Research Statement
Yangchen Roy

My research interest lies in the exploration of the empirical in linguistics to inform and refine the theoretical. I work on syntax and its interfaces with morphology, semantics, and pragmatics and explore these using child language acquisition data, adult language processing data, and fieldwork of understudied languages.

Most of my theoretical and empirical investigations have focused on dependency relations, given that these do not always encode a straightforward signifier-signified relationship, but a computation often systematic and contingent on a division of labour between the components of the grammar and processing mechanisms. In particular, I have studied the following phenomena: pronominal and reflexive anaphora (Bangla and Malayalam), non-(quintessential)-φ feature agreement as politeness agreement (Bangla), allocutive agreement (Bangla), and 1st person addressee agreement (Spiti), nominal-verb and noun-modifier agreement (Gaddi) from a distributed morphology perspective, and long-distance binding (Malayalam and Spiti). I have also extensively researched negation and NPIs (Bangla) and case morphology (Gaddi). My other research interests, driven by my endeavour to ensure that work in formal and experimental linguistics be channelised to help solve grass-root level language issues, are endangered language documentation, grammar writing, archiving and curation, and early language pedagogy informed by acquisition research. Here I discuss some of my past and current research, which motivate the research projects I have planned for the future.

1 Quintessential and non-(quintessential) φ-features

Studying different linguistic features, their agreement patterns and their acquisition can reveal much about their types (say, morphosyntactic, morphosemantic, and morphopragmatic), which types participate in which dependency relation, and where in the grammar agreement unfolds (the syntax, the morphology or a combination).

South Asian languages are known to morphologically manifest more than the quintessential φ-features, person, number, and grammatical gender/noun class. Other features, their agreement patterns, and what they reveal about how the syntax interfaces with morphology, semantics, and pragmatics are under-researched. I have researched non-(quintessential)-φ features in Spiti (Tibeto-Burman) and Bangla (Eastern Indo-Aryan). In **Roy (2022a)** I show that the Spiti 1st person agrees with the level of formality/intimacy between the speaker and the addressee. In Spiti, all first-person pronouns, irrespective of their syntactic position, manifest this allocutive agreement, with morphological case marking not blocking agreement. This contrasts with the nominal-verb quintessential featural agreement found in many South Asian languages, where morphological case marking on a nominal prevents any verbal element from agreeing with it in featural values. I argue that allocutive agreement found in Spiti is evidence for an elaborate nominal spinal periphery, with contextual agreement features housed in *GroundPs* (Wiltschko, 2021). This, along with my finding that Bangla-speaking children lack knowledge of politeness features till the age of 6;11 **(Roy, 2022b)** feeds my argument for a context layer above the nominal layer where politeness features are born and licensed. I argue that politeness is built up of more than one feature — [FORMAL] and [STATUS] — of which [STATUS] has gradient values, as proposed in Portner et al. (2019). The context-dependent relationship between the interlocutors or between the interlocutor and the other determines feature values and thus pronoun choices. I claim that the pronoun agrees with the [STATUS] and [FORMAL] features hosted in the head of the contextual layer.

Further evidence of a contextual layer at the periphery of spines is provided by **Bhattacharjee & Roy (2022)**. Investigating the distribution of the Bangla markers *go* and *re*, we find that both markers appear on utterances directed towards familiar addressee-interlocutors, with *go* appearing when the speaker is younger than the addressee and *re* appearing when the speaker is older. As in most languages with addressee agreement (Antonov 2015), *go* and *re* manifest in root clauses and not embedded clauses. We argue that *go* and *re* are allocutive markers and that their properties are evidence for a context/speech-act layer above the CP layer. Thus contextual layers as licensing domains can be motivated at both the nominal and the clausal peripheries. Studying allocutive agreement by investigating its acquisition and comparing it to quintessential φ-feature agreement would be a meaningful next step.

A study comparing the acquisition of different features, including their agreement, would help reveal the differences between the two. Malayalam (Dravidian), for instance, allows for a reference mechanism that I call *"person switch"*. This is essentially a case of non-pronominal 3rd person nominals functioning as the 1st or 2nd persons. This raises the question of whether *Person* is in such cases quintessential φ or not. Therefore is it a valued feature within the CP, as it is in its regular avatar, or is it, in its 1st person avatar, valued in the contextual/pragmatic layer above the CP? Is the timeline of acquisition of both kinds of 3rd person the same? This is an investigation worth pursuing. Further, studying which features in speakers'

featural inventories are affected when they are bi-/multi-lingual, heritage or immigrant speakers would also contribute to an understanding of the location of features in the grammar.

2 Antecedent-Anaphora Agreement vs. Controller-Target Agreement and their acquisition paths

In my research on pronominal and reflexive anaphora, I investigate the general mechanisms — Merge and Agree — of the language apparatus employed in resolving reference, and those that are contingent on cross-linguistic variation. My research reveals that studying the theoretical and empirical difference between anaphor-antecedent agreement and controller-target (nominal-verb and noun-modifier) agreement can unravel the roles of *Merge* and *Agree* in the grammar. This is a crucial pursuit since most literature argues for Agree driving both anaphora resolution and controller-target agreement in the syntax.

Reference resolution, which is an assignment of values to nominals based on assumptions about a given world i.e. a semantic assignment, should not technically be dependent on a purely syntactic/morphosyntactic phenomenon like agreement or on mechanisms where antecedent-anaphor linking piggybacks onto some other linkage, like case-licensing (Reuland, 2011). *Agree*-based accounts risk massive look-ahead (Upward-Agree), especially in cases of long-distance binding. Often features that have no morphological manifestation in natural language are posited to drive the derivation (Adger & Ramchand 2005, Hicks 2009), leading to the multiplication of theoretical devices with little empirical basis. *Merge*, on the other hand, is the simplest operation, and has backing from acquisition research as being acquired early, and as being a part of the genetic endowment of general human cognition (think the addition algorithm).

In **Roy (2023)** I designed three novel comprehension tasks (142 adult Bangla and 100 adult Malayalam speaking participants) and five semi-structured elicitation tasks (10 adult controls and 19 children for Bangla) to test adult and child competence of pronominal and reflexive anaphora in Bangla and Malayalam. The adult comprehension "Match-Mismatch" Tasks 1 and 2 aimed to determine if object pronouns are disallowed to be locally bound and if pronouns and reflexives are both permitted in the spec DP position. Task 3 aimed to check whether a context of contrast motivates the choice of a reflexive over a pronoun, since reflexives have only one reading (i.e. the bound interpretation), unlike pronouns, which additionally have deictic and coreference interpretations. The elicitation tasks (all picture-based storyboards on the *Explain Everything* interactive whiteboard app) were applied to both adults and children. Task 1 targeted the object position, and Tasks 2 and 3 the Spec-DP pronoun. Tasks 4 and 5 tested for the "avoid ambiguity" principle. The results reveal that while children have adult-like knowledge of pronominal anaphora, they struggle with resolving polite pronouns, producing grammatical but infelicitous utterances (by lowering the pronouns' politeness level). The fact that they struggle with polite pronouns, and not with non-polite ones suggests that their grammar abides by Principle B. The mismatch in the agreement of politeness features does not cause a crash in the derivation, meaning Agree operating on (at least) these features (leading to morphological agreement) is not necessary for binding to follow. Disagreement and binding can co-exist.

Children's ability to decipher anaphora at a very early age (<3;0), and their established knowledge of binding and c-command (and thus *Merge*), motivated me to ask whether it is *Merge* rather than *Agree* that drives anaphora resolution. For anaphora resolution, the anaphor and the antecedent must be linked. I argue that this linking happens in the syntax via internal and external *Merge*, such that this dependency is available to LF for conversion into logical syntax (i.e., *binding*). This linking in the syntax would be subject to constraints on the syntax (i.e., *locality*), which will translate into restricting anaphora resolution to phasal chunks. What could be a consequence of linking is the matching of fully valued features between the anaphor and the antecedent, rather than the other way around, where the matching of features (via *Agree* or similar) links the anaphor and the antecedent leading to valuation. One is then questioning if the mechanism at the bottom of anaphor-antecedent feature matching and that at the bottom of controller-target agreement is the same. Studying the timeline of the acquisition of anaphor-antecedent agreement and that of nominal-verb agreement and comparing the two can help explore this and I intend to do this as a long-term project.

3 Language documentation, archiving & building acquisition sketches

I have documented, with colleagues, the linguistic landscape of India's largest residential research university. Our data collection methods included a questionnaire to elicit data on the language choices of students, faculty, university staff, and support service providers. We also conducted interviews with community members on what their language means to them and what their experience has been of using it in the past and present. The survey and the interviews revealed that our university spaces mirror the linguistic diversity of the subcontinent and are evidence of thriving multiglossia. However, minority language speakers are discouraged in schools from using their native tongue in and outside the classroom, shaping their views of their language. The output of this research can be found at the **Fling@JNU's Langscape 2017 page**.

Since 2015, I have been part of two documentation teams at Jawaharlal Nehru University, one on Gaddi (Indo-Aryan) and the other on Spiti (Tibeto-Burman). I have studied the morpho-syntax of tense, aspect, mood-modality, and case in Gaddi, based on data collected by the team using the "Swadesh worldlist", Abbi (2001), and novelly designed TAM and Case questionnaires. An output of this work is the book *Gaddi Grammar*, currently under revision for the *Grammars of World and Minority Languages Series* of the University College London Press. My documentation of the Pin Valley variety of Spiti began in 2019. Our team of linguists lived in the mountains and collected language data from a group of six speakers and a group of Spiti monks over a period of 10 days. We collected over 68 hours of data — 2168 sentences, 1807 words, 4 songs, and 7 narrations. We designed elicitation tools to collect data on post-positions and case markers, tense, aspect, mood, evidentiality, pronouns, reflexives, and reciprocals, negation, questions, focus and topic, numerals, and quantification. An output of this work is a **Hindi-English bilingual questionnaire (Roy, 2019)**, which I used to study the pronominal, reflexive, and reciprocal systems of the language.

In the future, I aim to incorporate a new method in my endangered language documentation research — using acquisition sketches of endangered languages (i.e. *Acquisition Sketch Project* (Hellwig et al., 2023)), to gauge language vitality. Simultaneously comparing the morpho-syntactic properties of adult and child language can also help deliberate on the variation found in two linguistic systems. This is a method that can also benefit research on cross-linguistic variation of morphosyntactic features. For example, "Child Hindi" allows multiple-agreement in the verbal domain (Pareek, 2017). The verb can agree with the subject while the auxiliary may agree with the object. This is completely impermissible in "Adult Hindi", where the entire verbal complex must agree with the highest morphologically unmarked nominal. However, this kind of multiple agreement, which "Child Hindi" permits, can be found in Gujarati (Western Indo-Aryan). Therefore, documenting children's longitudinal interaction amongst themselves and with other members of the language community can serve as a rich language database produced cost and time effectively. It can be used to build language resources and to inform linguistic theory.

4 Early-language pedagogy informed by language acquisition: Building language games and apps for children

Another outlet of my research in formal and experimental linguistics is language pedagogy. Through my research on the acquisition of morpho-syntactic features, it is evident that some of these features do not require formal (i.e. explicit) instruction in an elementary school language classroom. But in my visits to different classrooms, I have observed that instructors tend to (unnecessarily) teach those features using linguistic jargon. However my research (and that of others) has shown that children would only require help with the ones they struggle with, and only in case the language in question is not the child's first language. In all other cases, the child should be left alone to hit the natural milestones of the development of a linguistic feature/structure. Therefore, I believe that insights from research on the formal properties of language, as well as language acquisition, should be used to build language games and routines for second language classroom. Based on my work on Bangla polite pronoun acquisition, I have designed the prototype of a game, *Shajao Roshogollar Payesh* "Decorate the Pudding". In this, children replace pictures of characters of various ages or related them in tiny lift-the-flap books stuck to a board, with pronouns. The board is a popular sweet dish from Bengal (Roshogollar Payesh) and the pronouns children use to do the replacing are written on "toppings" for the sweet dish — pistachios, cashew nuts, almonds, and raisins. Inside the flap is an event that the character is performing (testing for subject polite pronoun knowledge) or one where something is being done to the character (testing for object polite pronoun knowledge). Children describe the inside of the flap after "adding the topping". The game can be adapted for both a high-resource classroom and a low-resource one, or even as a game to play at home. These kinds of games can be highly useful in boosting children's confidence in multi-lingual classrooms, where the medium of instruction is often not the child's home language, disadvantaging and marginalizing them. I intend to build more such games in the future. Some could be built as computer resources and tablet apps in collaboration with other researchers.

References: Abbi, A. (2001) *A Manual of Linguistic Fieldwork And Structures of Indian Languages* **Adger, D., & Ramchand, G. (2005)** *Merge and move: Wh-dependencies revisited.* **Antonov, A. (2015)** *Verbal allocutivity in a crosslinguistic perspective* **Bhattacharjee, N. & Roy, Y. (2022)** *On Bangla go and re,* **Chomsky, N. (2019)** *The UCLA lectures,* **Hellwig, B. et al. (2023)** *The acquisition sketch project,* **Hicks, G. (2009)** *The derivation of anaphoric relations,* **Pareek, B. (2018)** *Nominal features in Hindi language acquisition: A study of agreement and modification* **Portner, P. et al. (2019)** *The speaker-addressee relation at the syntax-semantics interface,* **Reuland (2011)** *Anaphora and language design,* **Richards, M. (2020)** *Rethinking principles A and B from a free Merge perspective,* **Roy, Y. (2019)** *Questionnaire for Eliciting Pronouns, Reflexives and Reciprocals in South Asian Languages: English-Hindi Bilingual Set,* **Roy, Y. (2022b)** *On the context layer of Bangla pronouns,* **Roy, Y. (2022a)** *On the participant layer at the periphery of Spiti pronouns and reflexives,* **Roy (2023)** *Exploring Anaphora Binding and its Acquisition: The View from Bangla and Malayalam,* **Sundaresan, S. (2020)** *Distinct featural classes of anaphor in an enriched person system,* **Wiltschko, M. (2021)** *The grammar of interactional language.*