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.

→ 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.

Chomsky:
- Language learning
- I-language (Competence)
  E-language (Performance)
- Language generation

Saussure:
- social perspective
- focus on performance.

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

(Chomsky)

Class with 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 contexts.
    - 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.

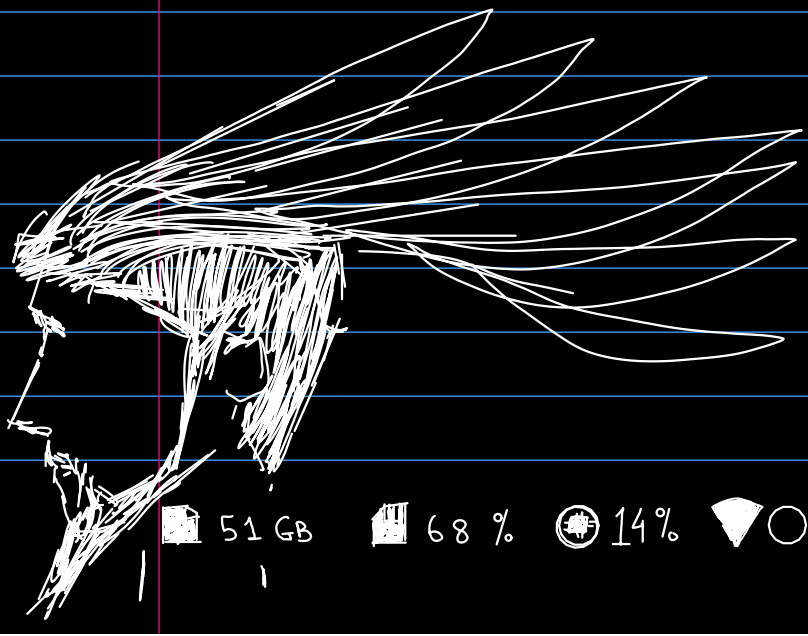
nativisation of new words.

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 :

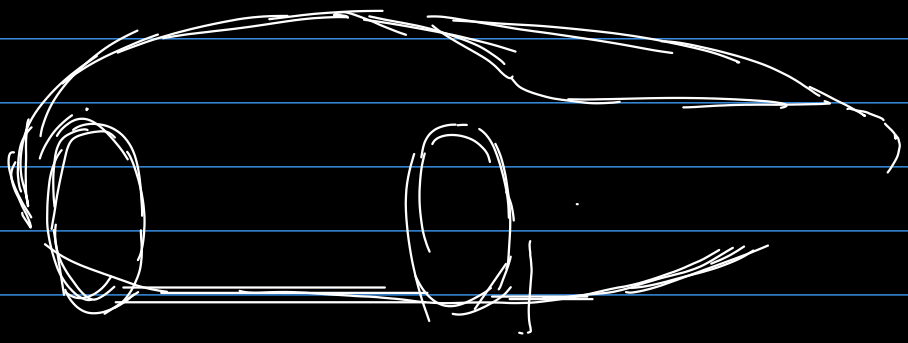
Inductive and Deductive.

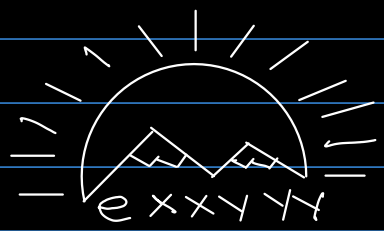
/

Observation + analysis

Corpus based vs Corpus driven.

deductive inductive.





17.01.2020

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
  - " " categories } ↓ infinite sentences.

Analytic vs Synthetic Languages

low morpheme – word ratio

high morpheme – word ratio → word order not very significant
→ morphology "

Analytic — Fusional — Agglutinative — Synthetic

Polysynthetic

Hard to get even the word boundaries

- Inuktitut
- Yupik
- Eskimo

/ forms of
- infixation
- circumfixation

} very complex.

Syntactic Categories

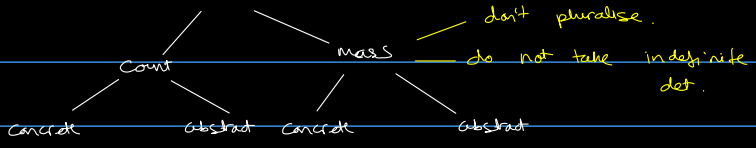
▷ Lexical Categories – imp for getting semantics

▷ 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



Grammatical Features

number

gender

case

○ Verbs

major category. // carry tense / aspect / modality / agreement

English is Verb – median

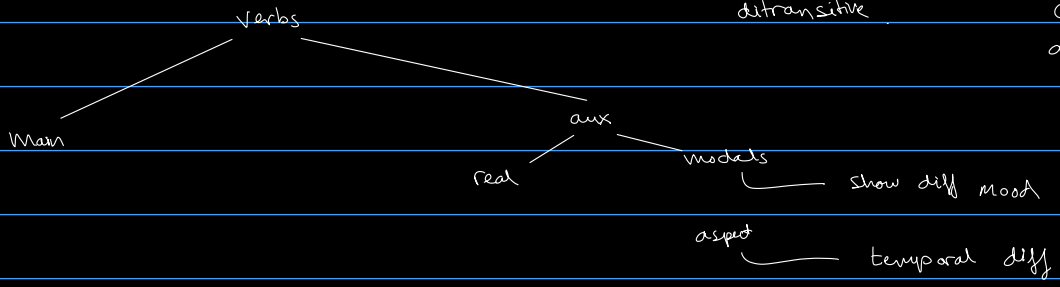
intransitive

transitive

ditransitive.

Stative verbs.

Cannot use with diff aspects / modalities.



Grammatical Features

tense

aspect

mood

agreement

portmanteau morpheme

○ Adjectives — dubious status in Indian languages

- attributive
- predicative

○ 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.

○ Pronouns

— like Nouns

— unlike " , different plural formations. | Morphologically.

Syntactically — they replace NPs

also Semantically

> Personal Change shape in subject/object

> Possessive

> Reflexive

> Relative

> Interrogative

>Demonstrative

In most (IE), reflexive pronouns have duplicative structure.

— part / inflected

— " 2 repetition

not in morphologically rich languages.

>| Indefinite

Copula

Due to complexities, cannot map languages easily.

○ Prepositions

Space
Time
Cause
Instrument
Accompaniment
Concession
Exception
Addition

Case

Subject
object
d-obj
i-obj
infinitive marker

instrument
associative
ablative
locative
genitive

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

○ Conjunctions

time

○ Determiners

look at universal dependency cats.

Definite | Demonstrative | WH
Indefinite | Possessive | Negative

Phrasal Categories

1. Noun Phrase

Thematic Roles — Phrase Structure Trees

PP attachment always shows ambiguity.

2. Verb Phrase.

Any Sentence must have 1 NP 1 PP

Structure of Clause

Relation between phrases.

independent clause / matrix clause.

dependent clause / ??

Complementizers.

All relative clauses are not complementizers.

Adverbial Classes

Sentence Composition Types

Complexity in Syntax

1. Ambiguity
   - attachment ambiguity.
   - correlation ambiguity.
2. Garden path sentences
3. Recursiveness

24.1.20.

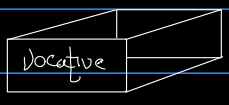
Assignment 1:

- take 100 contiguous sentences, (min length: 5 words)
- annotate with UD/AnCorra.

explicative subject. universaldependencies.org/u/dep/obj.html

Differential object marking (Aissen)

Related to hierarchies.



Agent

all cases linked to the VP.

dislocated