

## Linguistic Data 2

7.1.20

- Language intro
- Linguistic Data: — looking at data, analytically looking at how these work.
  - how information is encoded in different languages.
  - even a single occurrence of a token should be dealt with.

attested vs Possible

- Data in 2 forms:
  - spoken
  - written

We will deal with a lot of text (and annotation)

— Active vs Passive

: more in English than Indian languages.

makes Indian languages more "free word order"

→ markers are used in ILs, so fewer passive.

eg: John saw Mary. | John Mary ne dekhlo.  
Mary saw John. | Mary John ne dekhlo.

Demonstration of  
"structures in  
languages may not  
exist across"

— segue into different ways of studying linguistics.

: talks about Chomsky vs Saussure.

- language learning
- I-language (Competence)
- E-language (Performance)
- language generation

- social perspective
- focus on performance

- for LDs need to understand which theory/style to follow.

(Chomsky)

Class will focus on syntax + semantics.

ANCORRA for annotations.

UDs.

- antecedent
- context

Free flow of info

- pro drop languages
- ellipsis
- ambiguity in general.
- extra linguistic knowledge
  - disambiguating with known context.
  - world knowledge
  - cultural knowledge
  - pragmatic knowledge.

# Linguistic Data and Theories.

10.1.20.

- what do we mean by competence.
- " " " " " acquisition
- " " " " " performance. } in class, most interested in this.

Studying for Ling. but CL esp.

characteristics  
universals  
typology.

Applications

- MT
- QA
- Summarization
- ...

Complementizer.

Grammaticalization.

MENTAL GRAMMAR.

nativisation of new words.

This course is the same as ITL1, ITL2, CL1, CL2,

Ling1 so far.

inflection of words // categories of words that inflect.

Nouns of Space and Time (NST)

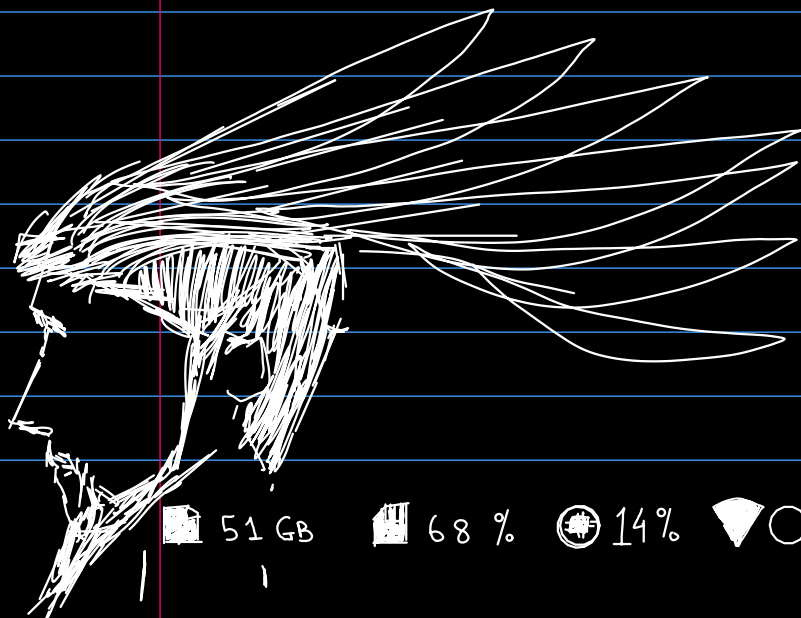
↳ in IL, inflect sometimes

} functionally adverbs  
} morphologically NST.

Why do we need to understand Grammar?

- Well formedness, ill-formedness.
- Acceptable vs unacceptable.

Challenges: unacceptable regular usage.



Theory Building:

Inductive and Deductive

/  
Observation  
+ analysis

Corpus based vs Corpus driven  
deductive inductive

