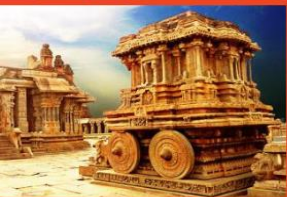


The Mother Tongue Parallel Text Corpus of India Vol. I



THE MOTHER TONGUE PARALLEL TEXT CORPUS OF INDIA VOL. I



34

Annotated, quality language data (both-text & speech) and tools in Indian Languages to Individuals, Institutions and Industry for Research & Development - Created in-house, through outsourcing and acquisition.

**Linguistic Data Consortium for Indian Languages
Central Institute of Indian Language
Mysuru, India-570006**

CENTRAL INSTITUTE OF INDIAN LANGUAGES
Manasagangotri, Mysuru, Karnataka, India, 570006
www.ciil.org

The Mother Tongue Parallel Text Corpus of India Vol. I

Dr. Narayan Kumar Choudhary, Writik Bhattacharya, Dr. Saritha S.L., Dr. Amudha R., Dr. Sajila S., Dr. Satyaendra Kumar Awasthi, Dr. Rejitha K .S., Amom Nandaraj Meetei, Dr.Vijayalaxmi F Patil, Dr. Shahnawaz Alam, Yumnam Premila Chanu, Saurabh Varik, Dr. Mansoor Khan, Chetan Baji, Sonali Sutradhar, Umesh Chamling Rai, Bhageshree K Khandale, Dr. Zargar Adil Ahmad, Dr. Modugu Kasimbabu, Dr.Kamaraj S., Syeda Mustafiza Tamim, Dr. Shalinder Singh, Hemlata Daimary, Poulami Das, Shivangi Priya, Neha Dixit, Anand Jain, Abhishek Avtans, Vishal Zanzrukia, Akanksha Srivastava, Prangshu Manjul, Ankita Tiwari, Prof. Shailendra Mohan

e-ISBN: 978-93-48633-08-8

CIIL Publication No.: 1501

First published: AD 2025 March :: Phalguna 1946 Śaka

© Central Institute of Indian Languages, Mysuru 2025

Publisher: Prof. Shailendra Mohan, Director, CIIL

Printing & Publication Unit

Aleendra Brahma, Officer-in-Charge, Printing Press & Publication Unit

B Manju Bhargavi, Unit-in-Charge, Printing Press & Publication Unit

M.N. Chandrashekar, Compositor, Printing Press & Publication Unit

Cover design: Umesh Chamling Rai

Contents

The Mother Tongue Parallel Text Corpus of India.....	1
1.1 Parallel Corpus	1
1.2 Languages in India	1
1.3 LDC-IL Parallel Corpus	2
1.4 Corpus Framework.....	2
1.5 Summary of The Mother Tongue Parallel Text Corpus of India	6
1.6 Reference.....	18

Tables

Table 1: The linguistic features and corresponding sentence count of parallel corpus	6
Table 2:Contributors to the Mother Tongue Parallel Text Corpus of India	17

THE MOTHER TONGUE PARALLEL TEXT CORPUS OF INDIA

Narayan Kumar Choudhary, Rejitha K .S.

1.1 PARALLEL CORPUS

A parallel corpus consists of an original text in one language along with its equivalent translations into one or more other languages. The most basic form of a parallel corpus involves only two languages, where one corpus is an exact translation of the other. However, some parallel corpora encompass multiple languages. Moreover, the direction of translation does not necessarily remain consistent. When multiple languages are involved, their interrelationships can vary in nature.

In parallel corpora, the target language texts often exhibit influence from the source language. This influence manifests as deviations in translation that are induced by the structure and characteristics of the source language. However, these deviations are not typically classified as errors in the conventional sense. Instead, they reflect specific lexical and syntactic choices of the source language or source text. Consequently, such target texts may be considered unrepresentative of the natural linguistic patterns of the target language.

A parallel corpus serves as a fundamental resource across various linguistic and computational fields. It plays a crucial role in lexicography, translation systems, and information retrieval, making it an essential tool for cross-linguistic studies. In grammatical and lexicographical research, parallel corpora facilitate contrastive and typological analysis by providing direct comparisons between languages. It supports knowledge acquisition for machine translation in computational linguistics, enhancing the development of automated translation systems. In language learning and teaching, parallel corpora offer authentic contrastive language data, aiding learners in understanding linguistic structures and variations across languages. It enables multilingual information retrieval, including cross-language information retrieval, allowing users to access and process information across different languages efficiently.

1.2 LANGUAGES IN INDIA

India is one of the most linguistically diverse countries in the world, shaped by various historical, geographical, and socio-cultural factors. Numerous communities have migrated or invaded India, contributing to its rich linguistic background. India's diverse geographical features have facilitated the development of distinct linguistic varieties. The coexistence of multiple religious and ethnic groups has promoted a multilingual environment, making linguistic diversity a defining characteristic of India.

The languages of India primarily belong to four major language families: Indo-Aryan, Dravidian, Austro-Asiatic, and Tibeto-Burman. According to the 2011 Census, India is home to a total of 121 major languages and 270 mother tongues.

India does not have a national language. Encouraging linguistic diversity is essential for preserving the country's rich cultural heritage. Nurturing India's multilingualism is crucial for maintaining harmony and inclusivity.

1.3 LDC-IL PARALLEL CORPUS

LDC-IL planned to develop a parallel corpus on 270 mother tongues of India. All the languages share similar fundamental features. So LDC-IL framed a structure to accommodate all the linguistic features of languages. The sentences are in English; however, they are not direct translations. Rather, they are constructed based on the syntactic, morphological, and semantic characteristics observed in Indian language structures. The primary source language is English, and it is first translated into various scheduled languages. This approach ensures that if translators of other mother tongue require a scheduled language as their source, it is made available to facilitate the translation process. This method helps translators who may have limited proficiency in English. Even though the original source sentences are in English other language linguistic features are incorporated. As a result, it may not follow the Standard English syntax. Some sentences may have been translated from Hindi or Punjabi, leading to the presence of indirect or passive constructions. Such structures can pose challenges in translation, particularly for languages where these constructions are rare or less commonly used.

It provides new insights and knowledge about the relationships between languages while also highlighting their differences. These differences and interrelations exist at various linguistic levels, including morphology, lexicon, syntax, and semantics. For instance, morphological features such as gender markers are present in Hindi but absent in Malayalam. Additionally, certain lexical items may not be available in a language if the corresponding concept or cultural element does not exist. For example, an equivalent word for "king" is not found in Tripuri language because the concept of a kingdom is unfamiliar to that community. Similarly, indirect questions or positive imperative sentences may lack direct equivalents in some languages.

All syntactic features, including sentence types, compounding, and parts of speech, are comprehensively covered. For example, within the interrogative category, various question forms such as yes-no questions, wh-questions, and tag questions etc. are included in LDC-IL parallel corpus. It is a valuable resource for comparative and cultural studies, as they facilitate the analysis of linguistic structures and cultural influences across languages.

1.4 CORPUS FRAMEWORK

The corpus is designed to capture a broad range of structural features across languages, allowing for cross-linguistic comparisons and theoretical analyses. The sentences are framed primarily based on *Lingua Descriptive Studies: Questionnaire* (Comrie & Smith, 1977), a widely used framework for linguistic description. Additionally, insights have been incorporated from reference works such as *A Manual of Linguistic Field Work* and *Structures of Indian Languages* (Abbi, 2001), *Language Documentation Handbook* (Bhattacharya et al., 2016), and *Comprehensive Questionnaire for Tribal Studies* (Sivashanmugam & Thayalan, 2012). This diverse set of sources ensures that the questionnaire captures a broad spectrum of linguistic phenomena.

This parallel corpus each language pair consists of 5,332 sentences, systematically constructed to capture 152 grammatical categories. These categories include a broad spectrum of linguistic features, including variations in tense, aspect, mood, voice, case, agreement, and syntactic constructions, thereby providing comprehensive coverage of the language's grammatical framework. Proper names within the dataset have been adapted to more familiar Indian names where applicable. The English sentences are derived from Indian language structures, with translations analyzed in the relevant grammatical contexts in each language. This ensures that the dataset remains authentic to the linguistic characteristics of Indian languages. The linguistic features and corresponding sentence count of the LDC-IL parallel corpus are outlined below:

Sl.No.	Grammatical features	Sentence count
1	Direct Speech	28
2	Indirect Speech	43
3	YES-NO Question	54
4	Wh-Question	135
5	Declarative Sentence	58
6	Tag question	35
7	Alternative question	28
8	Indirect Question	9
9	Constituent of the main clause questioned	37
10	Constituent of subordinate clause questioned	65
11	Noun Clause and reported speech	12
12	Constituent of noun phrase that can be questioned	45
13	Element of Postpositional Phrase can be questioned	19
14	Element of coordinating structures that can be questioned	31
15	Element within coordinated structure can be questioned	8
16	Element of sentence constituent that can be questioned.	33
17	Element of the sentence that can be questioned as echo- question	7
18	Yes-no echo question	46
19	Question word echo Question	42
20	Response particle 'yes' and 'no'	7
21	Positive imperative form	102
22	Negative imperative form	33
23	Negative sentence	11
24	Person/Number Combination	19
25	Simple Sentence	7
26	Interrogative as request	7
27	Noun Clause	160
28	Adjectival Clause (Relative Clause)	256
29	Sub-Adverbial Clause	171
30	Internal Structure of the sentence	251
31	Adjectival Phrase	25
32	Adverbial Phrase	25
33	Postpositional Phrase	39
34	Arguments of Postposition	5
35	Noun Phrase	79
36	Coord-General expression	63
37	Coord-Coordination- and	52
38	Coord-Coordination- but	13
39	Coord-Coordination -or	33
40	Coord-Coordination and accompaniment	98

Sl.No.	Grammatical features	Sentence count
41	Neg-Sentence negation	42
42	Neg-Negation of Verbal sentence	9
43	Neg-Constituent Negation	28
44	Neg-Multiple negatives	29
45	Neg-Universal Negation	73
46	Anaphora-Means of Expressing anaphora	38
47	Anaphora-Personal Pronoun	12
48	Anaphora-Reflexive Pronoun	14
49	Anaphora-Reciprocal Pronoun	7
50	Anaphora-Domains of anaphora	13
51	Reflexive-Means of expressing reflexivity-invariable form	52
52	Reflexive-Verbal affix	7
53	Reflexive-Position of the reflexive pronoun	4
54	Reciprocal-Means of expression	51
55	Reciprocal-Superlative	4
56	Comparison-Correlative comparison	19
57	Comparison-Superlative	5
58	Comparison-Comparison	1
59	Equative-Comparison	1
60	Equative-Means of expression	39
61	Possession-Possessive Noun Phrase	77
62	Emphasis-Non-contradictory emphasis	20
63	Emphasis-Contradictory emphasis	10
64	Emphasis-Particle	15
65	Emphasis-Constituent emphasis	15
66	Emphasis-Movement of emphasised element	24
67	Emphasis-Clefting	24
68	Emphasis-YES-NO Question	4
69	Statement-Simple Sentence	3
70	Emphasis-Simple Sentence	5
71	Emphasis-Reflexive Pronoun	1
72	Emphasis-Exclamatory	1
73	Sub-Subordinate clause to Main clause	1
74	HSMRs-Subordinate clause to Main clause	45
75	Topic-General expression	1
76	Interrogative-Wh-Question	1
77	Neg-General expression	1
78	MST-Greetings	17
79	MST-Exclamatory	11
80	MST-Brief answer and response	7
81	MST-Nonfinite verb in question	2
82	MST-Sentence Adverbial	5
83	MST-other	7
84	STs-Simple Sentence	1
85	InfM-Nominative Case	3
86	InfM-Accusative Case	11
87	InfM-Dative Case	7
88	InfM-Benefactive Case	2
89	InfM-Locative Case	29
90	InfM-Movement of emphasised element	2
91	InfM-Sociative Case	3
92	InfM-Simple Sentence	3

Sl.No.	Grammatical features	Sentence count
93	InfM-Ablative Case	6
94	InfM-Instrumental Case	5
95	InfM-Possessive Case	1
96	InfM-Expression of syntactic function- Intransitive or Transitive Verb	162
97	InfM-Nonlocal semantic function	310
98	InfM-Local semantic function	242
99	InfM-Location in time	162
100	InfM-Number Marking system	7
101	InfM-Noun class	12
102	InfM-Noun Phrase-Definiteness	17
103	InfM-Noun Phrase-Indefiniteness	17
104	InfM-Noun Phrase-Referential and non-referential Indefiniteness	11
105	InfM-Noun Phrase-Genericness	12
106	Pronoun-Personal Pronoun	23
107	Pronoun-Indefinite pronoun	12
108	Pronoun-Emphatic pronoun	2
109	Pronoun-Reflexive Pronoun	18
110	Pronoun-Reciprocal Pronoun	6
111	Pronoun-Interrogative Pronoun	7
112	Pronoun-Relative Clause	1
113	Pronoun-Demonstration	22
114	VInf-Voice	108
115	VInf-Present tense	78
116	VInf-Past tense	16
117	VInf-Future tense	38
118	VInf-Aspect	300
119	VInf-Mood	226
120	VInf-Causative	57
121	VInf-Defective verb	8
122	VInf-Finite verb	8
123	VInf-Nonfinite verb	7
124	VInf-Lexical verb as auxiliary	2
125	VInf-Verb Agreement-Person number gender	37
126	Adjective-Attributive Adjective	33
127	Adjective-Predicative Adjective	42
128	Adjective-Comparison	39
129	Adverb-Comparison	16
130	Adverb-Quality	14
131	Adverb-General expression	30
132	Adposition-Postposition	64
133	NumInflect-Ordinal with Nominal suffix	4
134	NumInflect-Numeral with Personal Noun	10
135	NumInflect-Quantifier	60
136	DevM-Prefixing	7
137	DevM-Derived Noun	9
138	DevM-Noun from verb	2
139	DevM-Nominal marked for tense	5
140	DevM-Deverbal Noun	5
141	DevM-Noun from Adjective	5
142	DevM-Derived verb	5
143	DevM-Derived Adjective	8
144	DevM-Derived Adverb	2

Sl.No.	Grammatical features	Sentence count
145	Comp-Noun noun compound	3
146	Comp-Part of	2
147	Comp-Time and season	3
148	Comp-Coordinate compounds	3
149	Comp-Echo compounds	2
150	Comp-Complex compound.	4
151	Comp-Verb compound	14
152	Comp-Conjunct Verb	4

Table 1: The linguistic features and corresponding sentence count of parallel corpus

This representative corpus contributes significantly to linguistic research and computational applications, offering a robust foundation for future studies in Indian language processing.

The corpus is developed using an online platform called TranKit. Freelance translators and reviewers are enlisted for each language after successfully passing a proficiency test. The coordination and assignment of translation and review tasks are managed by the LDC-IL staff.

In the first phase of corpus development, a total of 32,193 words of English source have been compiled. Currently, the corpus creation process has been completed for 147 mother tongues.

1.5 SUMMARY OF THE MOTHER TONGUE PARALLEL TEXT CORPUS OF INDIA

The Mother Tongue Parallel Text Corpus of India Vol.I comprising English and 147 mother tongues of India. Each corpus comprising a total of 5,332 sentences, systematically structured based on 152 grammatical categories. The following table gives the word count of each language along with the Translator/Reviewer who contributed to respective language.

Sl.No.	Language	Word Count	Translator/Reviewer
1	Assamese	26067	Angelina B. Dihingia
			Priyanshe Adhyapak
			Bikash Chetia
			Puja Das
			Udipta Saikia
2	Bengali	26880	Arpita Poddar
			Monalisa Paul
			Agnisha Majumder
			Kuntala Ghosh Dastidar
3	Bodo/Boro	25110	Dr. Bridul Basumatary
			Farson Daimary
			Dr. Bihung Brahma

Sl.No.	Language	Word Count	Translator/Reviewer
			Mahananda Brahma
4	Dogri	32008	Kuldeep Kumar
			Rahul Singh
			Diksha Choudhary
			Jatinder Singh
			Suhasi Dhawan
			Skyali Gupta
5	Gujarati	29044	Himani Kanojia
			Natwarlal Modha
			Dr. Rozy Patel
6	Hindi	27578	Neha Dixit
7	Kannada	21967	Sachin S
			Chaitra M N
			Anusha M H
			Rashmi H S
			Nayana H C
			Shauri BP
			Tejaswi G
8	Kashmiri	30821	Zahid Bashir Lone
			Sumeera Bashir
			Sajad Hussain Dar
			Javaid Ahmad Sofi
9	Konkani	25376	Ashalata Navelkar
			Yashwant Gawas
			Siddhi Gawas
			Pooja Tople
			Shweta Parab
			Sunita Gurudas Kanekar
10	Maithili	31316	Sharda Jha
			Dharmbir
11	Malayalam	20677	Dr. Resmi K. S.
			Anoop G. P
			Arya Navya A V
12	Manipuri	26625	Dr. Nameirakpam Amit
			Thaodem Romen Meitei

Sl.No.	Language	Word Count	Translator/Reviewer
			Dr. Bachaspatimayum Premabati Devi
			Dr. Thounaojam Harimohon Singh
13	Marathi	26327	Neha Satish Bandekar
			Arvind Ashok Tangadi
			Pratibha Jitendra Dongare
			Apurva Arvind Sambrekar
14	Nepali	27408	Rupesh Rai
			Padam Chhetri
			Sharmila Sharma
15	Odia	26127	Aditya Nayak
			Parthasarathi Dash
			Debamitra Mishra
16	Punjabi	33813	Rajwinder Kaur
			Sukhvinder Singh
			Manpreet Kaur
			Deepika Rani
17	Sanskrit	25812	Dr. Vinayak Bhat
			Varun R
			Vishwanath M V
18	Santhali	32916	Thakur Prasad Murmu
			Tarapada Soren
			Jayanta Tudu
19	Sindhi	33191	Heena Agnani
			Anita Narayan Kukreja
			Mohit Kukreja
			Mehek Wadhwani
20	Tamil	22483	Dr. Prabakaran P
			Dr. V. Alagumuthu
			Dr. Manimala J
21	Telugu	22822	Dr. Rajarao Dunna
			Ravinder Awgani
			P. Sujatha
22	Urdu	33414	Dr. Syed Majid Ali
			Dr. Md. Farhan
			Dr. Tasneem Zahera Haidry
23	Anal	30719	Seltun heberson anal
			DY Shuthahring Anal

Sl.No.	Language	Word Count	Translator/Reviewer
			Daryal Thawar Anal
24	Angami	28560	Rüübino Peseyie
			Thepfulenuo Mere
			Vetsino
			Vibeituonuo Mere
25	Apatani	33873	Bamin Lure
			Bamin Yalung
26	Are	22138	Darshini S
			Shruthi B B
			Likhith C Y
27	Awadhi	32074	Dr. Sumedha Shukla
			Vinod Kumar Tiwari
28	Bagheli/Baghel Khandi	32159	Abhyuday Pratap Singh
			Alankrita Singh
			Priyanshu Kushwaha
29	Bagri	32003	Sukhvinder Kaur
			Babbu Kaur
30	Bagri Rajasthani	32721	Dr. Mohan Lal
			Dr.Surinder Kumar
			Gurleen Kaur
			Ranjeet Kaur
31	Balti	32919	Nazir Hussain
			Anayatullah
32	Bhadrawahi	34686	Divya Rathore
			Hareesh Kumar
33	Bharmauri/Gaddi	32367	Shubh Karan
			Sunita Devi
34	Bhojpuri	31896	Anupama Tiwari
			Ravi Prakash
			Rajesh Kumar
35	Bilaspuri Kahluri	32913	Ajay Kumar
			Monika Kumari
36	Brajhasha	32885	Shyam Ratan
			Pramod Rathor
			Omprakash
37	Bundeli/Bundel khandi	32567	Anjana Kishanpuri
			Shivanee
38	Chakru/Chokri	28504	Thukuvelu Sakhamo

Sl.No.	Language	Word Count	Translator/Reviewer
			Huvo Thuluo
			Kukhrunelu Theluo
39	Chambeali/Chamrali	31823	Dr. Abdul Kreem
			Dr. Hardeep Singh
			Shahnaz Akhtar
40	Chang	27589	Mosha Mongko
			Narajungla Chang
			Bumo Takum
41	Chhattisgarhi	31972	Jayant Kumar Sahu
			Sanjeev Tiwari
			Dindayal Sahu
42	Chirr	30217	Nutsula
			Yokhumcha
43	Chungli	27987	Tianaro T Lemtur
			Pangernungla
44	Churahi	32193	Hasina
			Fazal Deen
			Mubarak Mohd
45	Coorgi/Kodagu	23540	Thashma K P
			Lavina G S
			Dr. Bhamini Raghavaiah K
46	Deori	24625	Navajyoti Deori
			Lila Kanta Deuri
47	Dhundhari	32067	Ramji Lal Bairwa
			Puran Mal Bairwa
48	Dimasa	23672	Dr. Krithika Barman
			Dr. Monali Longmailai
49	Gangte	39388	T Zoui Gangte
			Paulalson Gangte
			Lamgoulal Gangte
50	Garhwali	29449	Mamta Sharma
			Pradeep Bailwal
			Deepak Bijalwan
51	Garo	24478	Thamalisha W Sangma
			Ringchira G Momin

Sl.No.	Language	Word Count	Translator/Reviewer
52	Gojri/Gujjari/Gujar/	29139	Naseem Akhtar
			Mushtaq Khalid Chaudary
			Tariq Hussain Abrar
53	Gujari	34633	Azhar Nasir
			Asma Naseer
54	Halabi	35448	Shakuntala Tarar
			Vikram Kumar Soni
55	Handuri	32947	Hari Ram Dhiman
			Vivek Dhiman
56	Hara/Harauti	34546	Kishan Kumar
			Atul Kumar Jain
57	Haryanvi	35611	Dr. Rajinder Kumar
			Ravi kumar
			Kuldeep Gulia
			Ankit Maan
58	Hindi Multani	32085	Dr. Lata Khera
			Gurdatt Singh
59	Irula/Irular Mozhi	22901	Suresh P
			Srinivasan K
60	Kabui	32809	Lanbonlung Longmei
			Kamei Langangmei
61	Kachchhi	30804	Fatima Imtiaz Dhafrani
			Imtiaz Dhafrani
62	Kangri	31096	Ravinder Singh
			Nitish kumar
63	Karbi/Mikir	30399	Khorsing Teron
			Welsing Hanse
64	Khandeshi	27333	Gajanan Suresh Wankhede
			Urmila Patil
65	Khari Boli	32441	Aashi Agarwal
			Vaibhav Singh
			Anshul Garg
66	Khasi	41763	Dr. Kelleney Kitbok Suting
			Dr. P Marlon Brando Rani
			Rikynti L Ryntathiang

Sl.No.	Language	Word Count	Translator/Reviewer
67	Khezha	29372	Kewetsou Wetsah
			Koneite U Tsuzu
			Neikhrou Tsuhah
68	Khiemnungan	30306	Songmao M
			Heme
69	Khortha/Khotta	31157	Siddharth Sanket
			Dr. Shilpa
			Dr. Ritu Ghansi
			Basant Kumar
70	Kisan	30275	Nelson Ekka
			Premica Ekka
			Rewalina Ekka
71	Kodava	25049	Revathi
			K M Kusum
			Kuttappa M S
			M V Seetha
			Vidya M D
72	Kokbarak	27063	Jabanika Tripura
			Prasanta Tripura
			Salu Jamatia
73	Kolami	23108	Athram Mothiram
			Athram Rajkumar
			Tekam Tulsiram
74	Koli	27357	Gitanjali Mahesh Pagdhare
			Dr. Vanashri Pradip Phalake
			Oscar Philip Kinny
75	Kom	34843	K Songneihrang Kom
			L Chungneihmun Kom
76	Konda	24491	Janni Somi
			Boyi Gopala Rao
77	Konyak	34229	H Y Nyolong Konyak
			Eswa
78	Koya	22990	Dr. Suryanarayana Kalthi
			Irakam Ramesh
			Payam Srinu
79	Kudubi/Kudumbi	24206	Santhosh Kumar
			Shwetha K

Sl.No.	Language	Word Count	Translator/Reviewer
80	Kuki	32695	Lhaineilam
			Hengougouin Misao
			Chongpi
			Sehtinmang Kholhou
81	Kurmali Thar	28846	Hardeo Narayan Singh
			Bijraj Mahto
			Gyaneshwar Singh
82	Ladakhi	5611	Lobzang Tsering
			Dr. Neema Tashi
			Dr. Tsering Dolker
83	Lepcha	35457	Churmit Lepcha
			Nimkit Lepcha
84	Liangmei	30097	Lungphubou Abonmai
			Wichamdinbo
85	Limbu	23296	Kausila Subba
			Karnahang Limbu
			Pahang Limboo
86	Lotha	31941	Thungchopeni N Murry
			Yanpomo R Humtsoe
			Nzanmongi Z Ezung
			Dr. Yantsubeni Ngullie
87	Lyngngam	36404	Wandalin S Dkhar
			Balarihun Dkhar
			Resha Nongsiang
			Tyngshainhun Kharsnar
88	Magadhi/Magahi	32532	Dr. Anamika Kumari
			Madhulika
89	Malvi	32505	Parul Upadhyay
			Jaya Rajput
			Prachi Solanki
90	Mao	28806	Ashiihrii Khoziio
			Payia Maheo
			Kholi
91	Mara	35074	N Beithasia
			N Beizatha
			Loisy Khithie
92	Maram	31091	Ph Pungdila Grace

Sl.No.	Language	Word Count	Translator/Reviewer
			H Keren
			Lymba Solomon T
93	Maring	25055	Kanshouwa Susie
			Dangsawa Moran Maring
			K Florida Maring
			Ch Abose Moyon
			Shimlung Makunga Maring
94	Mech/Mechhia	25928	Ananta Champromary
			Dr. Bikash Narjinary
95	Mewari	31398	Dr. Reena Menariya
			Dr. Ripudaman Singh Ujjwal
96	Mewati	31265	Hanif Khan
			Ammar Khan
97	Miri/Mishing	23494	Dimbeswar Doley
			Bastav Lagachu
			Kalinath Panging
98	Mishmi	37171	Miju Mena
			Tiba Apralo
99	Mizo	37561	Dr. Wendy Lalhminghlui
			Catherine Lalhruaizeli
			Vanlalhruaizeli Chawngthu
100	Mongsen	31853	Amongkumba I kichu
			Imchalemla Longchar
101	Monpa	30449	Nima Dorjee
			Pema chotton
102	Mundari	31365	Dev Kumar Munda
			Amar Topno
			Hercules Singh Munda
103	Muwasi	30740	Arjun
			Sachin
104	Nawait	29111	Mohammad Zubair Jukaku
			Mohammed Rayees
105	Nimadi	32167	Nikita
			Ajay Patel
			Shivpal kalam
			Rohit Patel

Sl.No.	Language	Word Count	Translator/Reviewer
106	Nissi	32369	Taba Yami
			Likha Bai
			Nabam Tai Hina
107	Nocte	36632	Dr. Doli Tesia
			Dr. Dancha Tongluk
			Dr. Chalit Sumnyan
108	Pahari	34454	Syed Murtuza Hussain
			Dr. Syed Mahir Hussain Jafri
109	Paite	32627	Ching
			Ngaithianhoih
			Nempalching Hangshing
			Eddie Hnunrousiam Valte
110	Palmuha	31392	Parth Ranjan
			Shreekant Dubey
111	Pania	21702	Nandakumar S
			Anjali Bhaskaran
112	Paola	30276	L V Vaveiru
			K Dosou
			R S Japhet Khailunii
113	Pawari/Powari	31897	Dr. Tufan singh pardhi
			Dr. Shobha
114	Phom	26931	Rebecca Wandoi Phom
			V Shau-au
			Lily Phom
115	Pnar/Synteng	41979	Egira Shadap
			Juhhi Nanghuloo
			Lammidaka Pohleng
116	Pochury	31075	Chukhutho Nyusou
			Lojirho
			Lolia Ngouri
			Wuojisie Nyusou
117	Purkhi	31518	Sajjad Ali
			Mohd Issa
118	Rai	25530	Taraman Rai
			Dil Kumar Rai
			Bir Bahadur Rai
			Pratima Rai

Sl.No.	Language	Word Count	Translator/Reviewer
119	Rajasthani (Marwari)	31098	Dr. Madan Gopal Ladha
			Dr. Neeraj Daiya
			Dr. Sharada Krishna
120	Reang	36123	Rusan Reang
			Rotnojoy Reang
121	Rengma	33818	Sinyhunlo Kent
			Akenye Seb
			Kenyuni Kent
122	Rongmei	37767	K Gloria Phaomei
			Elisha Panmei
123	Sadan/Sadri	31618	Radheshyam Lakra
			Bibha Rani Topno
124	Sambalpuri	27776	Suman Rani Panda
			Astha Ayusha Mall
125	Sangtam	28852	Salome Sangtam
			Sethrila Thongtsar
126	Saurashtra/Saurashtri	22676	K. S. Senthilkumar
			K S Malini
127	Sema	30798	Pikali L Assumi
			Honikali Lohe
128	Shina	27916	Mohd Shafi
			Mukhtar Ahmad
129	Sirmauri	30265	Archna Kumari
			Nantram
130	Sugali	22780	Dr. Prasad Naik Mude
			Bukke santhoshnaik
131	Surjapuri	30396	Sanowar jahan
			Zoofishan khanam
			Khushnuma Begum
132	Talgalo (galo)	33204	Nivam Rekhung
			Jirken Bogo
133	Tangkhul	28408	Dr. Yatmi Luikham
			Dr. Maireiwon Ningshen
			Rinsomi Luikham
			Somiwon A Shishak
134	Thado/Thadou	33731	Dr. Mangyung Hemminlan Haokip
			Dr. Zamminlien

Sl.No.	Language	Word Count	Translator/Reviewer
			Mongngaichong Touthang
			Boishi Neijalhei khongsai
135	Tibetan	5497	Dr. Lhamo Tso
			Lobsang Gendun
136	Tikhir	30057	Shiutsu K Thongliu
			Longtsu k lams
137	Tripuri	32707	Kwrwng Tripura
			Salka Tripura
			Sushil Tripura
			Bhabesh Tripura
138	Tulu	23221	Sai Geeetha
			Yogitha Shetty
			Soumya Rao
			Dr. Rajashree
			Pradyoth Hegde
139	Vaiphei	39388	P Damlalson Vaiphei
			Genthianson Vaiphei
			P Felboi Vaiphei
140	Wagdi	31761	Hoshang Panchal
			Kushagra Panchal
141	Wancho	34532	Lemnon Wangjen
			Omwang wangjen
142	Yerava	24423	Poovaiiah M P
			Accamma M S
143	Yerukala/Yerukula	22359	E Mahalaxmi
			S Pravallika
144	Yimchungre	28489	Nagayimla Yimchunger
			Hantsula
145	Zeliang	33850	Lungchiepyile
			Machipeung Thou
			Kangzangyile Khate
146	Zemi	36443	Atule Heikha
			Ichilungle
			Iraicule
			Haidobabe Hingleu
147	Zou	35471	M Joseph Khamgoulia Zou
			Thangsianlal William

Table 2:Contributors to the Mother Tongue Parallel Text Corpus of India

1.6 REFERENCE

1. Comrie, Bernard & Smith, Norval. 1977. *Lingua Descriptive Studies: Questionnaire*. Amsterdam: North-Holland Publishing Company.
2. Abbi, Anvita. 2001. *A Manual of Linguistic Field Work and Structures of Indian Languages*. Muenchen: Lincom Europa.
3. Bhattacharya, Krishna et al. (Eds.) 2016. *Language Documentation Handbook*. Mysore: CIIL.
4. Sivashanmugam, C. & Thayalan, V. 2012. (Comp.) *Comprehensive Questionnaire for Tribal Studies*. Coimbatore: Dept. of Linguistics, Bharathiar University.