal. The constructions that I have listed here are arguably also in the same ballpark when viewed in the right way.

For starters, all five of the constructions listed in §1.1 are what I call "funny things complementizers do to relate to noun phrases around them." This description holds most obviously of the upward C agreement seen in (3): one funny thing that Cs can do to relate to NPs around them is agree with them. It also holds fairly straightforwardly for switch-reference constructions like (1), once one recognizes that that the outermost affix on the verb in a head-final language like Shipibo is often the realization of a C-like head. Then SS and DS markers can be taken to be complementizers that "relate funnily" to the subject NPs just above and below them by indicating whether they are coreferential or not. Allocutive marking also falls under this description if one believes, following Speas and Tenny (2003), that the person to which a sentence is addressed can be represented syntactically by a null pronoun in the periphery of the sentence (see also Oyharcabal 1993 for an earlier although less general version). Then allocutive marking can be seen as the result of C relating to this special NP near to it by agreeing with it.

Logophoric pronouns and indexical shift are less obvious cases of complementizers relating the NPs around then in funny ways. After all, no complementizer or clause-peripheral morpheme was mentioned in the preliminary characterization of those two constructions. But the notion of a complement clause was referred to in the characterizations of these constructions, and complement clauses have complementizers. Indeed, it turns out that in languages like Ibibio which complementizer is present partially determines whether logophoric pronouns are possible in the complement clause or not.[6] For example, the normal declarative complementizer *ke* does allow a

[6] The connection between a particular complementizer and the possibility of logophoric pronouns is even closer in Ewe, according to Clements (1975: 157): logophoric pronouns are possible only inside a CP headed by the complementizer *be*. See Chapter 5 for discussion.

logophoric pronoun inside the complement clause to be used to refer to the matrix subject, as in (6a), but the complementizer *naña* used with perception complements does not, as in (6b).

(6) Ibibio (fieldwork, Willie Willie)

*a. Okon a-maa-kit **ke** Emem a-maa-yip ebot ímọ̀.*
Okon 3SG-PST-see that Emem 3SG-PST-steal goat LOG
"Okon$_i$ saw that Emem stole his$_i$ goat."

*b. Okon a-maa-kit **naña** Emem a-yip ebot ọmọ/*imọ.*
Okon 3SG-PST-see how Emem 3SG-steal goat 3SG/*LOG
"Okon$_i$ saw Emem steal(ing) his$_i$ goat."

This then does qualify as a funny thing that complementizers do to relate to the NPs around them: the complementizer *ke* in Ibibo enables a special pronoun in its complement to refer to the subject above *ke*.

Indexical shift is similar. For example, in Mishar Tatar null first person pronouns can refer to the matrix subject (which need not be first person) when the embedded clause is headed by the verbal ("say") complementizer *diep*, as in (7a), but not with other forms of complementation, as in (7b) (Podobryaev 2014; see also Shklovsky and Sudo 2014 on related Uyghur).

(7) Mishar Tatar (Podobryaev 2014: 84, 89)

*a. Alsu [pro kaja kit-te-m **diep**] at'-tɤ?*
Alsu (1SG) where go.out-PST-1SG that say-PST.3SG
"Which place did Alsu$_i$ say that he/I$_{i,sp*}$ went?"

*b. Marat Alsu-ga [(minem) kil-gän-em-ne] at'-tɤ.*
Marat Alsu-DAT 1SG.GEN come-NMLZ-1SG-ACC say-PST.3SG
"Marat$_i$ said to Alsu that I$_{*i,sp*}$ came."

It is not outlandish, then, to think that these five constructions might be theoretically related, just as it is not outlandish to think that the forelimbs of different mammals might be theoretically related.

A more specific property that the five constructions have in common that encourages the idea that they are related is that they center on CPs in complement positions. For example, one can have an agreeing C in addition to an ordinary C in a complement clause in Ibibio ((8b)) but not in (say) a relative clause ((8a)).

(8) Ibibio (fieldwork, Willie Willie)

*a. Okon a-sʌk a-yem [ngwet odo [(***a-bo /*a-te**) se [(ami) ng-k-i-nọ Enọ]]].*

Okon 3SG-PROG 3SG-seek book the 3SG-C/3SG-C REL I 1SG-PST-1SG-give Eno

"Okon is looking for the book that I gave to Eno."

*b. Okon á-ké-n-dọ̀kkọ́ [(a-bo/a-te) ké [Emem í-maá-ghá Enọ]].*

Okon 3SG-PST-1SG.O-tell **3SG-C/3SG-C** that Emen 3SG-like-NEG Eno

"Okon told me that Emem does not like Eno."

Similarly, logophoric pronouns which refer to a c-commanding subject in a higher clause are licensed in complement clauses in Ibibio, but not normally in relative clauses, as shown in (9).

(9) Ibibio (fieldwork, Willie Willie)

*a. *Okon a-maa-duọk ngwet odo se imọ i-k-i-dep.*

Okon 3SG-PST-lose book the REL LOG 3.LOG-PST-3.LOG-buy

("Okon$_i$ lost the book that he$_i$ bought." OK with … *anye a-ke-dep* 'he 3SG-PST-buy')

*b. Okon a-maa-dọkkọ ayín ọ̀mọ̀ ke ímọ̀ i-ma-i-yip ngwet.*

Okon 3.SG-PST-tell son his that LOG 3.LOG-PST-3.LOG-steal book

"Okon$_i$ told his$_{i,k}$ son that he$_i$ stole the book."

In the same way, shifted 'I' can refer to the matrix subject when it is in a complement clause in Magahi, but not when it is in a relative clause, as shown in (10).

(10) Magahi (fieldwork, Deepak Alok)

*a. Santee kalpanaa kark-ai [ki ego sudar laiki hamraa-se biaah kart-ai].*

Santee imagine do-3.NH.S that one.CL beautiful girl me-INS marriage do-3.NH.S

"Santee$_i$ imagines that a beautiful girl will marry me$_{i,sp*}$."

*b. Santee, ego sudar laiki je hamraa-se biaah kart-ai, okra baare-me kalpanaa kark-ai.*

Santee one beautiful girl REL me-INS marry do-3.NH.S her

about-LOC imagine do-3.NH.S
"Santee$_i$ imagined a beautiful girl who will marry me$_{*i,sp*}$."

As for switch-reference, distinctive same subject marking is allowed on a small range of complement clauses in Shipibo, as in (11). However, it is never allowed on relative clauses. Hence no SS marker like *-ax* or *-i/kin* is seen in (12), but only the invariant nonfinite affixes *-a* (perfective), *-ai* (imperfective), and *-ti* (infinitive) (Valenzuela 2003). In (12b) the understood subject of both the matrix clause and the relative clause are the same and in (12a) the two subjects are different, but this has no effect on the verb morphology.

(11) Shipibo (Fieldwork)
*Maria-nin=ra [kenti be-**kin**] peo-ke.*
Maria-ERG=EV pot bring- SS.ERG start.TR-PFV
"Maria started to bring the pots."

(12) Shipibo (Valenzuela 2003: 477, 471)
*a. Jain-xon-ribi=ra a-kan-ti iki, [kawin [jan-n yawa raka-n-**ti**]]*
there-ERG-also=EV do.TR-3PL-INF AUX rush.mat 3-LOC peccary.ABS lying-TR-INF
"Also they have to make [rush mats [(for people) to place the peccaries on]]."

*b. [Ja [wexa-anan-**a**]joni rabé] raká-kan-ai westiora oxé.*
that cut-RECP-PFV.PTCP man two lying-3PL-IPFV one moon
"[The two men [who had cut each other]] lay for one month."

Finally, allocutive marking is allowed on relative clauses in Magahi, as in all finite clauses. However, *shifted* allocutive marking, where the agreement on the embedded verb reflects the relationship of the agent of the matrix clause to the goal of the matrix clause rather than that of the speaker to the addressee, is not allowed in relative clauses ((13a)). In contrast, this sort of shifted allocutive marking is allowed in CP complements in Magahi ((13b)).

(13) Magahi (fieldwork, Deepak Alok)
*a. Santee-aa <u>baabaa-ke</u> khabar kahk-au je Ram okraa kahk-**au/*o.***
Santee-FM grandfather-DAT word tell-3.NH.S:NH.AL REL Ram 3SG.ACC tell-NH.AL/*H.AL
"Santee$_i$ told grandfather the message that Ram told him$_i$." (said to a peer) (Bad is honorific /o/ reflecting Santee's relationship to his grandfather.)

*b. Santee-aa baabaa-ke kahk-au ki Ram Sita-ke dekhl-au/o hal.*
Santee-FM grandfather-DAT told-3.NH.S:NH.AL that Ram Sita-ACC saw-NH.AL/H.AL was
"Santee told grandfather that Ram saw Sita." (said to a peer) (OK is honorific /o/ reflecting Santee's relationship to his grandfather.)

So all five constructions show a systematic contrast between complement clauses, which are directly related to the main verb, and relative clauses that modify the object, which are only indirectly related to the main verb.

Another property that these five constructions have in common is that they take place in fully finite CP complements, not nominalized clauses. For indexical shift, this has been pointed out by Shklovsky and Sudo (2014) for Uyghur (see also Mishar Tatar in (7) above; it is also true for Magahi). In (14b) the complement of 'say' includes the nominalizing affix *-lik*, bears accusative case, and its subject has genitive case. These are all signs of nominalization and are all differences between (14b) and (14a). Correlated with this, (14b) does not allow a shifted indexical reading of 'I', whereas (14a) does.

(14) Uyghur (Shklovsky and Sudo 2014: 383)
*a. Ahmet [men ket-tim] di-di.*
Ahmet 1SG leave-PST.1SG say-PST.3
'Ahmet$_i$ said that he/I$_{i,*sp*}$ left.'

*b. Ahmet [mening kit-ken-lik-im-ni] di-di.*
Ahmet 1.SG.GEN leave-REL-NMLZ-1SG-ACC say-PST.3
"Ahmet$_i$ said that I$_{*i,sp*}$ left."

Similarly, in Ibibio logophoric pronouns are possible in CP complements like (15b) but not in true derived nominalizations like (15a). Note that the "subject" of the derived nominal in (15a) follows the nominalized verb, as possessors of nouns do in this language, whereas the subject of the clause in (15b) precedes the verb in accordance with the language's basic SVO word order.[7]

[7] However, logophoric pronouns are possible in gerundival complements in Ibibio, which have a mix of nominal and verbal properties; see Chapter 5. I do not go into how nominal is too nominal for these rare constructions here.

(15) Ibibio (fieldwork, Willie Willie)

*a. Okon i-kit-te [n-dudue eka ọmọ/*imọ].*
Okon 3SG-see-NEG NMLZ-commit.fault mother his/*LOG
"Okon$_i$ does not see his$_i$ mother's mistake/fault."

*b. Okon i-kit-te [ke eka imọ a-maa-due].*
Okon 3SG-see-NEG that mother LOG 3SG-PST-commit.fault
"Okon$_i$ does not see that his$_i$ mother committed a fault."

In addition, there is no possibility of adding an upward agreeing C to a nominalized/gerund-like constituent in Ibibio, whereas these can be added to normal CPs. This contrast is shown in (16).

(16) Ibibio (fieldwork, Willie Willie)

*a. Okon a-maa-tre [(*a-bo) u-koot ngwet].*
Okon 3SG-PST-stop (*3SG-C GER-read book
"Okon stopped reading the book."

*b. Okon á-ké-n-dọ̀kkọ́ [(a-bo/a-te) ké [Emem í-maá-ghá Enọ]].*
Okon 3SG-PST-1.SG.O-tell 3SG-C/3SG-C that Emen 3SG-like-NEG Eno
"Okon told me that Emem does not like Eno."

Similarly, (17) shows that there is no allocutive marking on nominalized (infinitive or participial) complements in Magahi, analogous to what can be seen in complement clauses in examples like (13b) (Alok 2021). (In this case the allocutive marking on the complement clause would resume the marking that is possible on the matrix verb—one of the two options attested in (13b)—since there is no goal argument of the matrix verb to support a shifted version.)

(17) Magahi (Fieldwork, Deepak Alok)

*a. Santee-aa [jaa-yel-(*au/*o/*ain) chaha h-au/o/ain.*
Santee-FM go-INF-(NH.AL/H.AL/HH.AL) want be.3.NH-NH.AL/H.AL/HH.AL
"Santee wants to go."

*b. Ham okraa dhekhe-se bach-l-i-**ain/o/au.***
I 3SG.NH.ACC see.PTCP-INS avoid-PFV-1SG-HH.AL/H.AL/NH.AL
"I avoided seeing him."

Finally for switch-reference, Imbabura Quechua presents a relevant minimal contrast. It has SR markers on subjunctive complement

clauses, but not on nominalized ones, as in (18). (18a) has a nominalized complement of the verb ‘want’, as shown by the fact that the complement as a whole is marked for accusative case whereas its object does not need to be marked accusative (Cole 1983, Cole 1985, Hermon 1985). This version does not have SR marking. In contrast, (18b) has a nonnominalized subjunctive complement, as shown by the fact that the complement as a whole does not bear accusative case and the object inside it must be accusative. This version is marked for SR; in (18a) the suffix is SS *-ngapaj* rather than DS *-chun* because the wanter is the same as the desired seer (see (20b) for a DS analog). There is a systematic complementarity in this: no clause type is marked for both SR and morphological case in Quechua.

(18) Imbabura Quechua (Hermon 1985: 25; Cole 1985: 37)

*a. [Aycha-(ta) miku-na-ta] muna-ni.*
meat-(ACC) eat-NMLZ-ACC want-1.S
“I want to eat meat.”

*b. Muna-y-man [ñuka mama-ta riku-ngapaj].*
want-1.S-COND my mother-ACC see-SBJV.SS
“I want that I see my mother; I want to see my mother.”

So all five phenomena are things that happen in nonnominal CPs rather than in nominalized constituents.

One other important property that four of the five constructions have in common is that they are subject-oriented rather than object-oriented in ways that may seem surprising from a theoretical perspective. For example, when the verb selecting a CP with an agreeing C has an object as well as a subject, the C agrees with the subject, not the object, even though the object seems to be structurally closer to C. This is shown in (19) from Lubukusu (see also (3) from Kinande). Previous research on upward C-agreement like Diercks (2013) and Carstens (2016) wrestles with how to explain this.

**(19)** Lubukusu (Diercks 2013)

*a.* ***Baba-ndu*** *ba-bol-el-a Alfredi* ***ba****-li a-kha-khil-e.*
CL2-people CL2.S-say-APPL-FV CL1.Alfred CL2-that CL1.S-FUT-conquer-FV
“The people told Alfred that he will win.”

*b.* ***Alfredi*** *ka-bol-el-a baba-ndu* ***a****-li ba-kha-khil-e.*
CL1.Alfred CL1.S-say-APPL-FV CL2-people CL1-that

CL2.S-FUT-conquer-FV
"Alfred told the people that they will win."

Similarly, SR marking on a complement clause shows whether the embedded subject is coreferential with the subject of the matrix clause, not with the object of the matrix clause, even though the object seems to be closer to the SR head in the complement clause. This can be seen in (20) from Imbabura Quechua. (20a) in which the lower subject is coreferential with the matrix subject has the so-called SS (or proximate) marker *ngapaj*, but (20b) in which the lower subject is coreferential with the matrix object does not; it must have the so-called DS (or obviative) marker *-chun* instead. This is a disanalogy between SS-marked clauses and controlled infinitives that Hermon (1985: 122-124) struggles with.

(20) Imbabura Quechua (Cole 1985: 37; Hermon 1985: 123)

*a. Muna-y-man [ñuka mama-ta riku-ngapaj].*
want-1.S-COND my mother-ACC see-SBJV.SS
"I want that I see my mother; I want to see my mother."

*b. Juan-da kunvinsi-rka-ni [(pay) Kitu-man ri-**chun**].*
Juan-ACC persuade-PST-1.S he.NOM Quito-to go-SBJV.DS
"I persuaded Juan that he go to Quito." (not **ri-ngapaj)*

One can also envision a connection here with the fact that logophoric pronouns are subject-oriented in that they can refer to the subject of the matrix clause, but not to the object of the matrix clause, as seen in (21) from Ibibio. Ordinary third person pronouns in languages like English are not required to refer to a subject in this way.

(21) Ibibio (fieldwork, Willie Willie)

*Okon á-ké-dọ̀kḳọ́ Edem ké Emem í-maá-ghá ímọ̀.*
Okon 3SG-PST-tell Edem that Emem 3SG-like-NEG LOG
"Okon$_{i}$ told Edem$_{k}$ that Emem does not like him$_{i,*k}$."

Similarly, a first person indexical pronoun 'I' in Magahi can refer to the matrix subject of a verb like 'tell', but not to the matrix goal.

(22) Magahi (fieldwork, Deepak Alok)

*Santee-aa Banteeaa-ke kahl-ai ki ham tej h-i.*
Santee-FM Bantee-FM-DAT told-3.NH that I intelligent be-1
'Santee$_{i}$ told Bantee$_{k}$ that he/I$_{i,*k,sp*}$ is/am intelligent.'

The odd construction out for this last similarity is allocutive marking.

This is object-oriented—anti-subject-oriented—rather than subject-oriented, in the sense that allocutive marking on the embedded verb can change when the status of the person referred to by the indirect object changes, as shown in (23). (All three sentences can be said to the same addressee, say a peer of the speaker.)

(23) Magahi (Fieldwork, Deepak Alok)

*a. Santee-aa **Bantee-aa-ke** kahk-ai ki Ram-ke Sita-se baat kareke chah-**au**.*
Santee-FM Bantee-FM-DAT told-3.NH.S that Ram-DAT Sita-INS talk do.INF should-NH.AL
"Santee told Bantee that Ram should talk with Sita."

*b. Santee-aa **baabaa-ke** kahk-ai ki Ram Sita-ke dekhl-**o** ha-l.*
Santee-FM grandfather-DAT told-3.NH.S that Ram Sita-ACC saw-H.AL be-PFV
"Santee told grandfather that Ram saw Sita."

*c. Santee-aa **profesar** saaheb-ke kahk-ai ki Ram Sita-ke dekhl-**ain** ha-l.*
Santee-FM professor HH-DAT told-3.NH.S that Ram Sita-ACC saw-HH.AL be-PFV
"Santee told the professor that Ram saw Sita."

So four of the five "funny things that Cs do to relate to the NPs around them" are subject-oriented in ways that may not be theoretically expected, whereas the fifth is anti-subject-oriented.

There are, then, some readily accessible similarities between the five constructions that render it not implausible that they would have a partially unified analysis. They might be different functions or adaptations of the same underlying structural "skeleton".

# 4. Overview of the analytic framework

I now sketch upfront my high-level hypothesis about what this shared structural skeleton consists of, which provides the unifying threads that tie together the various topics discussed in this work.

The first crucial assumption is that some Cs (or better some C-like heads in an articulated left periphery, after Rizzi 1997) license a pronoun-like DP in the periphery of the clause headed by C. These DPs are rather hard to detect by conventional means in that they are obligatorily phonologically null, they have minimal lexical-semantic

content, and they do not create islands for the extraction of material out of the CP (in contrast to *wh*-phrases in the CP periphery, which can create *wh*-islands). Because they are hard to detect by these means, I often refer to them as "ghostly DPs", giving the discussion a bit of color. I also refer to them as operators, having in mind null operators in the syntactic sense, like the null DPs that undergo *wh*-movement in post-Chomsky (1977) analyses of many constructions (although the ghostly DPs posited here are base-generated in the CP periphery, rather than arriving there by movement). All five constructions involve such ghostly DPs in the C-space, I claim. This is at the root of why the five constructions are possible in CP complements but not in derived nominals or nominalized clauses: nominal constituents do not have the C-like heads that license these ghostly DPs. At the same time, some of the differences between the constructions can be attributed to exactly which head in the C-space licenses the ghostly DP, and to exactly what formal features the ghostly DP has. The ghostly DPs involved in indexical shift and allocutive marking have first and second person features; the ghostly DPs involved in logophoric constructions have a language-particular [+log] feature; the ghostly DPs involved in so-called indexiphor constructions (Deal 2020) have both first person and [+log] features; the ghostly DPs involved in simple upward C-agreement and switch-reference have no intrinsic features of their own.

The second crucial assumption is that any of these ghostly operators may, and in many cases must, undergo obligatory control. The result of this is that they are bound by a designated argument of the verb that selects the CP as its complement. As such, the ghostly operators are somewhat analogous to PRO, the necessarily null DP that is licensed by infinitival Tense in English and many other languages (see Landau 2013 for an overview). This assumption helps to explain why the five phenomena happen in complement clauses but not (for example) in relative clauses: the extra nominal projections that intervene between CP and the matrix verb in a relative clause construction block obligatory control, causing structures with some of the ghostly operators to crash (if the DP does not have interpretable features of its own) and others to show different antecedent-taking behavior (as in the classical distinction between obligatory and nonobligatory control). I also claim that obligatory control is what causes four of the five constructions to be subject-oriented as outlined in (19)-(22): the subject and not the object can control the ghostly DP operator because of a condition governing which what thematic role the controller must

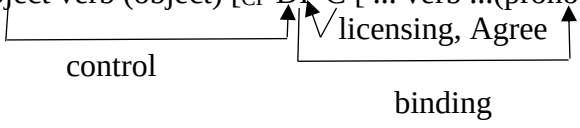
have in order to match the thematic role of the controllee—a property of obligatory control that can be seen for ordinary PRO in the phenomenon of "control shift."

The third crucial assumption, which almost goes without saying, is that the ghostly DP operators can bind pronouns inside the TP complement of the C that licenses them, as long as the pronoun is compatible with the ghostly DP in features. This can happen in any of the constructions, but it is most noticeable in the logophoric pronoun constructions and the indexical shift constructions, because logophoric pronouns and indexicals are specified as needing to be bound by a ghostly DP operator with the proper features. A variant of this has a ghostly DP that is an A-binder rather than an A-bar binder, resulting in long-distance anaphor constructions like the one found in Japanese. Another variant allows the ghostly DP to add features to its bindee, resulting in indexiphors/monstrous agreement constructions like the ones found in Telugu (Messick 2023) or Donno Sɔ (Deal 2020). I also argue that certain kinds of crossover effects can emerge in logophoric constructions, analogous to those that appear when a *wh*-operator binds pronouns and variables of different kinds.

Finally, I assume that a C can enter into Agree with a ghostly operator in the CP periphery—either the very same C-head that licenses the ghostly operator, or a nearby one. An obvious result of this in some languages is that phi-features that the ghostly DP has intrinsically or gets from its controller are transferred to C. This happens in upward C-agreement constructions and in allocutive constructions. A less obvious result of this is that some languages create a pointer from C to the ghostly DP ("Agree-Link" in the terms of Arregi and Nevins 2012) and via the ghostly DP to its controller. This is an ingredient in some switch-reference constructions, I claim.

The basic skeletal template for all these constructions, then, is summarized in (24), following in essence Koopman & Sportiche's (1989) approach to logophoricity in Abe.

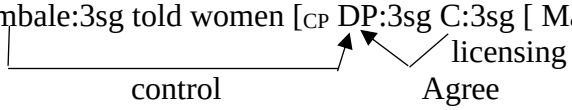
(24) [Subject verb (object) [$_{CP}$ DP C [ ... verb ...(pronoun) ... ]]]




The building blocks of this template—DP-licensing, control, binding, and Agree—are all provided by UG, analogous to the way that "universal mammal anatomy" provides a template of bones which

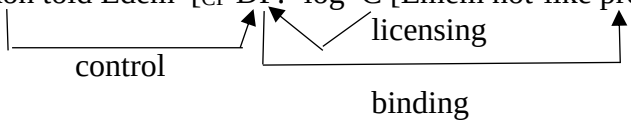
have become expanded or contracted in different mammals to facilitate different functions. For example, upward C-agreement involves licensing a featureless ghostly DP, control of that DP by the matrix subject, and Agree holding between C and the ghostly DP. (The ghostly DP may also bind a pronoun in the CP complement, but nothing special happens with that.) This is sketched in (25).

(25) [Kambale:3sg told women [$_{CP}$ DP:3sg C:3sg [ Maria buy...]]]



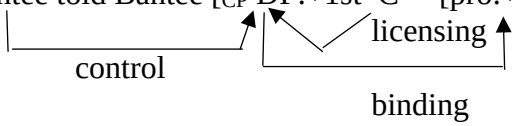

The logophoric pronoun construction involves C licensing a [+log] DP, which is controlled by the matrix subject. C does not agree with this DP, but the DP can bind special [+log] pronouns in its c-command domain. Indeed, [+log] pronouns must be bound by this operator.

(26) [Okon told Edem [$_{CP}$ DP:+log C [Emem not-like pro:+log]]]




Similarly, the indexical shift construction involves licensing a [+1] DP in the CP of an embedded clause, which is also controlled by the matrix subject. Again, C does not usually agree with this ghostly DP, but the DP can bind [+1] pronouns inside its c-command domain. Indeed, [+1] pronouns must be bound by such an operator—by an instance in the root clause if not by one in the embedded clause. (Note that complete phi-feature matching is not required for an argument of the matrix clause to control a ghostly DP operator.)

(27) [Santee told Bantee [$_{CP}$ DP:+1st C [pro:+1st be smart]]]




Variants of this construction result in indexiphoric constructions in languages like Telugu and long-distance anaphoric constructions in languages like Japanese. Finally, switch-reference marking on complement clauses in languages like Imbabura Quechua, Washo, and Choctaw involve licensing a featureless ghostly DP which is controlled by the matrix subject. C then undergoes Agree-Link but not

Agree-Copy with both the ghostly DP operator and the highest DP inside its c-command domain (the embedded subject). These Agree-links are interpreted as coreference (or disjoint reference) at LF, resulting in a same subject (or different subject) construction.

(28) [ I persuade Juan [$_{CP}$ DP C [ he go to Quito]]]

licensing

control Agree-link (2x)

disjoint, so DS

I must clarify, however, that this theory is proposed only for SR in complement clauses in the minority of languages that allow that. For languages that allow SR only adjunct clauses, I do not posit a null DP operator, but make use of direct Agree between the C of the adjunct clause and an NP argument in the matrix clause, as in Baker & Camargo Souza (2020), Arregi & Hanink (2022), and Clem (2023). Overall, switch-reference constructions are only partially integrated with the other rare constructions discussed in this work.

One important further assumption for the overall framework is that the C-space in languages of the world can license pairs of ghostly DP operators, not just solitary ones; the CP-periphery can contain an object-like null DP as well as a subject-like one. Whereas the thematic-role-matching condition on obligatory control implies that only the matrix subject can control the subject-like ghostly DPs that I have presented so far, the same condition implies that only the matrix object can control these additional object-like ghostly DPs. This assumption allows me to incorporate allocutive agreement in Magahi into the account. I claim that Magahi has two ghostly DPs in the periphery of finite CPs, a subject-like one that is [+1] and an object-like one that is [+2]. Shifted allocutive agreement arises when C licenses the [+2] DP, the matrix object controls it, and a head near C agrees with it. This is sketched in (29) as a representation for (23c).

(29) [Santee told prof:HH [DP$_1$:+1 DP$_2$:+2,HH C:+2,HH [ Ram saw ...]]]

Agree

Control

The benefits of including object-like ghostly DPs go well beyond allocutive agreement. The same [+2] operator in Magahi also accounts for the indexical shift of a second person pronoun like ‘you’, such that it refers to the goal object of a matrix verb like ‘tell’ in (30). Here the

pronoun ‘you’ matches $DP_2$ in features (+2, NH) and is bound by it. $DP_2$ is in turn controlled by indirect object of ‘tell’, so ‘you’ ends up referring Bantee.

(30) Magahi (Fieldwork, Deepak Alok)
*Santee-aa **Bantee-aa-ke** kahk-ain ki Ram **toraa** dekhl-i-au hal.*
Santee-FM Bantee-FM-DAT told.3.NH.S-HH.AL that Ram you.NH.ACC saw-1.S-NH.AL be.PFV
“Santee$_i$ told Bantee$_k$ that Ram saw him/you$_k$.” (said to a teacher)

Similarly, a few African languages have special addressee pronouns that appear in embedded clauses and refer to the goal of the matrix verb in addition to logophoric pronouns that appear in embedded clauses and refer to the subject of the matrix verb. Mupun is one such language, as shown in (31). Here the addressee pronoun *gwar* can only refer to the goal of the saying event denoted by the matrix verb, just as the logophoric pronoun *ɗì* can only refer to the agent of that event. These addressee pronouns can be analyzed as pronouns that need to be bound by a [+log] object-like operator ($DP_2$ in a structure like (29), but with different features) that exists alongside the [+log] subject-like operator already discussed ($DP_1$ in a structure like (29)).

(31) Mupun (Frajzyngier 1993: 125)
*Datar sat n-Dapus nə ɗì naa la reep gwar.*
Datar say P-Dapus that LOG see ASP? girl ADDR
“Datar$_i$ told Dapus$_k$ that he$_i$ saw his$_k$ daughter.”

Moreover, one African language, Kipsigis, may have double upward agreement on C: C can agree with both the matrix subject, as in Lubukusu, and the matrix object (Diercks & Rao 2019).[8] This possibility follows assuming that Kipsigis also has a pair of ghostly DP operators and C in this language can undergo multiple Agree, collecting phi-features from both of these operators.

[8] Here I follow Diercks & Rao’s (2019) analysis. In contrast, Dreimel & Kouneli (in press) argue that *lɛ* is still synchronically the verb ‘say’ in examples like (32). See Chapter 2 for some discussion.

(32) Kipsigis (Diercks & Rao 2019)

*a. Ko-i-mwaa- an* ***i-lɛ-ndʒan*** *ko-Ø-ɪt laɣok.*
PST-2SG.S-tell-1SG.O 2SG-C-1SG PST-3SG-arrive children
"You (SG) DID tell me that the children arrived."

*b. Ko-i-mwaa- tʃi* ***a-lɛ-ndʒi*** *ko-Ø-ɪt laɣok*
PST-1SG.S-tell-3.O 1SG-C-3 PST-3.SG-arrive children
"I DID tell him/her/them that the children arrived."

So most of the constructions have subject-oriented and object-oriented versions, although the object-oriented ones are generally much rarer and may depend on the subject-oriented version being present as well.[9]

# 5. Methodological remarks

## 5.1. Language selection and sources

The design of this investigation and the way it is carried out is not intended to be especially innovative, but rather a further example of how generative linguistics has been fruitfully pursued for years. However, some comments on two general issues may be of some use: how I picked the languages I focus on and my sources for those languages, and how I see the relationship between syntax and semantics in this domain.

First on languages and sources. For the most part, this is not the kind of topic that can be pursued by pulling out descriptive grammars on a wide range of languages and seeing what they say about these topics. In part, that is because many languages do not have any of the five constructions, and fewer still will have more than one. In addition, these are phenomena that take place primarily in embedded clauses, hence only in sentences with a relatively high degree of complexity. Such complex sentences are often described rather incompletely in

[9] The one construction that I do not have an object-oriented version of is switch-reference. It is conceivable that the super-rare subject=object construction found in the Panoan language Amahuaca (Clem 2023) is an instance of this, but I do not investigate that here. Another possibility is that the fact that SR alone among these constructions happens freely on adjunct clauses inhibits the development of an object-oriented version. See Chapter 3 for discussion of a possible subject-oriented analog of allocutive agreement in Dargwa.

descriptive grammars. To carry out this kind of project, then, one needs access to sophisticated native speakers of languages that have the relevant constructions (sophisticated in that they are able to concentrate on examples of a certain complexity, not in that they have been indoctrinated into a particular theory). Fortunately, for all but one of the five constructions outlined in this chapter, I have had the opportunity to work with native speaker linguists in a detailed fashion over the course of several years, being informed by what I have learned about one construction in one language when asking questions about another construction in another language. Therefore, each subtopic has a strong home base in a particular language—one that I hope is typical and/or particularly revealing concerning the topic in question. The primary languages and sources are as follows:

1. Upward C-agreement: Ibibio (Willie Willie), Lubukusu (Justin Sikuku), Kinande (Philip Mutaka)
2. Allocutive agreement: Magahi (Deepak Alok)
3. Indexical shift: Magahi (Deepak Alok) (also a little Sakha, Nadya Vinokurova)
4. Logophoric pronouns: Ibibio (Willie Willie), Yoruba (Oluseye Adesola), a little Edo (O.T. Stewart)
5. Switch reference: Yawanawa (Livia Camargo Souza*), Shipibo (three consultants in Lima, fall 2012).

A pattern in this is that the people above have almost all been students or postdocs at Rutgers University, where it has been my privilege to work with them closely in various capacities. It is not for nothing, then, that this book is dedicated to my students; it would have been impossible to pursue this project without their involvement and inspiration. The one person on this list who is not a native speaker of the relevant language is Livia Camargo Souza, but she has worked on Yawanawa for many years (starting before she came to Rutgers), made many trips to the Amazon to study it, and developed relationships where she could ask speakers questions remotely when needed. Switch-reference is also a special case in that the SR languages that are most relevant to this study are ones in which SR marking is found on complement clauses as well as on adjunct clauses. That is true in a limited respect in Yawanawa and Shipibo, which will help us to break into the topic in an effective way, but once

that is done I will be more dependent on secondary sources for this topic than for the others. In addition, there are two constructions that are related to logophoricity and indexical shift that I have not discussed separately in this overview for which I have had similar access to linguists who are native-speakers of strategic languages. These are the indexiphoric/monstrous agreement construction of Telugu (Sreekar Rogatham) and the long-distance anaphor construction of Japanese (Shiori Ikawa). Even if the grand unification that I am attempting is not deemed successful, I hope that my individual descriptions and analyses will be found to have value.

Two of these languages are particularly strategic in having not just one of the rare constructions but two of them. As such, they provide very special opportunities to see how the constructions compare and interact within the same language. One of these is Magahi, which has both allocutive agreement and indexical shift. These was serendipitous, and the two constructions turn out to be deeply interdependent, making use of the same pair of ghostly DP operators, as mentioned above. The other is Ibibio, which has both upward C-agreement and logophoric pronouns. This fact motivated me to reconnect with Willie Udo Willie via the network of Ken Safir's Afranaph project. In this case, the two constructions turn out to be largely independent of each other, although parallel, using different ghostly DP operators at different levels of the CP space. Magahi and Ibibio thus constituent the empirical heart of the book.

Beyond these languages where there has been the opportunity to do significant primary research, the most useful sources have been articles or book (chapters) which study one of the constructions in some language in significant detail. In each case, I compare the results that I have achieved with my collaborators with others reported in the literature on the various construction, focusing on "classics" from the literature (e.g. Ewe for logophoric pronouns; Amharic, Zazaki, and Nez Perce for indexical shift; Basque for allocutive marking, etc.). This comparison gives some sense of what is and is not stable about a particular construction across the languages that are known to have it. However, the "sampling" is not very systematic or complete. My emphasis has been to consider cases that are relatively well-described and analyzed rather than pursuing typological or areal balance. I imagine a unifying project proceeding at the following three levels:

- Studying construction A in language A´.

- Studying construction A in languages A′′, A′′′, A*, ….
- Studying construction A in language A′, B in B′, C in C′, …

Indeed, there is some material at all three levels in this work. But the emphasis is on the first level and the third level, with investigation at the second level being more haphazard and opportunistic. The hope here is that the languages that I can focus on in depth are fairly representative of the phenomena in question, either by blind luck or (better) because UG is such that almost any instance of one of these phenomena is a good instance of the phenomenon. If so, then not much will be lost in moving fairly quickly—although not immediately—from the first level to the third level of study. In proceeding this way, I am assuming that if one spends all one's time and energy on the first level and the second, one might miss something interesting and important on the third level of study. That is a mistake that I am not willing to make, even if it increases somewhat my risk of making other mistakes.

## 5.2. Syntax and semantics

I also need to say something about how I see the relationship of syntax and semantics in this domain (and indeed in most domains). The five constructions that I have mentioned have gotten uneven amounts of attention from specialists in the two subfields. Some of them have been discussed more by syntacticians than by semanticists. That is particularly true for the two constructions in which agreement plays a prominent role: upward C-agreement and allocutive agreement. Other topics have been discussed more by semanticists than by syntactians. That is particularly true for the two constructions that feature bound pronouns: indexical shift and logophoricity (including LD anaphors). At least one of these topics, switch-reference, has been the subject of a turf war between syntactically-oriented approaches (e.g. Finer 1984, 1985) and semantically-oriented approaches (Sterling 1993, McKenzie 2012). I take this diversity of approach to be part of the "fun" of working in this area.

One basic assumption that is relevant to staying oriented in this is that no one group owns any of these topics. We should not say that syntacticians own upward C-agreement, or that semanticists own indexical shift. It is true almost by definition that any construction in any natural language has both a syntactic structure and a semantic interpretation. Therefore, it can be studied from both perspectives. In

particular cases, one kind of analysis may be relatively trivial while the other kind is deep and illuminating. But we will in general only find that out by pursuing both kinds of analysis for every construction, and what looks trivial from within one frame of reference can prove to be deep and illuminating in another frame of reference.

Not only does every natural language construction have a syntax and a semantics, the two are often parallel to one another to a large extent. When that is the case, a descriptive generalization, or even an explanation, can be stated in either terms. It is still, in a significant sense, the same generalization/explanation, given the homomorphic relationship between the two subdisciplines in this domain. Sometimes the generalization may make more sense conceptually when couched in one set of terms rather than the other, such that one has a stronger sense of why it holds. That is great when it happens, but I take it to be a secondary priority, less urgent than finding new generalizations and explanations stated in whatever terms. We can also expect there to be some cases in which a syntactic account of a construction and a semantic account are complementary, not homomorphic, where one can explain some features of the construction about which the other is relatively silent. That is great too. Occasionally there will be actual conflicts, where a syntactic account and a semantic account are genuinely different, and one works better than the other. But I take that to be a less common situation, to be resolved by doing both our syntax and semantics better. In the ultimate scheme of things, every natural language construction must have both a syntax and a semantics, and they must be consistent with one another, given what a language is.

If that is the right lay of the land, how do I position myself within it? The answer is: as an unabashed syntactician reasonably (but not perfectly) literate in the semantic issues. Therefore, in this investigation I lead with the syntax and with generalizations and (partial) explanations phrased in the syntactic idiom. That is what I have been trained to do, and what I profess to be reasonably good at. There is a general flow in this work from the more syntactically-studied topics (Chapter 2 on upward C-agreement, Chapter 3 on allocutive agreement) toward the more semantically-studied topics (Chapter 4 on indexical shift, Chapter 5 on logophoricity), as I start in my comfort zone and see where that leads. (Although, somewhat ironically, I claim that the ghostly DP that C agrees with in cases of upward agreement makes a relatively clear semantic contribution, not considered carefully in the literature.) In pursuing this study, I want to

emphasize the similarities and interconnections across the five constructions, and a bit beyond. Suppose, then, that I start with constructions thought to be syntactic, give them a syntactic analysis, and go on to argue that constructions thought to be semantic are similar to them in important ways, such that they should be explained using the same theoretical resources or risk failing to capture a generalization. Is this being imperialistic for syntax in a way that other researchers should feel threatened or offended by? Hopefully not. Keep in mind that in asserting a generalization or explanation in syntactic terms, I do not mean to be denying that there is also a generalization or explanation in semantic terms that may be parallel to the syntactic version, that may be even deeper than the syntactic version, or that may overlap with it and cover things that the syntactic version does not. We should all spin the best yarns we can, working out how they fit together as we go.

I also do not in general present any worked out formal semantics for the syntactic structures that I propose for the various constructions (until Chapter 8, which has a rather different character). That might seem quaint of me in the 2020s—almost like I was trained in the 1980s. However, my experience has been that it is never impossible to provide a reasonable semantics that goes with a reasonable syntax for a certain construction. One does what one has to do, and formal semantics is not such a tightly constrained enterprise that it cannot find a way, as I understand the situation. To me, it seems fairly clear what the syntactic representations I present are supposed to mean in informal terms; hopefully it will be clear to others too. Perhaps someone will even feel motivated to work out the semantic details at some point. But me doing that would be like a dog challenging a dolphin to a swimming contest: it would be possible, probably, for the dog to get to the other end of the pool, but it would be relatively slow, clumsy, and not the dog doing what it does best. (The dog has the wrong sort of forelimbs!) I take my strengths to be syntax, collecting data on interactions from less studied languages, and constructing a big picture. So I concentrate on doing that, rather than on details about the syntax-semantics interface in this domain.

# 6. What’s where: midlevel results

The basic organization of the rest of the book flows out of what I have said so far. The next six chapters are each organized around one of the crosslinguistically rare constructions that has been introduced in this chapter, with a general movement from constructions which have been

more discussed in the syntax literature toward constructions that have been discussed as much or more in the semantics literature. Chapter 2 takes up the topic of upward C-agreement in African languages. Chapter 3 considers allocutive agreement, which can function as another form of upward C-agreement in Magahi, given that (unlike Northern dialects of Basque) allocutivity is freely marked on complement clauses as well as root clauses. Chapter 4 turns to indexical shift, focusing again on Magahi. This is a natural next step given that allocutive agreement and indexical shift interact in Magahi in systematic ways. Chapter 5 discusses logophoric pronouns, with an emphasis on Ibibio—another natural step, given that logophoric pronouns in CP complements in West African languages refer to the matrix subject in much the same way that shifted first person indexicals do in Magahi and other languages with indexical shift. This chapter also compares logophoric pronouns in Magahi with long distance reflexives in Japanese, which have often been said to be in some sense logophoric elements. Chapter 6 extends the discussion to indexiphoric constructions, mentioned but not exemplified in this chapter so far. These are constructions in which what seems to be a logophor or long-distance anaphor triggers first person agreement on the embedded verb. As such, it combines elements of indexical shift and logophoricity in what looks like a hybrid form. (35) is an example from the Dravidian language Telugu.

(33) Telugu (Messick 2023)
*Raju [**tanu** parigett-ää-**nu** ani] cepp-ää-Du.*
Raju 3SG run-PST-1SG that say-PST-3.M.SG
"Raju$_i$ said that he$_i$ ran."

Chapter 7 then considers whether switch-reference belongs in this family of constructions. I argue that the answer is sometimes: SR on adjunct clauses typically does not involve a ghostly DP operator controlled by the matrix subject, but SR on complement clauses does in those languages that have it. Finally, Chapter 8 takes up a theoretical topic: it takes a closer look at the principles of obligatory control by which an argument of the matrix verb controls a ghostly DP at the periphery of its CP complement, a notion that is used quasi-descriptively in Chapters 2-7. There I pursue a unified analysis of this phenomenon and normal cases of the control of the PRO subject of infinitival clauses, even though the choice of which argument of the matrix verb controls the null DP seems different in some cases.

This book is organized more on the basis of the constructions being analyzed than on the midlevel theoretical discoveries being made. There are some theoretical discoveries and innovations, however, and I would be happy for interested readers to find them and appreciate them. Here is an overview of the main ones.

One midlevel result concerns agreement. Three of the six constructions involve full-fledged agreement, including the transfer of phi-features: upward C-agreement, allocutive agreement, and the indexiphor construction in (35). Comparison of these constructions reveal that they are often subject to an additional condition on what can control the ghostly DP operator: in two of the three constructions, the controller must itself be the goal of an agreeing T. I call this the T/Agee condition. I claim that this constraint shows us something about how Agree works. It testifies to Arregi and Nevin's (2012) distinction between Agree-Link, which creates a pointer from a probing functional head to its NP/DP goal, and Agree-Copy, which in a distinct step carries out the transfer of phi-features from the goal to the probe. The main discussion of this is at the end of Chapter 2, with a brief reprise in Chapter 3 and an extension in Chapter 6.

Another midlevel result has to do with the typology of ghostly DP operators: what kinds of features they can have and how this affects their behavior in ways that go beyond simply determining the features of the pronouns that they bind. The most important distinction is between ghostly DPs that have interpretable features and those that do not. I argue that ghostly DPs that do not have interpretable features need to undergo obligatory control almost immediately (on the next phase) or they run afoul of the principle of Full Interpretation. In contrast, ghostly DPs that have interpretable features can survive outside of contexts of obligatory control—in adjunct clauses, subject clauses, relative clauses, and root clauses. This topic is touched upon in most of the chapters, but the first key discussion is toward the end of Chapter 3 (comparing the goal of allocutive agreement to the goal of upward C agreement) and the fullest discussion is toward the end of Chapter 5, by which point we have a critical mass of examples that is large enough to discern some patterns.

The operator-pronoun binding relationship gets some significant theoretical attention in Chapters 5 and 6. Chapter 5 discovers some new types of crossover violations in languages with logophoric pronouns, where the same ghostly DP operator may or may not be allowed to bind both a logophoric pronoun and an ordinary pronoun,

depending on the exact feature values of the items involved and the details of the c-command relationship. This is potentially relevant to the formulation of the crossover principle. Chapter 6 then argues that a ghostly DP operator can add phi-features to the pronoun it binds, creating complex phi-feature bundles that result in indexiphoric constructions like (35).

The biggest theoretical challenge and opportunity is saved for last and gets a chapter of its own (Chapter 8). This concerns the fundamental nature of the control relationship between an argument of the matrix verb and the ghostly DP operator in the periphery of its CP complement. In some important ways, this is analogous to the relationship between PRO and its controller in ordinary cases of obligatory control. In particular, both fall under the "obligatory control signature" of Landau (2013), as I show throughout. However, there are some differences too, particularly as to which argument of the matrix verb is chosen to be the controller of the null DP in the complement clause. I take these challenges to be opportunities, considering what a generalized Control Theory could look like that has both ordinary control theory and what is needed for ghostly DP operator constructions as special cases. I present a new theory of obligatory control in which it is induced by event identification and thematic uniqueness in the sense of Carlson (1984). This is the most theoretical ambitious piece of the unified analysis of this set of crosslinguistic constructions.