NLP: A Paninian Perspective

Felicitation to Prof KV Ramakrishnamacharyulu

Rajeev Sangal
Vineet Chaitanya

2 August 2012

Outline of Talk

- **A bit of history**
- **Research themes**
- **Meta-themes which guided the work**

Prof KV Ramakrishnamacharyulu

Key member of the Akshar Bharati team

- **The journey of Akshar Bharati started 25 years ago, around 1984, in association with Prof KV Ramakrishnamcharyulu**
- **Connected with many institutions**
  - **Started at IIT Kanpur, connected with Rashtriya Sanskrit Vidyapeetha, Tirupati, Univ. of Hyderabad and IIIT Hyderabad**

Who is *Akshar Bharati*?

Personification of a group working on:

- **Computer processing of Indian languages**
- **Giving due importance to the traditional Indian theories of language**
- **Team work (leading to personification)**

Spirit of Akshar Bharati

The following relate to the spirit:

- **Bourbaki**
- **Durga**

Spirit of Akshar Bharati - Bourbaki

Nicolas Bourbaki:

- **Pseudonym for a group of mainly French mathematicians starting in 1935**
- **Wrote a series of books presenting an exposition of modern advanced mathematics**
- **Founding all of mathematics on set theory**
- **Rigour and generality**
- **Led to the discovery of several concepts and terminologies**

Spirit of Akshar Bharati - Durga: Mahishasur Mardini

Created by everybody giving their best:

- **Shiva - the trident**
- **Vishnu - the disc (chakra)**
- **Varuna - the conch**
- **Agni - the spear**
- **Yama - the cudgel**
- **Vayu - the bow**
- **Surya - the arrows**
- **Indra - the vajra**
- **Kubera - the mace**
- **Brahma - the water pot**
- **Kala - the sword**
- **Vishwakarma - the axe**
- **Himavaan - a mountain lion as her vehicle**

Major Event in 1985

Rick Briggs' paper appeared on Sanskrit and Knowledge Representation in AI Magazine in Dec. 1984. Featured on cover.

- **Paper created quite a stir in the world!**
- **CSI (Computer Society of India) decided to hold an international conf. called KRIS-85 (Knowledge Representation and Inference in Sanskritam). People involved:**
  - **Swami Parmanand Bharati, Head, Sankar Mutt, Bangalore. (Earlier Prof of Theoretical Physics at IITM)**
  - **Prof HN Mahabala, President, CSI (Also Prof of CS at IITM. Earlier Prof of CS at IITK)**

KRIS-85 Preparatory Workshop

Preparatory Workshop held at Bangalore Aug. 1985. VC (Vineet Chaitanyaji) attended it. Met:

- **Prof. KV Ramakrishnamacharyulu, faculty member, Kendriya Sanskrit Vidyapeeth (later RSVP), Tirupati**
  - **Asked to translate Rick Briggs' paper from English to Sanskrit**

KRIS-85 Conference

A stellar gathering of scholars from India and abroad. Dec. 1985. ECG Sudarshan, among others.

- **AI, NLP researchers**
- **Vaiyaakaranas, Naiyaayikas, and**
- **Plain Computer Scientists, and plain Sanskritists**

People - KV Ramakrishnamacharyulu

- **Faculty member, scholar in Vyakarana, Kendriya Sanskrit Vidyapeeth, Tirupati.**
  - **PhD in Sanskrit Vaiyakarana Bhushansara – a text on shaabdabodha**
  - **Participated in Sanskrit Bharati (spoken Sanskrit)**
- **Got interested in looking at Computer and Sanskrit. Got encouragement (blessings) from scholars in the field:**
  - **Tatachar-ji, Renowned scholar of Nyaya, Vice Chancellor, RSVP Tirupati**
  - **Peri Suryanarayana Shastri, Renowned scholar of Vyakarana**
  - **Prahaladachar-ji, Nyaya scholar, later Vice Chancellor, RSVP Tirupati**
- **Regularly visited IIT Kanpur from 1986 onwards. Part of Akshar Bharati since then.**

Research Themes

Research Themes

1. **Panini applied to modern Indian languages (ILs)**
2. **Parsing of sentences**
3. **Theoretical comparisions with contemporary computational frameworks**
4. **Machine Translation**

* **Relation of Sanskrit with other Theoretical Sciences besides Computer Science**

Theme 1: Panini Applied to Modern ILs (1984-90)

- **Morphology through paradigms**
- **Vibhakti represents verbal or nominal ending together with function words (also called local word grouping)**
- **Karaka charts plus transformations**
  - **Key insight: Kiparsky's paper - Abhihite. Verbal endings control noun vibhakti in Sanskrit.**
    - **Modern Indian languages follow the same principle!**
  - **Notion of subject vs. karta, and agent vs. karta are different!**
    - **When to translate a term and when not to!**
      Analogy: Energy vs. power, or Force vs. pressure.
- **Lakshan charts plus ontology**

Parsing Problem

- **laDake ne laDakii ko phool diyaa.**
- **boy -erg girl -dat flower gave**
- **The boy gave a flower to the girl.**

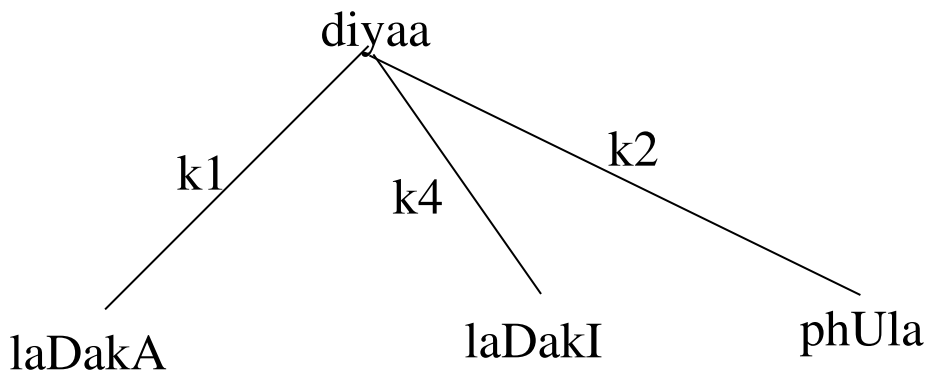


Figure: Modifier-Modified Dependency Tree (Visheshan-visheshya bhaava in shaabda bodha)

Vibhakti to Karaka-Role Mapping

For verb 'de' (give)::

- **Vibhakti: Karaka role**
- **ne: karta (like agent thematic role)**
- **ko: sampradaana (like beneficiary)**
- **0: karma (like patient)**

Mapping Between Vibhakti and Karaka Level

Karaka chart for de[tA] (give[habit.])

karaka	necessity	vibhakti
karta	mandatory(m)	0
karma	m	0
sampradan	m	ko
karana	opt	se

Problem in Mapping

- **laDakaa laDakii ko phool de rahaa hE.**
- **boy-0 girl -ko flower-0 give -ing is**
- **The boy is giving a flower to the girl.**

- **laDake ko phool denaa padaa**
- **(give had-to)**
- **The boy had to give a flower.**

- **laDake se phool diyaa nahiiM gayaa**
- **(give not could)**
- **The boy could not give the flower.**

Karaka Chart Transformation

Based on TAM (tense-aspect-modality) labels

TAM label	karaka chart transformation
_yA	karta.vibhakti:= ne
_na_paDA	karta.vibhakti:= ko
_yA_gayA	karta.vibhakti:= se

The transformation is:

- **Independent of verbal root**
- **Depends on TAM only**

Published in Journal of Indian Linguistics 1991, KBCS-91, COLING-1990.

Theme 2: Parsing (1988-93)

- **Integer programming based solution (VC, PB)**
  - **Theoretically equivalent to the assignment problem or bi-partite graph matching problem**
  - **Solution in polynomial time**
  - **ACL-1993 paper. Could not go, no funds.**
- **Nested constraints (projective constraints) (NLPRS-1995 paper)**

Constraints

1. **For each of the mandatory karakas in a karaka chart for each demand group, there should be exactly one outgoing edge labelled by the karaka from the demand group.**

2. **For each of the desirable or optional karakas in a karaka chart for each demand group, there should be at most one outgoing edge labelled by the karaka from the demand group.**

3. **There should be exactly one incoming arc into each source group.**

Constraint Parser using Integer Programming

1. **For each demand group i, for each of its mandatory karakas k, the following equalities mush hold:**

$$M_{i,k} : \sum_{j} x_{i,k,j} = 1$$

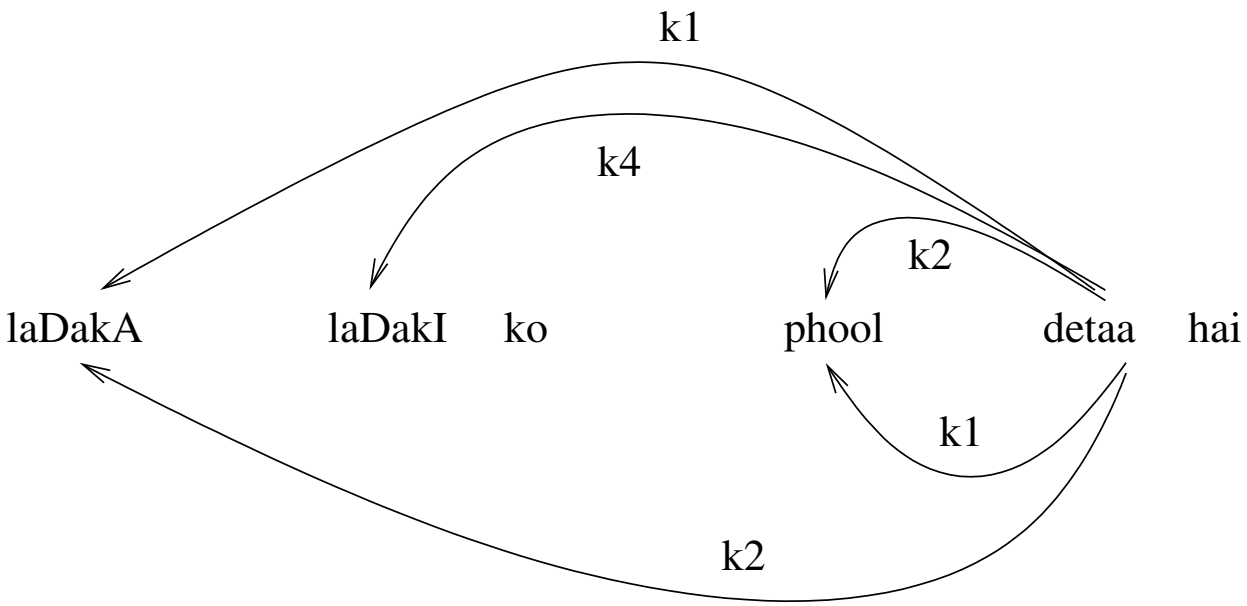
2. **For each demand group i, for each of its optional or desirable karakas k, the following inequalities must hold:**

$$O_{i,k} : \sum_{j} x_{i,k,j} \leq 1$$

3. **For each of the source groups j, the following equalities must hold:**

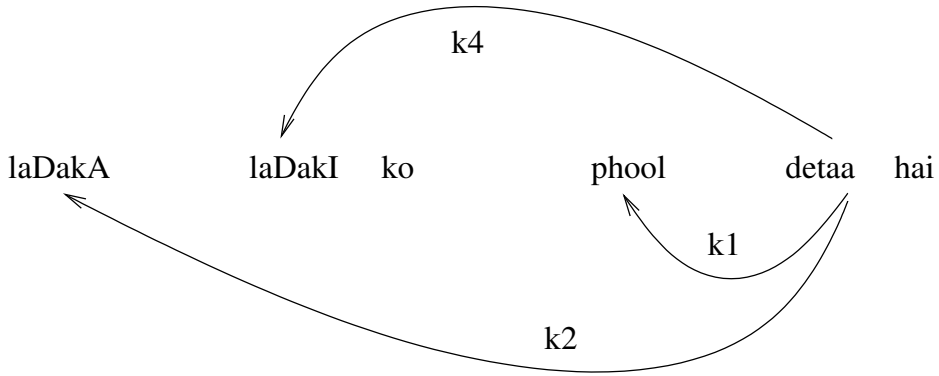
$$S_j : \sum_{i,k} x_{i,k,j} = 1$$

Example : Constraints Graph

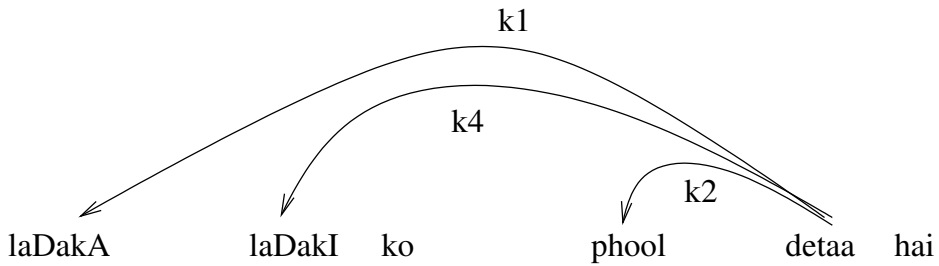


Example : Solutions

- **Solution 1**



- **Solution 2**



Theme 3: Theoretical Comparisions with other Computational Language Frameworks (1985-94)

Study of contemporary computational linguistic frameworks:

- **Lexical Functional Grammar (Bresnan)**
- **Generalized Phrase Structure Grammar (Gazdar)**
- **Tree Adjoining Grammar (Joshi)**
- **Government and Binding (Chomskyan)**

Ministry of IT (Dept. of Electronics) gave funds for teachers training programme for NLP, 1991-95.
IITK conducted many such courses over 4 years. Became a ground for debates with linguistis, and development of theory.

Major Theoretical Differences

- **Dependency grammar vs phrase structure grammar (PSG)**
  - **Rest of the computational world was using PSG**
  - **Today, DGs are used extensively**
- **Notion of karta distinct from subject or agent**
* **Guiding principle: Simplicity of framework, Closeness to intuition.**

Theme 4: Machine Translation (1986-89)

Hindi to Telugu MT system (lab. prototype) [Bhanumati, M Phil thesis, KU, 1989]

- **Ideas from CPG theoretical framework implemented, otherwise a toy system**
- **Idea of lakshan charts (discrimination nets for word sense disambiguation)**
- * **Project connected people to us**
  - **Thakur Dass, Kendriya Hindi Sansthan**
  - **Suraj Bhan Singh, Kendriya Hindi Sansthan. Through his book: Hindi kaa vaakyaatmak vyakaran**
  - **Vidya Niwas Mishra, VC, Sampurnanda Univ.**

Theme 5: Relation of Sanskrit with other Theoretical Sciences besides NLP

- **DST project led by Navjyoti Singh, with about a dozen groups/institutions.**
  - **Areas: Mathematics, philosophy, astronomy, mind, language, logic, etc.**

Meta-Themes

Meta-themes

1. **Application of traditional knowledge to solve contemporary problems**
2. **Look at own strength. (Academia looked too much towards the West. Needed change)**
3. **Academic research should connect with real life problems**
4. **Team Work without Ego**

1. Meta-theme: Application of Traditional Knowledge

- **Not just interested for the sake of history (though history is also important).**
- **Panini applied to modern Indian languages or teaching of Sanskrit (VC)**

\+ **National attention through KRIS-1985.**

2. Meta-theme: Look at Own Strength

- **Found that Westward-looking attitude influenced nature of research in academia**
  - **Take the best from everywhere, but do not be swept away.**
- **Found people working on fashionable problems. Our own problems - not addressed by academia.**
- **National attention: 1990 DST group project on theoretical traditional sciences**

3. Meta-theme: Connect with Real Life

- **Academia in India has a larger role than in the West, where industry is there to connect with real life problems.**
- **Real life always a hard test. Forces one to think differently, at times, and come up with creative solutions.**

4. Meta-theme: Team Work without Ego

- **Team work**
- **Working for common good**

Summary

- **Human aspects: People involved**
- **External conditions: Prevailing conditions, Events**
- **Research ideas/themes**
  1. **Panini applied to modern Indian languages (ILs)**
  2. **Parsing of sentences**
  3. **Theoretical comparisions with contemporary computational frameworks**
  4. **Machine Translation**
  * **Relation of Sanskrit with other Theoretical Sciences besides CS**
- **Meta-themes**
  1. **Application of traditional knowledge**
  2. **Rooted in the self**
  3. **Connect with real life problems**
  4. **Team work without ego**