

**Dr. Narayan Kumar Choudhary**  
Lecturer cum Junior Research Officer, Central  
Institute of Indian Languages, Ministry of  
Human Resource Development, Department of  
Higher Education, Government of  
India, Mysuru – 570006  
Email: nchoudhary.ciil AT gmail {dot} com  
Phone (O): 0821-234-5007 / 0821-234-5092  
Personal Weblog: [Narayan Choudhary's Blog](#)



## Research Interests

Computational Linguistics, Natural Language Processing (NLP) and Language Technology in general. Corpus Linguistics, Data Science, Part of Speech Annotation, Syntactic/Dependency Parsing, and related areas.

## Other interests

Language Typology, Language Documentation, Language Policy and Planning, General Linguistics.

## Current Institutional Responsibilities

Assistant Director (Admin) / Head of Administration (Since August 1, 2016 till date)  
Officer in-Charge, Linguistic Data Consortium for Indian Languages (LDCIL), Since May 2017 till date).

## Publications

### Books

1. [Proceedings of the Third Students' Conference of Linguistics in India](#) (SCONLI-3). 2011. ed. with Gibu Sabu M., Parimal Publishers, New Delhi.
2. [Indian Language Part-of-Speech Tagset: Hindi](#). 2010. Co-authored by Kalika Bali, Monojit Choudhury, Priyanka Biswas, Girish Nath Jha, Maansi Sharma. Linguistic Data Consortium, Philadelphia. (This is actually a PoS Annotated corpus of Hindi general domain text)
3. Choudhary, Narayan (ed.). 2019. [Linguistic Resources for AI/NLP in Indian Languages](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-295-8.
4. Choudhary, Narayan. 2018. [Cost Analysis of Linguistic Resources](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-283-5.
5. Ramamoorthy, L., Narayan Choudhary, Sonali Sutradhar, Arundhati Sengupta, Sankarshan Dutta, Priyanka Das & Saswati Karmakar. 2019. [A Gold Standard Bengali Raw Text Corpus](#). Central Institute

of Indian Languages, Mysore. 978-81-7343-204-0.

6. Ramamoorthy, L., Narayan Choudhary, Bridul Basumatary & Farson Daimary. 2019. [\*A Gold Standard Bodo Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-209-5.

7. Ramamoorthy, L., Narayan Choudhary & Sunil Kumar. 2019. [\*A Gold Standard Dogri Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-213-2.

8. Ramamoorthy, L., Narayan Choudhary, Mona Parakh, Purva S Dholakia., Gadhavi R Hiren & Maheshkumar R Solanki. 2019. [\*A Gold Standard Gujarati Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-214-9.

9. Ramamoorthy, L., Narayan Choudhary, Jitendra Kumar Singh, Richa, Anjali Sinha, Dheeraj Kumar Mishra, Arimardan Kumar Tripathi, Aditi Debsharma, Satyaendra Kumar Awasthi & Madhupriya Pathak. 2019. [\*A Gold Standard Hindi Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-219-4.

10. Ramamoorthy, L., Narayan Choudhary, Vijayalaxmi F. Patil, Chetan Suryakant Baji, Malini N Abhyankar, Rajesha N. & Manasa G. 2019. [\*A Gold Standard Kannada Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-226-2.

11. Ramamoorthy, L., Narayan Choudhary & Shahid Mushtaq Bhat. 2019. [\*A Gold Standard Kashmiri Raw Text Corpus.\*](#) A Gold Standard Kashmiri Raw Text Corpus. Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-230-9.

12. Ramamoorthy, L., Narayan Choudhary, Saurabh Varik, Rashmi Shet Tanawade & Yashwant D Gawas. 2019. [\*A Gold Standard Konkani Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-232-3.

13. Ramamoorthy, L., Narayan Choudhary, Arun Kumar Singh & Dinesh Mishra. 2019. [\*A Gold Standard Maithili Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-236-1.

14. Ramamoorthy, L., Narayan Choudhary, Saritha S.L., Rejitha K.S. & Sajila S. 2019. [\*A Gold Standard Malayalam Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-240-8.

15. Ramamoorthy, L., Narayan Choudhary, Amom Nandaraj Meetei, Yumnam Premila Chanu, Longjam Anand Singh & M. Bidyarani Devi. 2019. [\*A Gold Standard Manipuri Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-245-3.

16. Ramamoorthy, L., Narayan Choudhary, Gajanan R Apine & Apurva P Betkekar. 2019. [\*A Gold Standard Marathi Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-250-7.

17. Ramamoorthy, L., Narayan Choudhary, Samar Sinha, Jeena Rai, Umesh Chamling Rai & Rupesh Rai. 2019. [\*A Gold Standard Nepali Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-255-2.
18. Ramamoorthy, L., Narayan Choudhary, Raja Kumar Naik, Pramod Kumar Rout, Kshirod Kumar Das & Santosh Kumar Mohanty. 2019. [\*A Gold Standard Odia Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-258-3.
19. Ramamoorthy, L., Narayan Choudhary, G. Palanirajan, S. Thennarasu, Prem Kumar L. R, Amudha R., Prabakaran R., Vijayan N. & M. Ramesh Kumar. 2019. [\*A Gold Standard Tamil Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-266-8.
20. Ramamoorthy, L., Narayan Choudhary, Thirupal C Reddy & Gangaraju H. 2019. [\*A Gold Standard Telugu Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-271-2.
21. Ramamoorthy, L., Narayan Choudhary, Mansoor Khan, Shahnawaz Alam, Bi Bi Mariyam & Rushda Idris Khan. 2019. [\*A Gold Standard Urdu Raw Text Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-274-3.
22. Ramamoorthy, L., Narayan Choudhary, Sonali Sutradhar, Priyanka Biswas, Arundhati Sengupta, Sankarshan Dutta & Priyanka Das. 2019. [\*Bengali Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-206-4.
23. Ramamoorthy, L., Narayan Choudhary, Bridul Basumatary & Farson Daimary. 2019. [\*Bodo Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-211-8.
24. Ramamoorthy, L., Narayan Choudhary, Jitendra Kumar Singh, Richa, Anjali Sinha, Dheeraj Kumar Mishra, Arimardan Kumar Tripathi & Satyaendra Kumar Awasthi. 2019. [\*Hindi Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-221-7.
25. Ramamoorthy, L., Narayan Choudhary, Vijayalaxmi F. Patil, Chetan Suryakant Baji, Malini N. Abhyankar, Rajesha N. & Manasa G. 2019. [\*Kannada Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-228-6
26. Ramamoorthy, L., Narayan Choudhary, Saurabh Varik & Rashmi Shet Tanawade. 2019. [\*Konkani Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-234-7.
27. Ramamoorthy, L., Narayan Choudhary, Arun Kumar Singh, Dinesh Mishra & Atuleshwar Jha. 2019. [\*Maithili Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-238-5.
28. Ramamoorthy, L., Narayan Choudhary, Gajanan R Apine & Apurva P Betkekar. 2019. [\*Marathi Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-251-4.
29. Ramamoorthy, L., Narayan Choudhary, Samar Sinha, Jeena Rai, Umesh Chamling Rai & Rupesh Rai. 2019. [\*Nepali Raw Speech Corpus.\*](#) Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-255-2.

30. Ramamoorthy, L., Narayan Choudhary, Poonam Dhillon & Sarbjeet Kaur. 2019. [Punjabi Raw Speech Corpus](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-264-4.
31. Ramamoorthy, L., Narayan Choudhary & Rajesha N. 2019. [Telugu Raw Speech Corpus](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-272-9.
32. Ramamoorthy, L., Narayan Choudhary, Mansoor Khan, Shahnawaz Alam, Bi Bi Mariyam & Rushda Idris Khan. 2019. [Urdu Raw Speech Corpus](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-276-7.
33. Ramamoorthy, L., Narayan Choudhary, Poonam Dhillon, Sarbjeet Kaur & Sandeep Singh. 2019. [A Gold Standard Punjabi Raw Text Corpus](#). Central Institute of Indian Languages, Mysore. ISBN: 978-81-7343-262-0.

### Research Papers in Refereed Journals/Conferences

1. Parth Pathak, Pinal Patel, Vishal Panchal, Sagar Soni, Kinjal Dani, **Narayan Choudhary**, Amrisha Patel. 2015. [ezDI: A Supervised NLP System for Clinical Narrative Analysis](#). In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Denver, Colorado. (Accorded first rank in the shared task).
2. Neha Dixit and **Narayan Choudhary**. 2014. [Evaluating Two Annotated Corpora of Hindi Using a Verb Class Identifier](#). In Proceedings of ICON 2014. Goa University, Goa (To appear in ACL Anthology).
3. Neha Dixit and **Narayan Choudhary**. 2014. [Automatic Classification of Hindi Verbs in Syntactic Perspective](#). International Journal of Emerging Technology and Advanced Engineering, Volume 4, 8th Issue. (ISSN 2250 – 2459 (Online))
4. Parth Pathak, Pinal Patel, Vishal Panchal, **Narayan Choudhary**, Amrisha Patel, Gautam Joshi. 2014. [ezDI: A Hybrid CRF and SVM based Model for Detecting and Encoding Disorder Mentions in Clinical Notes](#). In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014 Shared Task, awarded third best result). Dublin, Ireland. ISBN 978-1-941643-29-7
5. **Narayan Choudhary**, Parth Pathak, Pinal Patel, Vishal Panchal. 2014. [Annotating a Large Representative Corpus of Clinical Notes for Parts of Speech](#). 2014. In: Proceedings of 8th Linguistic Annotation Workshop, Dublin, Ireland. ISBN 978-1-941643-29-7
6. **Narayan Choudhary**, Girish Nath Jha. 2011. [Creating Multilingual Parallel Corpora in Indian Languages](#). 2011. In Proceedings of the 5th Language & Technology Conference, Poznan, Poland. (awarded the Best Student Paper)
7. **Narayan Choudhary**, Girish Nath Jha and Pramod Pandey. 2011. [A Rule based Method for the Identification of TAM features in a PoS Tagged Corpus](#). In Proceedings of the 5th Language &

Technology Conference, Poznan, Poland.

8. **Narayan Choudhary**. 2011. [Web-drawn corpus for Indian Languages: A Case of Hindi](#). In Proceedings of Information Systems for Indian Languages. Volume 139, Part 2, 218-223. Springer Verlag.

9. **Narayan Choudhary**. 2008. [बोधात्मक भाषाविज्ञान](#), in Gaveshanaa, April-June, 2008 vol.:90/2008 Central Institute of Hindi, Agra. pp.:11-18 (This is a Hindi translation of the article “Cognitive Linguistics” from Encyclopedia of Linguistics by Gilles Falkner, 2006)

10. Narayan Choudhary. 2007. [Syllable Structure of Great Andamanese](#), November, 2006. In proceedings of National Seminar on Perspectives in Linguistics, Kashmir University, Srinagar, Kashmir. India. Pp. 141-146

11. **Narayan Choudhary**, Anvita Abbi, Girish Nath Jha. 2007. Morphological Analyzer for Great Andamanese Verbs: Implementing a Concatenative Template. In Vishwabharat ( April 2007 - January 2008 Journal) TDIL, New Delhi, pp.113-118 [http://tdil.mit.gov.in/april-jan-2008/8.8\\_Morphological\\_analyzer.pdf](http://tdil.mit.gov.in/april-jan-2008/8.8_Morphological_analyzer.pdf)

## INVITED TALKS

1. “NLP and Information Extraction”, SCONLI-07, Aligarh Muslim University, Aligarh, 8-10 February, 2013
2. Orientation Course in Computational Linguistics, Tezpur University, Assam. 21-23 December, 2014
3. Translation Process State, Knowledge Commission of Gujarat, Ahmedabad. 01-02-2016
4. Current Trends in Translation Industry, National Translation Mission, CIIL, Mysore. 03-03-2017
5. Translation Practices in the IT Industry, National Translation Mission, CIIL, Mysore. 04-03-2017
6. Introduction to Translation, National Translation Mission, CIIL, Mysore. 16-05-2017
7. ICT based teaching of English Training Programme for Secondary Teachers (AP, Telangana and TN), 3-09-2017 RIE, NCERT, Mysore
8. Ways of Empowering Language using Language Technology, Language Across Curriculum in Multi-Linguistic Context: Scope and Challenges, 29-31 January, 2018 RIE, Mysore

## AWARDS

1. [Ranked 1st in](#) “SemEval-2015 Task 14: Analysis of Clinical Text”, to occur in NAACL-2015, Denver, Colorado.
2. Ranked 3rd in “SEMEVAL 2014: Shared Task 7: Analysis of Clinical Text”, Dublin, Ireland
3. Best Student Paper Award for the paper titled “Creating Multilingual Parallel Corpora in Indian Languages” at LTC’11, Poznan, Poland
4. UGC-NET Lectureship Award, 2003 and 2004

## Past Research & Work Experiences

1. NLP Research Engineer, ezDI, LLC.: July, 2012 – February, 2016
2. Senior Linguist, Shallow Parser Tools for Indian Languages, JNU, New Delhi: May, 2012 – June, 2012
3. Senior Linguist, Indian Languages Corpora Initiative (ILCI), JNU, New Delhi: March, 2009 – May, 2010
4. Teaching Assistant, Centre for Linguistics and Special Centre for Sanskrit Studies, JNU, New Delhi: August, 2007 - July, 2009
5. Research Assistant, Centre for Linguistics, JNU, New Delhi: August, 2005-July, 2007
6. Project Associate, CSE, IIT Kanpur: May, 2005-June, 2005

## Educational Qualifications

1. Ph. D. Thesis Title: [Automatic Identification and Analysis of Verb Groups in Hindi](#). JNU. 2006-2011.
2. M. Phil. Dissertation Title: [Developing a Computational Framework for the Verb Morphology of Great Andamanese](#). JNU, New Delhi, 2006.
3. National Eligibility Test for Lectureship (NET -December, 2003; June, 2004) of the UGC
4. Masters (Linguistics), JNU, New Delhi, 2004. MA Thesis: [Word Order in Pnar \(Jaintia\)](#)
5. Bachelor of Arts (English Hons., Economics, History, Hindi), LNMU, Darbhanga, 2001.

## Memberships

[Linguistic Society of India](#), Life Member

[Association of Computing Machinery](#), Student Member, 2010-2013

## Other Info

### Computing Skills

Platforms: Well versed with Windows and Linux (Ubuntu/RHEL)

Development Environment: MySQL 5; PHP, JAVA, C++, Perl, Prolog, LISP, CSS, HTML

### Languages

Well Versed with Expertise: English, Hindi, Maithili

Academic Knowledge: Sanskrit, Pnar (Jaintia), Great Andamanese, Gujarati

### Hobbies

Reading, Writing, Music, Yoga, Swimming, Mountaineering.