



LRIL

**Problems in Developing Lexical
Resources for Computing**

Rita Mathur

Word list

- Tatsama words
- Tadbhava words
- Borrowed words from English
- Borrowed words from Parso – Arabic



Lexicon may consist of

- Basic form
- Inflectional forms may be generated
- Morphological form
- Word categorization



Word categorization

```
graph TD; A[Word categorization] --> B[Grammatical category Agreement]; A --> C[Word endings a, i or u ending]; A --> D[Source language Sanskrit English Perso-Arabic];
```

Grammatical category
Agreement

Word endings
a, i or u ending

Source language
Sanskrit
English
Perso-Arabic

Problems of spelling standardization

- Pronunciation
- Traditional usage
- Printing and writing convenience
- Individual preference
- Consistency with pattern



Solution

- Selection of the words
- Purpose
- Level
- Users
- Generalized rules
- Tagging



Standardization

- *Consistency*
- *Orthographic*
- *Pronunciation*
 - *Regional variation*
 - *Individual variation*
- *Need for computing*



Thank You

