

Assamese Field Methods and Language Documentation Report

Divya Kadav

University of Mumbai, Department of Linguistics

12/11/2024

Introduction

This report documents the field methods and language documentation techniques used to study Assamese. The data collection took place over a period of 1st August 2024 to 21st October 2024 in the classroom. The purpose of this research was to document and analyse the phonological and morphological analysis of Ujoni Assamese. We used abbi word lists, pictures and gave sentence prompts to collect a wide range of language data. In this report, we will look at the informant's background, literature review, methods used to collect data and documentation of language and the challenges.

About the Language

Name of the Language - Assamese

Geographic Location - Assam

Status - Classical language of India

Language Family - Indo - Aryan

Assamese

About 15.3 million people in north-east India speak Assamese. Most speakers (99%) live in Assam, with a few in neighbouring states. Assamese is the official language of Assam, a state in north-east India. It is often used as a common language in areas with many different languages. Assamese is related to Bengali and Oriya. There are also Assamese speakers in Bhutan and Bangladesh. Only half of Assam's population speaks Assamese. The rest speak other languages, including Bengali, Hindi, Nepali, Oriya, Bodo, Mishing, Karbi and Santali. Assamese developed as a border language in an area where many languages mixed. It went through major sound changes and its grammar was simplified. This is probably because Assamese is used to communicate with non-Indo-Aryan speakers. (Gutman.Avanzati. 2013)

Literature Review

In preparing for this fieldwork, I made use of the foundational resource A Manual of Linguistic Field Work and Indian Language Structures by Anvita Abbi. We, as a class in general have used Abbi sentence list and word list to form sentences to provide the informant. This manual is about studying language in the field with Indian examples. It explains how to get data from South Asian speakers. The book explains how to collect, analyse and present language data. She has given examples of Indian languages.

The book explains how language groups have influenced each other. There are tips for people collecting data in the book. The appendix includes questionnaires on language topics to help field workers. The manual provided with the methods required for conducting phonological and morphological analyses.

To better understand and see what previous studies have been done on Assamese, I used Assamese: Its Formation and Development by Kakati Banikanta. I also referred to Field Linguistics, A Guide to Linguistics Field Work by William J. Samarin. In his book, Formation and Development of the Assamese Language, Kakati Banikanta has provided an in-depth analysis of the various dialects of Assamese, like the Kamrupi dialects. Specific phonological and morphological characteristics observed in the Buddhist texts known as the 'Dohas' have been transmitted uninterruptedly from the early to the modern period in the Assamese language. The Assamese language and the dohas share certain morphological characteristics, including the dative case ending in -lai.

The Assamese Period is divided into three parts: the early Assamese, which is also split into the prevashnavite and the vaishnavite sub-periods; the middle Assamese, which had a period of Ahom court; and the modern Assamese. The author states that certain Assamese words exhibit similar or parallel formations in Westernmost languages, such as Marathi. In certain Kamrupi dialects, the η phoneme is pronounced as a bilabial semi-vowel [w]. Additionally, the author has discussed the alveolar nature of the liquids in Assamese. In Assamese, there is no differentiation between the phonemes /r/ and /l/, which are both pronounced as alveolar consonants. Assamese, as a Magadhan dialect, shows only one sound, the single liquid /l/. However, Assamese also features both the phonemes /r/ and /l/. The /r/ is dropped, which is a characteristic of modern Assamese. In addition to this, the author has discussed the phonology, morphology, and different dialects of Assamese. The Author has also mentioned how different lexical correspondence have influenced Assamese like Khasi, Kol, Malayan, Tibeto Burman, Bodo, The Thai or Ahom, Non Aryan Elements like Austric, Bodo and more (Kakati. 1941).

Informant Details

Name - Priyakhi Khanikar

Age - 28 years old

Gender - Female

Location - Assam, Dibrugarh District

Dialect/Variety - Upper Assamese (Ujoni Dialect)

Occupation - PhD Student at University of Mumbai, Department of Linguistics

Education - Her primary schooling was done in Assam. She obtained her Master's and MPhil degrees from Delhi University.

Language -

	Assamese	Bengali	Hindi	English
Read	✓	✓	✓	✓
Write	✓	✓	✓	✓
Understand	✓	✓	✓	✓

Mother Tongue - Assamese

Background - Priyakhi received instruction in the Assamese language for the first three years of her education. The three-year period of formal education included the basic level of Assamese. She began to read Bengali as the script was similar to Assamese, influenced by the films and literature she encountered. She began studying Bengali in the eighth grade. She pursued her entire academic education in an English-language environment. She has a proficiency in all four languages, which she makes use of in her engagement with media and articles. She attended Vivekananda Kendriya Vidyalaya, where she studied English, Hindi, and Sanskrit. The script is similar for upper and lower Assamese. In her hometown, Assamese is the primary language used in conversation with family and friends. She also reads the newspaper in Assamese. However, with migrants, she switches to the language preferred by the individual. The Brahmaputra River divides between the upper (ujoni) and lower (namoni) regions of Assam. So, the dialect or variety spoken by her is the ujoni variety of Assamese. In educational settings and among migrants, she alternates between languages or simply code switches according to the requirement.

Session Overview

Our session was conducted in the Indo Canadian Room, Ranade Bhavan in University of Mumbai. Our sessions with the informant involved sharing a predetermined set of words and sentences with her about two or three days before each session. We provided her the material in advance so she could take her time to reflect on the vocabulary and sentence structures. Each session was focused on a specific theme. For example, one session involved a word list comprising colours and sentence structure for comparative and contrastive purposes. This was done in order to allow her to prepare responses that were related to these specific topics. The students were divided into three groups.

1. Group A consisted of Divya Kadav, Palakshi Sathe, Tithi Desai, and Rishi Goswami.
2. Group B comprises Adya Arun, Ananya Kashelkar, Nivedya Suhas, and Amanda Fernandes.
3. Group C includes Tanisha Dhareshwar, Avantik Zinta, and Sanika Joag.

The initial session was done by Professor Avinash Pandey and Renuka Ozarkar. Each session lasted two hours, and was conducted on Mondays and Thursdays.

Methodology

On the day of the session, one of the members of the group who is incharge for their turn would assume the role of the elicitor. The elicitation process would begin with the speaker repeating each word or sentence requested on a three-time basis. Because of this approach we could get enough time to identify the sounds that had been pronounced and to write the responses.

We started with our session on simple vocabulary and then gradually moved on to the more challenging aspect. So the roles taken by each group member other than the elicitor was one would keep record of managing the timings, one would help to organise and compile the questions that we were going to ask her in a word file and send it to Renuka and Avinash for checking. Once it was checked, we would send the word file and pdf to the google groups for everyone to access. With time, Priyakhi became more comfortable in the sessions. Her responsiveness changed where she would repeat each sentence several times, with time she didn't need to repeat them so often. Also, with the glossing of the sentences, we would be able to understand the words. For the sentences where we could not give English sentences, we tried to compensate with Hindi sentences. So for the words or sentences that we were struggling to understand, we would try to use similar structure but different sentences to try to understand what it was.

As I collected data, my curiosity about certain linguistic features led to a rich dataset. For example, I was interested in compound words, and during the elicitation process we got to know that gol in golpata was mostly used by Bengali speakers and Assamese-speaking Muslims. In addition, my interest in classifiers expanded the dataset. I asked for expressions for 'it is raining' and received a variety of terms.

I maintained a field diary in which I noted the words and sentences in their original form, provided a gloss, and then wrote a translation. Sticky notes were attached to each session in order to clarify which one was being referred to. Furthermore, the dates and names of elicitors were also recorded for each session. The diary was organised into a numbered sequence of pages. I analysed the data using Google Docs. To maintain an online record, I made use of the Google Sheets application. The language of communication was English, with the occasional use of Hindi. The informant was asked a question in English, which she then answered in Assamese. On occasion, we would seek clarification from the informant regarding certain words that were unclear to us. At times, we encountered difficulties in managing time effectively. In these instances, we would ask the questions of those in the following session. We initially held an open session at the beginning of the process for those who wished to clarify any doubts they had, like sentences or regarding pronunciation. After this, we proceeded with our scheduled session.

Tools - The audio recording of the narration was made using a Zoom H6 audio recorder set at 44.1 kHz, 16-bit WAV format. We would carry our field methods diary with us for writing the session information, pen, pencil.

Session Prompts and Pictures - We used pictures to help her give visual descriptions of the colours we wanted to have words for in Assamese. The data regarding the questions and colours were printed out and provided to her. As a result, we were able to learn about the various colours used in their culture. We asked her about the colours and whether they had any for their traditional sarees.

To analyse the five most common consonants and vowels, I used R programming. Using this software, I performed a frequency analysis and identified the five most common consonants and vowels in the data set. To visually represent these findings, I created a graph using the ggplot2 package, which effectively illustrated the distribution of these phonetic elements.

In addition, I ensured proper data management and backup by storing all collected data in Google Sheets and maintaining documentation of the analysis process in Google Docs for future reference.

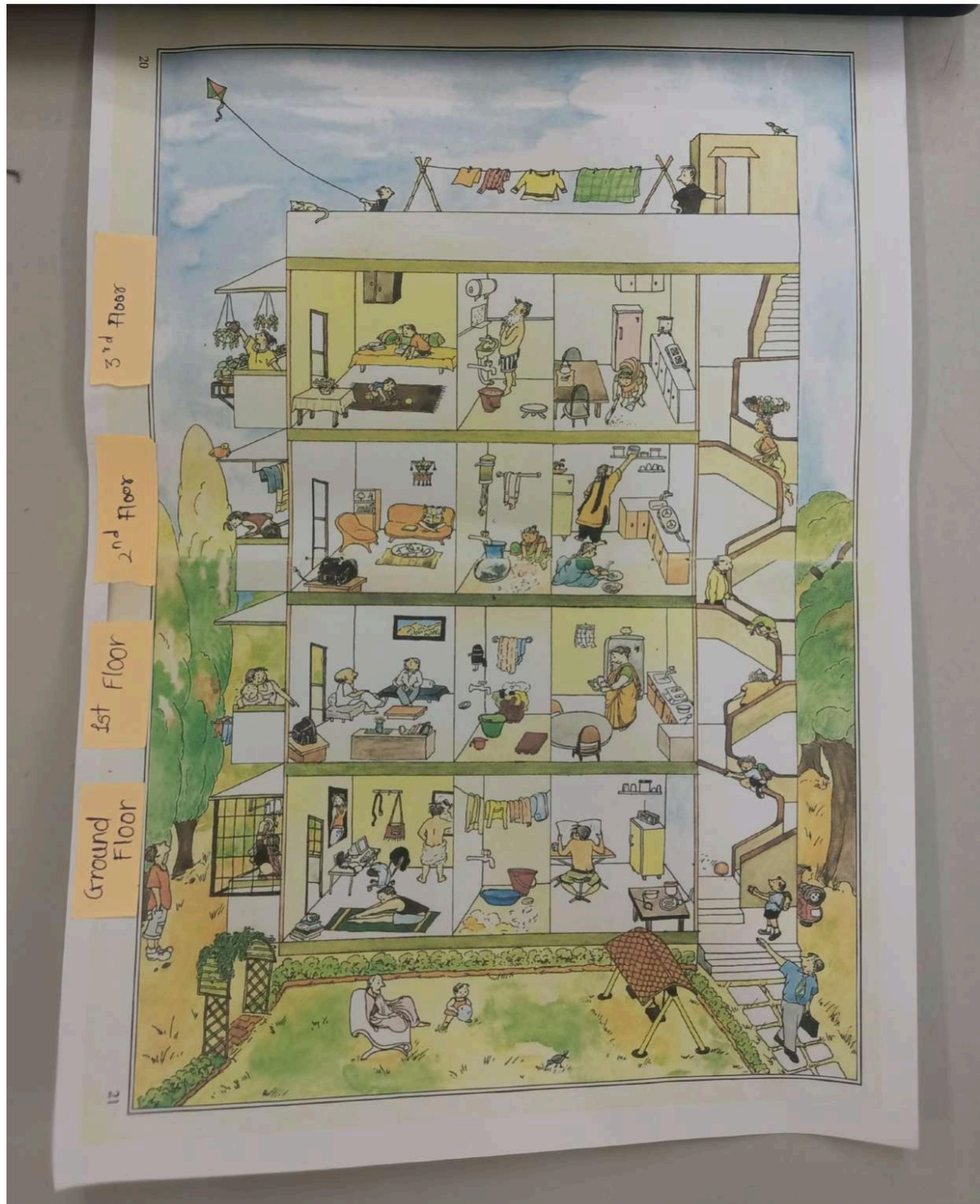
Some days we moved from a structured vocabulary-sentence approach to adding sessions for narration. For narrative sessions, we would ask her to talk about culturally significant issues or to describe imaginary scenarios. For example, we might ask her to talk about 'bihu', a traditional festival, or to describe a birthday party. Occasionally we asked situational questions, such as "How would you describe what happens if this event happened in a different way?"

Narration sessions needed a different setup because they were longer. Sometimes, we found that having her describe a picture or scene helped her feel less nervous about narrating. Our classroom was usually quiet, but sometimes we had to deal with noise from outside. Background noise sometimes affected the clarity of recordings. We adapted by moving the recorder closer to her.

Limitations

Despite the data being detailed, there were still some limitations. The dataset could have been more structured, with a negative dataset alongside each data sample. This would have helped to analyse grammatical structures more clearly and in more detail. Some sentences were complex, which made it harder to understand and collect data. A simpler set of sentences might have helped. The way we ask an informant to speak may not show how people really speak. Spontaneous sessions could have revealed more about the language.

Figure 1 - Narration Picture for Tense and Aspect



To give an example for narration, we provided Priyakhi with a picture showing a four-story building with a terrace, staircase and garden. Each floor, including the ground, first, second and third floor shows different people doing their daily activities. We used this image as a prompt for narration. We asked her to describe the actions taking place on each floor. We had written this narration in the Word-File which we had sent a few days before the session began. The details are given below:

1. Considering that we are going to conduct this Field Methods Lecture on 7th October 2024, below mentioned is the data we are looking for:
2. The actions happening inside the rooms on the ground floor (including the balcony) were taking place on the 6th of October.
3. The actions happening inside the rooms on the first floor (including the balcony) are currently on-going on 7th October, during the lecture.
4. The actions happening inside the rooms on the second floor (including the balcony) will take place on 8th of October.
5. The actions happening inside the rooms on the third floor (including the balcony and the terrace) are already complete (and not on-going).
6. The actions taking place on the stairs happen to the character everyday in their lives.

The difficulties encountered during this narration were due to the fact that the informant was unable to provide the narration in a single continuous session. The visual representation appeared to be a challenge for her, as it required careful consideration of the specific floor and the date on which the action occurred. To address this issue, we requested that she narrate each image as if it were taken on October 6th, then on October 7th, 8th, and so on. Additionally, we encountered a challenge with the Zoom Recorder, which we presumed had converted all the data into .wav files. However, one of the files turned out to be corrupted. From this, we learned that only one recorder is not enough. One should have multiple recording systems available to avoid such situations.

As Session 1 was our introductory session and we were in the process of learning the fundamentals of fieldwork, we were guided by Avinash and Renuka. The first data collected was related to animals. Prior to the session, Renuka had prepared a series of sentences, which she then provided to Priyakhi. The location was the Indo-Canadian Room. The following session was also based on the topic of birds. Renuka and Avinash were responsible for guiding the elicitation process. During our eighth session on kinship, we reviewed a list where the ego was identified as female. At this point I asked a question: what if the ego were male? I wanted to understand how changing the gender of the ego would affect the kinship terms used. I wanted to understand how changing the gender of the ego would affect the kinship terms used. During the third session, members of Group A worked together to identify the most appropriate word type, which was agreed to be 'colour'. Furthermore, it was decided that comparative and contrastive type sentences would be used, as they were considered to align well with the concept of colour. Similarly, with regard to body parts, we selected sentences for relativisation, participialisation and adjectives, on the basis that we

could construct sentences such as "Call the goldsmith who had a fever". In the ninth session, an attempt was made to fill in the gaps in the data observed until that point by including sentences. In the fourteenth session, we attempted to provide her with sentences to verify the existence of ergative constructions. Additionally, we requested that she tap for each word with her hands. We presented her with a selection of words to analyse the syllable structure. This was one of the techniques I had previously discussed in my presentation with my classmates.

Session Details

Session No.	Date	Type of data collected	Duration	Elicitor
1	01/08/2024	Words for animals and simple sentence	2 hours (9.30 am - 11.30 am)	Avinash Pandey and Renuka Ozarkar
2	05/08/2024	Words for Birds and simple sentences.	2 hours (9.30 am - 11.30 am)	Avinash Pandey and Renuka Ozarkar
3	8/08/2024	Words for colour and sentences for Comparative and Contrastive.	2 hours (9.30 am - 11.30 am)	Divya Kadav, Palakshi Sathe, Rishi Goswami, Tithi Desai
4	12/08/2024	Words for Pronouns and Provision and Food Category, Sentences for Conditional and Coordination.	2 hours (9.30 am - 11.30 am)	Adya Arun, Amanda Fernandes, Ananya Kashelkar, Nivedya Suhas
5	15/08/2024	Words for Verbs and Nature. Sentence for Negatives.	2 hours (9.30 am - 11.30 am)	Tanisha Dhareshwar, Sanika Joag, Avantik Zinta
6	22/08/2024	Words for Body-Parts. Sentence for Relativization, Participialization and Adjectives.	2 hours (9.30 am - 11.30 am)	Divya Kadav, Palakshi Sathe, Rishi Goswami, Tithi Desai
7	29/08/2024	Words for Ornaments and Clothing. Sentence type for Interrogatives and Imperatives.	2 hours (9.30 am - 11.30 am)	Adya Arun, Amanda Fernandes, Ananya Kashelkar, Nivedya Suhas

8	2/09/2024	Words for Kinship Terms and Relational Terms.	2 hours (9.30 am - 11.30 am)	Tanisha Dhareshwar, Sanika Joag, Avantik Zinta
9	23/09/2024	Words for Professions. Various Sentence Type to fill the gap.	2 hours (9.30 am - 11.30 am)	Divya Kadav, Palakshi Sathe, Rishi Goswami, Tithi Desai
10	30/09/2024	Words for Kitchen, items and other equipment. Sentences for Modality.	2 hours (9.30 am - 11.30 am)	Adya Arun, Amanda Fernandes, Ananya Kashelkar, Nivedya Suhas
11	3/10/2024	Words for Adjectives.	2 hours (9.30 am - 11.30 am)	Tanisha Dhareshwar, Sanika Joag, Avantik Zinta
12	12/10/2024	Sentences for Causatives [Abbi Sentences]	2 hours (9.30 am - 11.30 am)	Divya Kadav, Palakshi Sathe, Rishi Goswami, Tithi Desai
13	14/10/2024	Abbi Sentences for Cases, Agreement and Coindexing	2 hours (9.30 am - 11.30 am)	Adya Arun, Amanda Fernandes, Ananya Kashelkar, Nivedya Suhas
14	21/10/2024	Open Session	2 hours (9.30 am - 11.30 am)	Discussion was open to all the students.

Documented Data

Transcriptions - In the initial stages of the project, I made use of the International Phonetic Association (IPA) Chart by Kiel for the transcription of data. However, as I gained familiarity with the specific consonants and vowels used by the informant, I began conducting transcriptions independently, without the reference chart, during the data collection phase. To transcribe the data for online entry, I used the SIL Und-LATN font, which contains IPA fonts. This data was then imported into Google Sheets for storage and analysis.

Glossing - Our informant gave us rough translations of the sentences. We then talked with our colleagues to understand the language better. We compared our interpretations, shared insights and resolved any ambiguities in meaning or structure. We improved our glossing skills and understanding of Assamese by discussing different interpretations and grammar. When we found a phrase or construction difficult, we tried to understand it first. Then we went back to our professors, colleagues for more help. This helped us to understand the glossing. Other than this, presentations were conducted to help us know about the observations and generalisations we made. I have used Leipzig Glossing rules for morphological glossing.

Data Documentation - As previously stated, I have used a field diary to record my observations of the Assamese language. The notes have been structured in the following format: The first line contains the Assamese International Phonetic Alphabet (IPA) transcription, the second line contains glossing, and the third line contains the translation. In this manner, I used two diaries for my fieldwork. For online documentation, I drew upon my previous experience of data archiving, acquired during my internship at a maritime history society. I attempted to utilise these techniques, writing the data in a Google sheet and searching for specific items using the CTRL + F function. For the analysis of the data, I made use of Google documents.

Figure 2 - Documentation in Google Sheets

A1

	A	B	C	D	E	F	G	H	I	J
1	o	Note Section	ENG WORD	ASSAMESE	SENTENCE					
2	S1	Session 1								
3		Animals								
4		01/08/2024								
5		Elicitor - Avinash and Renuka								
6										
7										
8										
9										
10										
11										
12	1		Animal	dʒontu		habi-t	bohut	bonoria	dʒontu	ase
13						jungle=LOC	many	wild	animal	be.PRES.3
14						There are many wild animal in the jungle				
15						habi-t	bohut	bonoria	dʒontu	thake
16						jungle=LOC	many	wild	animal	be.PRES.3
17						There live many wild animal in the jungle				
18						habi-t	e-ta	dʒontu	ase	
19						jungle=LOC	one-CLF	animal	be.PRES.3	
20						There is an animal in the jungle				
21						amaɹ	e-ta	goɹu	ase	
22						1PL GEN	one-CLF	cow N.SG	have.1	
23						We have a cow				
24	2		Cow	goɹu		mul	e-ta	moh	ase	
25						1SG GEN	one-CLF	buffalo	have.1	
26						I have a buffalo				
27						amaɹ	du-ta	ku k'w	ase	
28						1PL GEN	two-CLF	dog.PL	have.1	
29						We have two dogs				
30						amaɹ	e-ta	ku k'w	ase	
31						1PL GEN	one-CLF	dog.SG	have.1	
32						We have a dog				
33	3		Dog	ku k'w						
34										
35										
36										
37										
38										
39										
40										

+

≡

Dataset ▾

Analysis ▾

Phonotactic Constraint ▾

Sheet10 ▾

Narration

References

1. Anvita Abbi. 2001. A Manual of Linguistics Field Work and Indian Language Structures. Lincom publishers. ISBN: 3895864013,9783895864018
2. IPA.2020. Kiel.
https://www.internationalphoneticassociation.org/IPAcharts/IPA_chart_orig/pdfs/IPA_Kiel_2020_full.pdf
3. Thapasya. Sansuma B.2024.Three Day Workshop Language Documentation Techniques and Technologies. School of Humanities and Social Sciences. IIT Indore.
4. Class and Discussion Notes of MA-I (Semester I, II, III) Linguistics.
5. Haspelmath. Bickel. 2015.Leipzig Glossing Rules. Committee of Editors of Linguistics Journals. Max Planck Institute for Evolutionary Anthropology and Department of Linguistics of the University of Leipzig.
6. Alejandro Gutman, & Beatriz Avanzati. (2013). Assamese.
<https://www.languagesgulper.com/eng/Assamese.html>

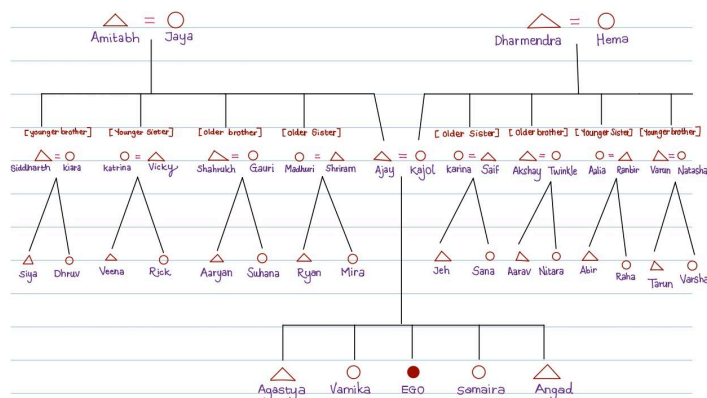
Appendices

Picture shown for Kinship Terms:

Group 3 - Avantik, Sanika and Tanisha

Kinship terms and Relational terms

02/09/2024



Group 3 : Avantik, Sanika and Tanisha

Word List: Kinship terms and Relational terms

Sr. no.	Questions
1.	What is the relationship between ego and Samaira? What would Ego call Samaria?
2.	What is the relationship between ego and Angad? What would Ego call Angad?
3.	What is the relationship between ego and Vamika? What would Ego call Vamika?
4.	What is the relationship between ego and Agastya? What would Ego call Agastya?
5.	What is the relationship between ego and Ajay? What would Ego call Ajay?
6.	What is the relationship between ego and Kajol? What would Ego call Kajol?
7.	What is the relationship between ego and Jaya? What would Ego call Jaya?
8.	What is the relationship between ego and Amitabh? What would Ego call Amitabh?
9.	What is the relationship between ego and Hema? What would Ego call Hema?
10.	What is the relationship between ego and Dharmendra? What would Ego call Dharmendra?

11.	What is the relationship between ego and Madhuri? What would Ego call Madhuri?
12.	What is the relationship between ego and Shahrukh? What would Ego call Shahrukh?
13.	What is the relationship between ego and Katrina? What would Ego call Katrina?
14.	What is the relationship between ego and Siddharth? What would Ego call Siddharth?
15.	What is the relationship between ego and Shriram? What would Ego call Shriram?
16.	What is the relationship between ego and Gauri? What would Ego call Gauri?
17.	What is the relationship between ego and Vicky? What would Ego call Vicky?
18.	What is the relationship between ego and Kiara? What would Ego call Kiara?
19.	What is the relationship between ego and Katrina? What would Ego call Katrina?
20.	What is the relationship between ego and Akshay? What would Ego call Akshay?
21.	What is the relationship between ego and Alia? What would Ego call Alia?
22.	What is the relationship between ego and Varun? What would Ego call Varun?

23.	What is the relationship between ego and Saif? What would Ego call Saif?
24.	What is the relationship between ego and Twinkle? What would Ego call Twinkle?
25.	What is the relationship between ego and Ranbir? What would Ego call Ranbir?
26.	What is the relationship between ego and Natasha? What would Ego call Natasha?
27.	What is the relationship between ego and Siya? What would Ego call Siya?
28.	What is the relationship between ego and Dhruv? What would Ego call Dhruv?
29.	What is the relationship between ego and Veena? What would Ego call Veena?
30.	What is the relationship between ego and Rick? What would Ego call Rick?
31.	What is the relationship between ego and Aryan? What would Ego call Aryan?
32.	What is the relationship between ego and Suhana? What would Ego call Suhana?
33.	What is the relationship between ego and Ryan? What would Ego call Ryan?
34.	What is the relationship between ego and Mira? What would Ego call Mira?

35.	What is the relationship between ego and Jeh? What would Ego call Jeh?
36.	What is the relationship between ego and Sana? What would Ego call Sana?
37.	What is the relationship between ego and Aarav? What would Ego call Aarav?
38.	What is the relationship between ego and Nitara? What would Ego call Nitara?
39.	What is the relationship between ego and Abir? What would Ego call Abir?
40.	What is the relationship between ego and Raha? What would Ego call Raha?
41.	What is the relationship between ego and Tarun? What would Ego call Tarun?
42.	What is the relationship between ego and Varsha? What would Ego call Varsha?
43.	What is the relationship between Kajol and Amitabh? What would Kajol call Amitabh?
44.	What is the relationship between Kajol and Jaya? What would Kajol call Jaya?
45.	What is the relationship between Ajay and Dharmendra? What would Ajay call Dharmendra?
46.	What is the relationship between Ajay and Hema? What would Ajay call Hema?

