- Language intro

- Linguistic Data: — looking at data, analytically looking at how these work.

  — how information is encoded in different languages.

  | attested vs Possible |

  — even a single occurence of a token should be dealt with.

- 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.

  → markers are used in ILs, so fewer passive.

  makes Indian languages more "free word order"

  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 LD, need to understand which theory/style to follow.

                                              (Chomsky)

Class will focus on syntax + semantics.

ANCORRA for annotations

UDs.

                                    Free flow of info
                                    • pro drop languages
                                    • ellipsis
                                    • ambiguity in general.
                                    • extra linguistic knowledge
                                      — disambiguating with known context.
                                      — world knowledge
                                      — cultural knowledge
                                      — pragmatic knowledge.

  • antecedent
  • context

– 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
        Q,A
        Summarization
        :
        :

Complementizer.

Grammaticalization.

MENTAL GRAMMAR.   This course is the same as ITL1, ITL2, CL1,CL2,

    nativisation of new words.   Ling1 so far.

inflection of words // categories of words that inflect.

Nouns of Space and Time (NST)   { functionally adverbs
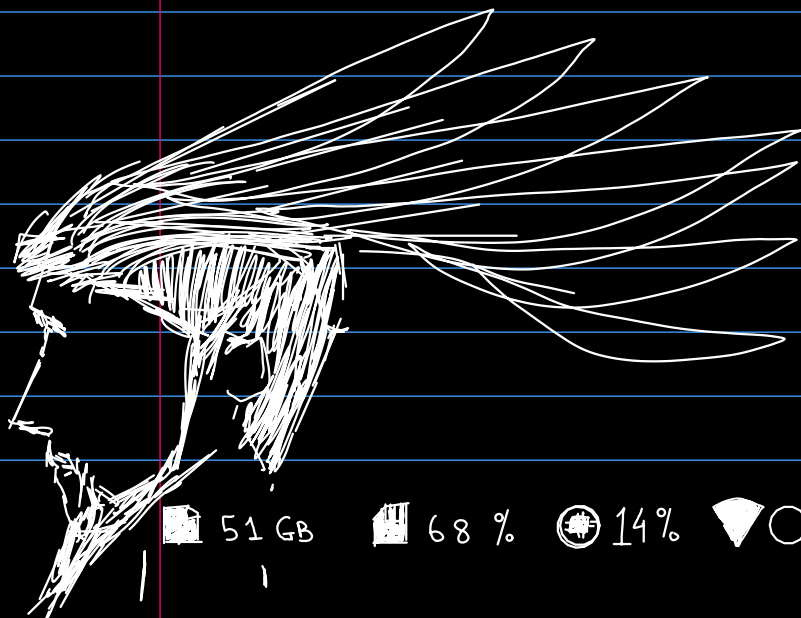   ↳ in IL, inflect sometimes   { morphologically NST.

Why do we need to
understand Grammar?

• Well formedness, ill-formedness,

• Acceptable vs unacceptable.
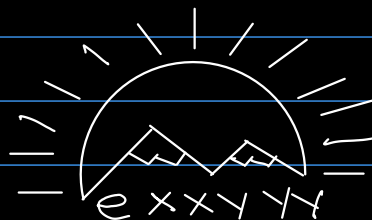
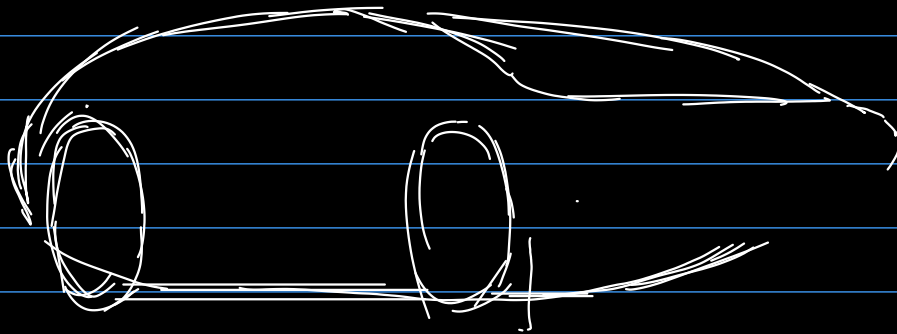
Challenges: unacceptable regular usage.

51 GB   68 %   14%   OnePlus5z   73%

Theory Building :

Inductive and Deductive.

Observation
+ analysis

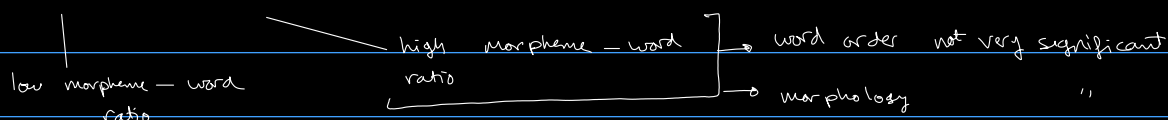Corpus based vs Corpus driven

deductive inductive.

e x x y y y

Study on Syntax

- find rule-governing system of sentence formation
  - understanding constituents of sentences.
- relationship between finite and infinite.
  - Not many types of phrases: NP, PP, AdjP, etc. } both finite ↓ infinite sentences.
  - " " categories

Analytic vs Synthetic Languages

low morpheme - word ratio

high morpheme - word ratio } → word order not very significant
→ morphology "

Analytic ←——————————————————————→ Synthetic

Analytic   Fusional   Agglutinative   Synthetic.

Polysynthetic

Hard to get even the word boundaries

- Inuktitut
- Yup'ik
- Eskimo

tons of
- infixation
- circumfixation } very complex.

Syntactic Categories

▷ **Lexical Categories** — imp for getting semantics

▷ **Functional** " — cannot add inflections.

Looking at each category ———— all categories (for each language) have rules

○ **Nouns** — tons of stuff.

  — proper nouns vs common nouns.
  : not pluralised
  : preceded by articles

  Count   Mass
  /   \   /   \
Concrete Abstract Concrete Abstract

— don't pluralise.
— do not take indefinite det.
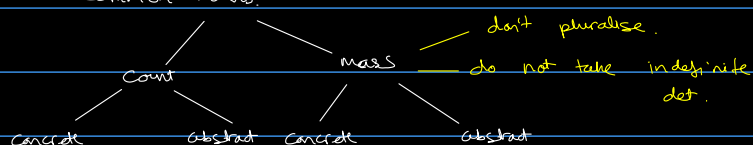
Grammatical Features

number
gender
case

O **Verbs**   major category.   // carry tense | aspect | modality | agreement

        English is verb - median
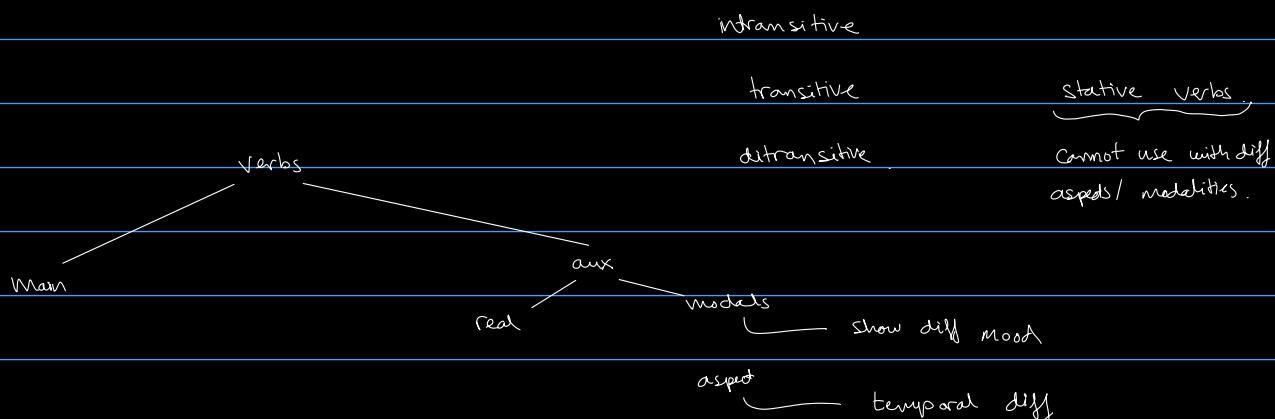
                                              intransitive

                                          transitive   stative verbs
                                                                        ‗‗‗‗‗‗‗‗
        Verbs   ditransitive   cannot use with diff
          ⟋ ⟍   aspects / modalities.
        ⟋   ⟍
    Main   aux
                  ⟋ ⟍
              real   modals
                            ⟋ show diff mood

                    aspect
                        ⟋ temporal diff

Grammatical   Features

    tense   portmanteau   morpheme

    aspect

    mood

    agreement

                    dubious   status in Indian languages
O **Adjectives** —

    — attributive
    — predicative

O **Adverbs** — can be (almost) anywhere in
                      the sentence.

        Modify — clause | Fortunately |, we didn't miss the train
                — verb   Peter walked | slowly | back to the car
                — adjective   That was | extremely | useful.

    O **Pronouns**

            — like Nouns | Morphologically.
            — unlike ", different plural formations. |

    Syntactically | they replace
                   | ≻ Personal   Change shape in subject/object
    also Semantically | NPs
                                        ≻ Possessive

                                        ≻ Reflexive

                                        ≻ Relative
            In most (ILs) reflexive pronouns have duplicative   structure.
                — part 1 inflected   ≻ Interrogative   not in morphologically rich
                — " 2 repetition   ≻ Demonstrative   languages

Due to complexities, cannot map languages easily.

O **Prepositions**

Space

Time

Cause

Instrument

Accompaniment

Concession

Exception

Addition

Case	Subject	instrument
	object	associative
	d-obj.	ablative
	i-obj	locative
		genitive
	infinitive marker	

When case marker has more than one function

we call it case syncretism

Functional categories are ambiguous due to their construction

Lexical categories are ambiguous because of their semantics ← needs a bigger context for disamb

O **Conjunctions**

time

O **Determiners**

look at

| Definite | Demonstrative | WH | universal dependency can |
| Indefinite | Possessive | Negative | |

Phrasal Categories

Due to complexities, cannot map languages easily.