

# Lexicography for Endangered Languages

Mayank Jain  
JNU, New Delhi

# What is a dictionary?

- A dictionary is a list of words with their definitions and other details such as pronunciation, morphological information, etymological information, illustration, usage, example sentences etc.

# Fundamental Question for lexicographers

- Users – who will use the dictionary?
  - The community
  - Professionals
  - Learners
  - others
- 
- Answers to this question will decide design and publication decisions for dictionary maker i.e. which type of dictionary should be made, what information should be given in a particular entry etc.

# Types of Dictionaries

- There are various types of dictionaries as mentioned below:
  - monolingual** – consists of one language
  - bilingual** – Consist of two languages
  - encyclopedic** (detailed information about head word)
  - compact** (small dictionaries)
- Many dictionaries also provide pronunciation information (PRONUNCIATION DICTIONARY)
- Some dictionaries exclusively give historical origin of the words (ETYMOLOGICAL DICTIONARY).
- Dictionaries of specific domain e.g. (LEGAL, MEDICAL etc)

# Dictionaries interacts with linguistic disciplines

- When we look at a dictionary entry, we see that lexicography uses the knowledge from different branches of linguistics.
- For example, phonetics and phonology (pronunciation); semantics, lexical relations, historical linguistics, derivational morphology etc.

# Structure of dictionary

MACROSTRUCTURE

MICROSTRUCTURE

# Macrostructure

Overall structural organization of whole dictionary, typically:

- **Front Matter** : introduction, user guidelines
- **Body** – entries and definitions (other information in an entry), organisation of dictionary (alphabetically vs. thematically organised dictionaries)
- **End Matter** – appendices and additional information, eg. personal names, place names, maps etc.

# MICROSTRUCTURE

**Internal Structure** of dictionary entry blocks, typically:

- headword (lemma, form to be looked up)
- pronunciation
- Variant(s)
- Morphological information
- Part of speech category or word class
- Meaning(s)
- Examples
- Etymological or historical notes
- Cross-references to related items
- Still pictures and audio files for entries
- Semantic domain
- Detailed cultural and other information about some entries
- Source
- Others according to the requirement.



# Headword

- Headword is typically **citation** form of a lexeme.
  - Citation forms will depend on the users of the dictionary. We should choose such citation forms which would be understood by users.

# PRONUNCIATION

Separate pronunciation of the head word is provided in IPA (If the headword is not in IPA)

# VARIATION

- Variations of the headword needed to be recorded in the dictionary. This variation may be because of different factors e.g. geography, age etc.
- Each variation will be entered in a separately.

# Morphological Information

- Morphological information about the headword may be provided, i.e. constituent morphemes of the word.

# Part of Speech (Word Classes)

- Apart from basic POS we may need to modify POS categories on language particular basis.
  - For example, In Assamese, a lexical category of classifier is a required.
- **noun, verb, adjective, adverb, pronoun, determiner, adposition, conjunction, classifiers, particle, clitic .....**

# Word Meaning and Definition

- Central task in a dictionary making is arriving at the meaning(s) of words.
- Glosses should capture the basic meanings (as far as possible) of the words.
- The explanation of the gloss (single word usually) is necessary as it is usually not able to capture the meaning of the word in the gloss language. For example, Khasi has a number of words for various types of movement which can not have equivalent in a single English or Hindi Word.

# Homonymy

- If two words have the same form, but their meanings are unrelated, we call them homonyms
- One should be careful of homonyms.
- 'Bear' (the animal) and 'bear' (to carry) are homonyms.

# What is a homophone?

- A type of homonym.
- If two words are pronounced the same, we call them homophones.
- 'Two' (2) and 'Too' (also) are homophones.



# What is a homograph?

- A type of homonym.
- If two words are spelled the same, we call them homographs.
- ‘Sow’ (to plant) and ‘sow’ (female pig) are homographs, but not homophones.
- ‘Sow’ (to plant), ‘so’, and ‘sew’ are homophones, but not homographs.

# Example Sentences

- A good example may explain better than a paragraph.
- In the best practice, Every sense of the entry should have a example sentence.
- Examples should be related to real life.
- Try to get the example sentence from the native speakers. Avoid creating example by yourself.

# Etymological or historical notes

- Documenting a word's history.
- If a word is historically inherited from a language e.g. in New Indo-Aryan languages, there are many words from which are historically related to Sanskrit.

# Lexical Relations

- Synonyms or antonyms
- Part-whole relationship, for example A Room has walls, A Tree has branches
- Generic to specific, e.g. Flower is a generic, rose is a specific member of 'flower'

# Cross-references to related items

- Words related to each other which are not captured by the lexical relations.
- For example, an entry of a 'tree' can be cross referred to the 'medicine' which can be made from that tree.

# Pictures and sound files for entries

- Pictures from the field can be entered in the dictionary. Also, illustrated pictures can be put in the dictionary. The pictures should be in the natural surrounding as far as possible.
- Sound files: Sound files of headwords and sentences can be incorporated in the dictionary.

# Detailed cultural and other information about some entries

- Some entries are culturally very significant have some culture specific details associated with them.
- Such details needs to be put in the dictionary.

# Semantic domain/ Theme

- The semantic domain(s) to which a word belong to have to be entered in the dictionary.
- For example words belonging to kinship terminology, Flora and fauna etc.



# Source

Source of the data should be entered in this field in a uniform format e.g. A possible format could be as given below:

Name of the field linguist/Name of language helper/Date/Place/Audio and video Recording information/Hard copy information

# Usage

- How the word is used in the society.

For example

- Is it a taboo word.
- Restricted to religious domain.

# Semantic domain/ Theme based dictionaries

- It is advisable to follow the theme based approach to create dictionaries of related to one theme.

For example,

- Kinship Terms
- Flora and Fauna
- Body part terms
- Ornaments and costumes

# Semantic domain/ Theme based dictionaries

- If you are working on a language for a **short time**, it is always preferable to create dictionaries of one subject area.
- **Choosing** a subject area should be done considering the importance of a field with respect to a particular field

For example, for a community living in a forest, the dictionary related to forest, trees, animals are suitable.

# Other fields

- According to the needs we may add some other fields in the dictionary for example literal meaning, scientific names etc.

Thank you