

# Machine Translation - The saara Approach

Kavi Narayana Murthy



School of Computer and Information Sciences  
University of Hyderabad

January 24, 2019

# Contents 1

- 1 Machine Translation
- 2 The saara Approach to Machine Translation
- 3 The saara Machine Translation Development Framework
- 4 Conclusions

## MT

- There are more than 7000 Languages in the World
- Relevant/Useful/Interesting Materials exist in many of these
- Manual Translation is Slow, Tedious, Impracticable
- Machine Translation (MT) saves Time, Effort and Money
- R&D in MT dates back to 1950s
- R&D in India started in early 1990s
- Quality is a big issue
- Lab-models exist but usable systems are rarely seen

## MT

- Rule Based Systems (based on Linguistics) have not been very successful - they quickly saturate
- Data Driven approaches such as Statistical MT (SMT) and Neural MT (NMT) require large amounts of Parallel Corpus - not always available
- Quality is the Big Challenge. MT output is rarely Post-Editable. 'Mistakes' cannot be easily corrected.

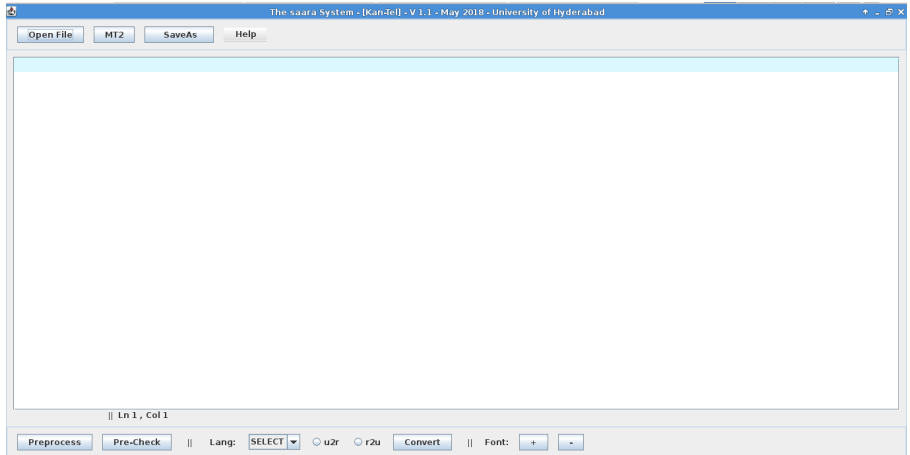
# Contents 2

- 1 Machine Translation
- 2 The saara Approach to Machine Translation**
- 3 The saara Machine Translation Development Framework
- 4 Conclusions

# The saara Approach

- Saara Approach: Simple. Fast. Easily Post-Editable. Mistakes in the Dictionaries can be easily corrected. Does not require Linguistics. Does not Require Parallel Corpus. MT systems can be developed quickly. Semi-Automatic. Still saves Time, Effort and Money. Suitable for Modern Prose only.
- Kan-Tel: Ready to Use. Tel-Kan: Coming soon.
- Other Language Pairs: The saara MT Development Framework (SMTDF)

# Kannada-Telugu Translator



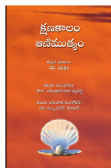
# Kannada-Telugu Translator

- Steps: Pre-Check, Translate, Post-Edit
- Pre-Check: Manual. Takes about 20 min/Page. One person can do 20 pages a day, 500 pages a month.
- Translate: Automatic. Speed: 1,00,000 Sentences per Second!
- Gives 85 to 95 % Translation Performance
- Post-Edit: 15 min/Page, One person can do 20+ pages a day, 500+ pages a month.
- Total Cost: Comparable or Less Than manual translation
- Quality improves with usage, time and cost come down



## Kannada-Telugu Translator: Case Studies

- “Kshana Hottu Aani Muttu” by Sri S Shadakshari. Translated using SAARA, Post-Edited by Dr. Pammi Pavan Kumar, and Released



- “Shakuni” by Dr. Tirumala Kulakarni - Translated and Post-Edited, being finalized

# Performance Evaluation

Score	Meaning
-1	No Output OR buffer clearance issue
0	Nonsense ( (If the sentence doesn't make any sense at all, it is like someone speaking to you in a language you don't know) )
1	Some parts make sense but is not comprehensible overall (e.g., listening to a language which has lots of borrowed words from your language, you understand those words but nothing more)
2	Comprehensible but has quite a few errors (e.g., someone who can speak your language but would make lots of errors. However, you can make sense out of what is being said)
3	Comprehensible, occasional errors (e.g., someone speaking Hindi getting all its genders wrong)
4	Perfect (e.g., someone who knows the language)

Table: Scheme of Evaluation - Comprehensibility

# Contents 3

- 1 Machine Translation
- 2 The saara Approach to Machine Translation
- 3 The saara Machine Translation Development Framework**
- 4 Conclusions

# SMTDF

- SMTDF: The saara Machine Translation Development Framework
  - ▶ Supports Fast Development of High Performance MT Systems - for any two languages
  - ▶ Knowledge of the Two Languages is Enough
  - ▶ ML Algorithms keep Learning as the system gets used
  - ▶ Includes a Meta layer with AI capabilities for monitoring, automatic adaptation, etc.

# SMTDF

The saara Machine Translation Development Framework - Version 1.0 - October 2018  
University of Hyderabad

User: ramal5

Telugu-Kannada

Logout

Please type/paste Telugu text below or Upload a Telugu plain-text file in utf-8 format and click Translate.

Upload input file:

Browse...

No file selected.

Translate

[User Manual](#)

[License Terms](#)

[Acknowledgements](#)

[Contact Us](#)

# SMTDF

The saara Machine Translation Development Framework - Version 1.0 - October 2018  
University of Hyderabad

User: ramal5

Telugu-Kannada

Logout

Sentence No: 1

Please check the following words carefully and correct if required. There may be spelling errors, meaning given may be wrong or inappropriate for the current sentence.

గోమయంతో ఆ ప్రాంతాన్ని ఆటంకపొదా అనాదిగా వస్తున్న అనాయ

EDIT

గోమయంతో	గోమయంతో
ఆ	ఆ
ప్రాంతాన్ని	ప్రాంతాన్ని
ఆటంకపొదా	ఆటంకపొదా
అనాదిగా	అనాదిగా
వస్తున్న	వస్తున్న
అనాయ	అనాయ

CONTINUE

User Manual

License Terms

Acknowledgements

Contact Us

# SMTDF

The saara Machine Translation Development Framework - Version 1.0 - October 2018  
University of Hyderabad

User: rama15

Telugu-Kannada

Logout

Sentence No: 5

Please check the translation carefully and reorder the words if required.

EDIT

ದಾಳುಗಳು , ಪುಸ್ತಕಗಳು ಮಾದ್ಯತೆ ಒಡವರಿಗೆಷ್ಟು ಹಾನಿ ಮಾಡುತ್ತಿದ್ದಾರೋ ಗೊತ್ತು !

Reordering - Select a word and click left or right to reorder.

ದಾಳುಗಳು , . ಪುಸ್ತಕ\_ಕೆಲಸಗಳು ಮಾದ್ಯತೆ ಒಡವರಿಗೆಷ್ಟು ಹಾನಿ ಮಾಡುತ್ತಿದ್ದಾರೋ ಗೊತ್ತು !

<< Shift Left Shift Right >>

POSTEDIT

NEXT SENT

[User Manual](#)

[License Terms](#)

[Acknowledgements](#)

[Contact Us](#)

# SMTDF

The saara Machine Translation Development Framework - Version 1.0 - October 2018  
University of Hyderabad

User: rama15

Telugu-Kannada

Logout

Sentence No: 7

Please check the following words carefully and correct if required. There may be spelling errors, meaning given may be wrong or inappropriate for the current sentence.

వర్షాలు కురువటం వలన నానారోజులైన ధాన్యాలు పెటలు మృత్యుగా మడుతాయి

EDIT

వర్షాలు	
కురువటం	
వలన	
నానారోజులు	నానా...బెడద
ధాన్యాలు	ధాన్య గిటు
పెటలు	బోలెగల
మృత్యుగా	మృత్యురూపిణి
మడుతాయి	బోలెయ్యడ

CONTINUE

User Manual

License Terms

Acknowledgements

Contact Us



# Contents 4

- 1 Machine Translation
- 2 The saara Approach to Machine Translation
- 3 The saara Machine Translation Development Framework
- 4 Conclusions

# Conclusions

High Performance MT is feasible provided it is semi-automatic (human assisted). MT systems can be developed quickly using the saara MTDF - only knowledge of the two languages is required, linguistic expertise and resources are not absolutely essential.

Thank You

*Thank You*

email: [knmuh@yahoo.com](mailto:knmuh@yahoo.com)

web: [www.LanguageTechnologies.uohyd.ac.in](http://www.LanguageTechnologies.uohyd.ac.in)