CS 671 NLP
PARTS-OF-SPEECH TAGGING AND SYNTAX

amitabha mukerjee
iit kanpur

Structure in language

पांच फिरंगी अफसरों __ फांसी
पर ____ दिया

what can go in the blanks?

what can NOT go there?

Syntax

Sentences are built from “words”.

boys like girls

germans drink beer

sentence = noun verb noun

Syntactic Composition

- Constituency : *like girls* = verb phrase VP
  head : *like* V
  constituent: *girls* N-plural
- Grammatical Function (maps to semantics?):
  subject: boys
  predicate: like
  arguments: boys, girls
- Hierarchy and Control

One Version of Constituent Structure

- Lexicon: *the a small nice big very boy girl sees likes*
- Grammatical sentences:
  - (the) boy (likes a girl)
  - (the small) girl (likes the big girl)
  - (a very small nice) boy (sees a very nice boy)
- Ungrammatical sentences:
  - *(the) boy (the girl)
  - *(small) boy (likes the nice girl)

Regularities : Wh-movement

- *I saw Ram*
  Who did you see?
- *Maine rAm ko dekhA*
  Tumne kisko dekhA?

29% of V-final languages have wh-movement
58% of V-medial languages have it

COMPOSITION / SYNTAX

What is Syntax?

- *Compositionality* Assumption: Larger phrases built up from smaller ones
- Construct rules for how words compose into phrases and sentences = **Grammar**
  - may also apply to morphemes

Why is Syntax Important?

- Grammar checkers
- Question answering
- Word sense Disambiguation
- Information retrieval (?)
- Machine translation
- Most NLP tasks

Theories of Syntax?

- Unfortunately, no consensus on a theory of grammar - aggressive debates :
  - Chomskyan – formalist, autonomous from semantics, we are born with syntax
  - Cognitive linguistics – semantics has a role, language is learned by discovering patterns in usage

Syntax : Composability

- Are sentences constructed by combining words? [decomposability]

- Or are words obtained by breaking up sentences? [holism]

- Possibly, in learning a language, babies understand the sentence *before* the words

Chomskyan (Generative) view

- Syntax is independent of meaning. Perception, action, etc. are not relevant to grammar
- Of course, language is compositional
- Lexicon = list of words → *arbitrary*
- Syntax: Words are composed via deterministic, formal rules → *systematic*

Chomskyan Language Acquisition

- Babies acquire language with very little guidance. (Poverty of Stimulus)
- Possible only if we have an innate *Language Faculty* with a built-in *Universal Grammar* (Nativism)
- Language learning = filling language-specific parameters in the UG

Autonomous Syntax

- Are grammaticality judgments based on form alone?

colourless green ideas sleep furiously

vs

furiously sleep ideas green colorless

→ ***autonomy of syntax*** argument

[chomsky 57]: syntactic structures

Autonomous Syntax : Assumptions

- Rules determining the syntax (form) of language are formulated without reference to meaning, or language use.

- Related : **Grammar is not statistical**

  "There appears to be no particular relation between statistical relations and grammaticalness" p.17 [chomsky 57]: syntactic structures

  see P. Norvig: On Chomsky and the Two Cultures of Statistical Learning [http://norvig.com/chomsky.html]

Cognitive Linguistic view(s)

- Syntax is dependent on, and guided by the intended meaning.
  - Grammatical structures also have meaning

- Meaning $\neq$ reference
  - “The eminent linguist”
  - “The blonde bombshell”

  May both refer to same person, but have very different connotations.

Cognitive Linguistic view(s)

- Syntax is not Formal, nor deterministic.
  Many phenomena are not sharply Yes-No:
  - Arbitrariness in the lexicon
  - Grammar – Lexicon continuum
  - Compositionality is partial

- Babies acquire language by relating phrases with their usage (meanings).

Language and Meaning

- 1955: J.L. Austin of Oxford – Lectures on Speech Acts → *How to do things with Words*
- 1957: Chomsky's *Syntactic Structures* :
  autonomy of syntax
- 1960: William Stokoe, Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf
- 1965: Rudolf Carnap, Meaning and Necessity
- 1987: Langacker: Cognitive Linguistics

Language = Speech Act

I pronounce you man and wife.

Translation

"Can't you see?"

language universal?

Redundant negation as agitation

Semantics – Syntax – Pragmatics divide

- CARNAPIAN division of the theory of language:
  - SYNTAX - relations between expressions
  - SEMANTICS - relations between expressions and what they stand for
  - PRAGMATICS - relations between expressions and those who use them
- [Peregrin 1998, The pragmatization of semantics] :
  - Internal Challenge: context – Deictic (pronouns, demonstratives); indef article “a” = introduces new element ; “the” = old item
  - External Challenge: **language is not a set of labels stuck on things**; not "what does a word mean?" but "how is it used?“ [Wittgenstein PI 53]
- Langacker : Composition based on Syntax + Semantics + Pragmatics

“Grammar” : many meanings

Narrow (traditional) sense :

- grammar = syntax + morphology
  (morphosyntax)

Broad (generative / cognitive) sense

- grammar = theory of language

Grammar as lexicon + syntax

Autonomous syntax: constructs based on :

arbitrary forms [lexicon] +

productive rules [syntax]

Cognitive grammar :

- lexicon-syntax division is not sharp, but graded.
- "generative rules" may not exist.
- grammar = continuum of constructions from:
  - very specific ("cat", "kick the bucket")
  - patterns (noun, transitive construction)
  - more general patterns (schemas)

[broccias 06] cognitive approaches to grammar

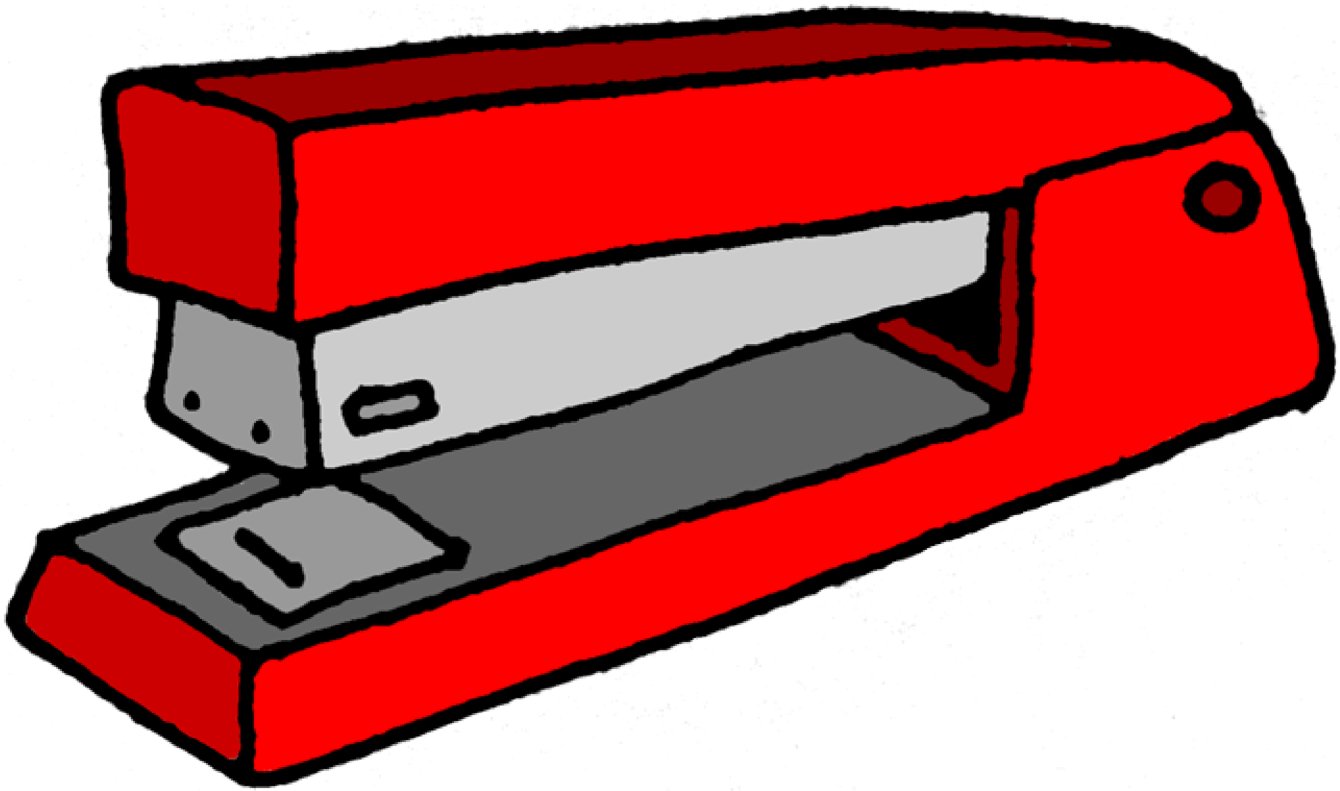
Non-arbitrary lexicons

"elephant"

jyoti vadhir vidyalay, bithur

Non-arbitrary lexicons

“stapler”



Given “staple”, “stapler” is not arbitrary!

Grammar as lexicon + syntax

Autonomous syntax: constructs based on :

arbitrary forms [lexicon] +

productive rules [syntax]

Cognitive grammar :

- lexicon-syntax division is not sharp, but graded.
- "generative rules" may not exist.
- grammar = continuum of constructions from:
  - very specific (*cat*, *kick the bucket*)
  - patterns (noun, transitive construction)
  - more general patterns (schemas)

[broccias 06] cognitive approaches to grammar

Traditional NLP models of syntax

पांच फिरंगी अफसरों को फांसी पर लटका दिया

- Language is compositional
- It’s not clear exactly the form of these rules, however, people can generally recognize them
- Rules of syntax may be probabilistic

Grounded Language

- grounded **lexicon**:
  relation between sounds and sensorimotor patterns

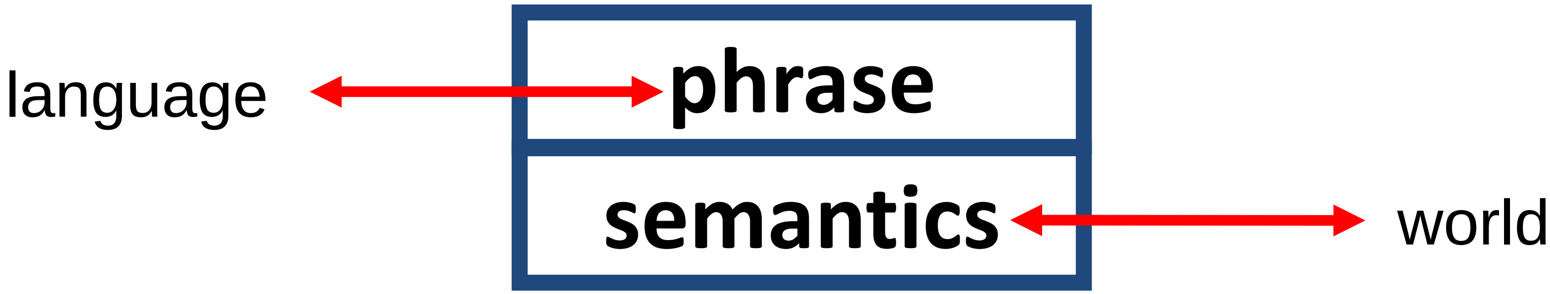
- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Symbol = Form-Meaning pair

- Symbols = (form) label + **meanings**.



symbol = label + semantics

[langacker 87]

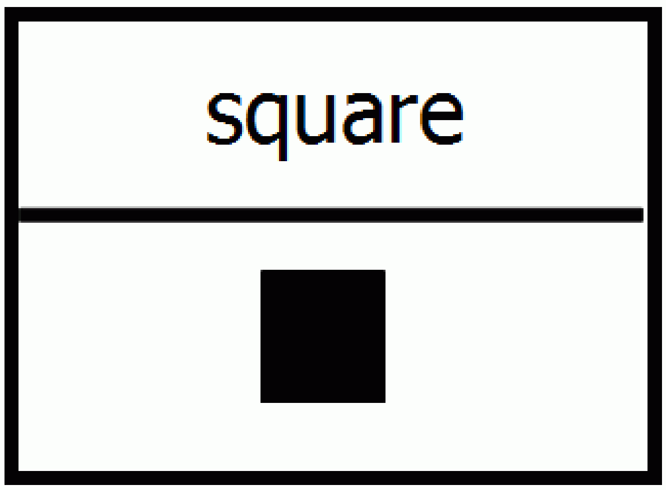
- Semantics : not static: evolves with language use
- *image schema* : map in perceptual space
- Linguistic label acts as index to concept
- Earliest image schemas = pattern on sensory data (chunk)

Difficulty

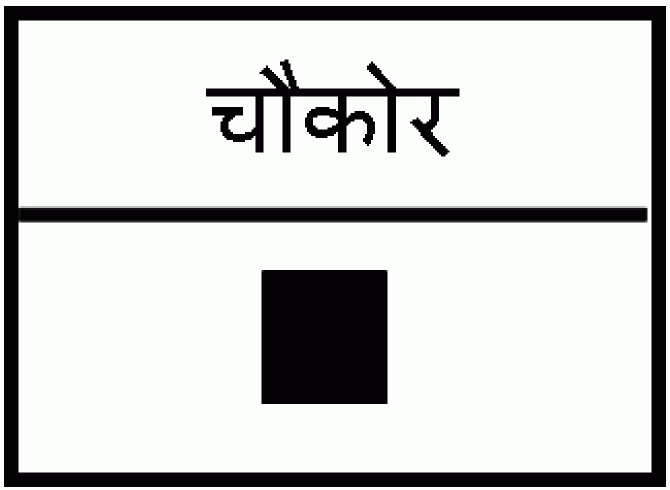
- What is meaning?
  Potentially unbounded set of relations arising in different usage situations

Lexicon

- grounded **lexicon**:



english lexicon

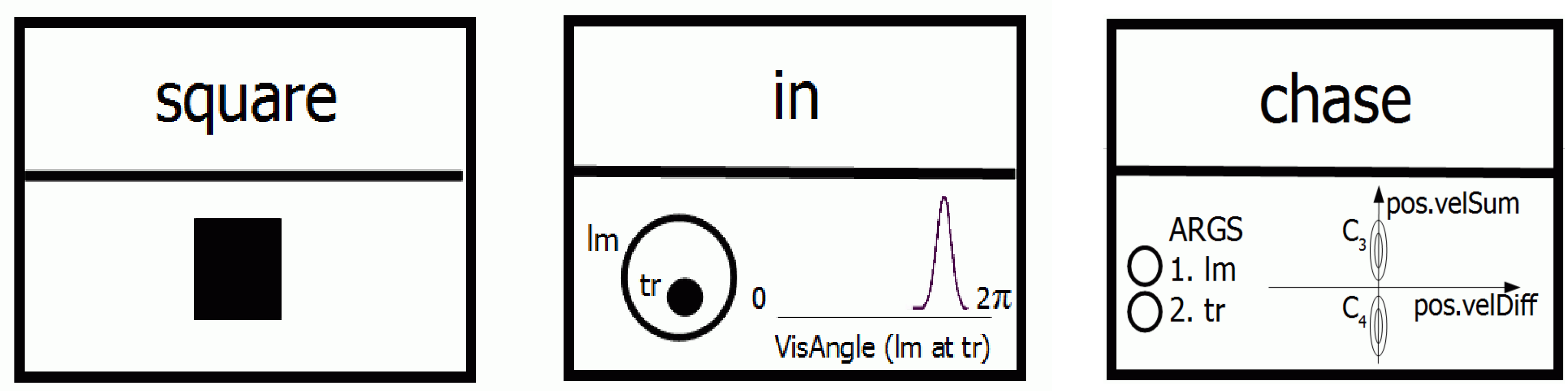


hindi lexicon

[langacker 87]

Lexicon

- grounded **lexicon**:



- semantic pole : perceptual patterns (image schemas)
  → probabilistic predicate + arguments

CS 671 NLP
PARTS-OF-SPEECH TAGGING AND SYNTAX

Modes of Learning

Grammar for NLP : Summary

- Syntax = systematcity in composing words
- Two views : Chomskyan vs Cognitive

- NLP approach: machine learning / probabilistic
  - Supervised: Based on annotated corpus with intermediate tags :
    - parts of speech (brown), parse tree (treebank),
    - semantic maps (framenet)
  - Unsupervised : Attempt to learn syntax + semantics from grounded input (embedded in context)
    - Given an input, provide a response. (No need to analyze)

Context Free grammar

- Syntax = systematcity in composing words
- Grammar G = (V, Σ , R, S)
  - V = variables (non-terminals)
  - Σ = vocabulary (terminals)
  - R = finite relation from V to $(V \cup \Sigma)^*$

  S = start symbol

- Productions or rewrite rules :

  S → NP VP NP → Det N VP → V N

  NP → N VP → V

Context Free grammar

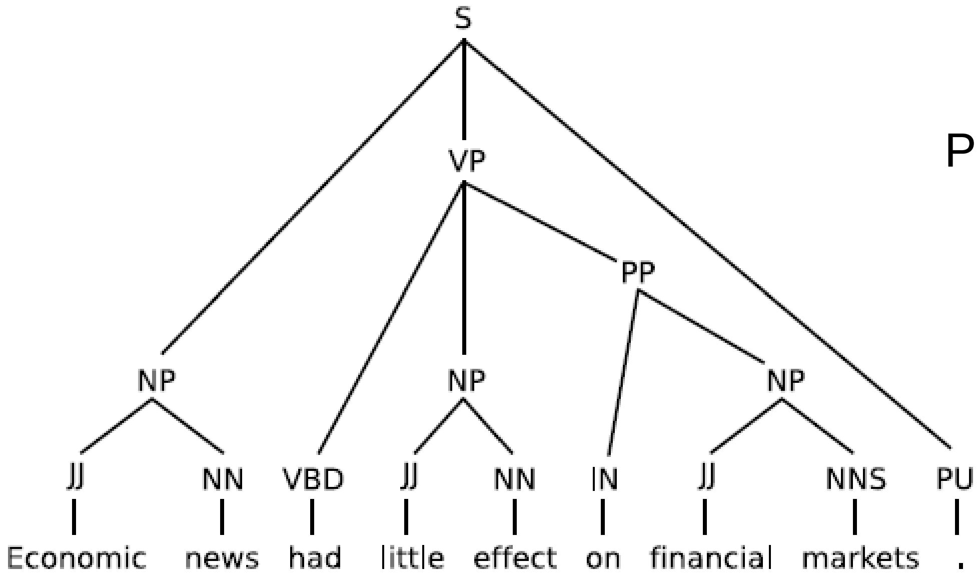
Can generate sentences:

boys like girls
germans drink beer

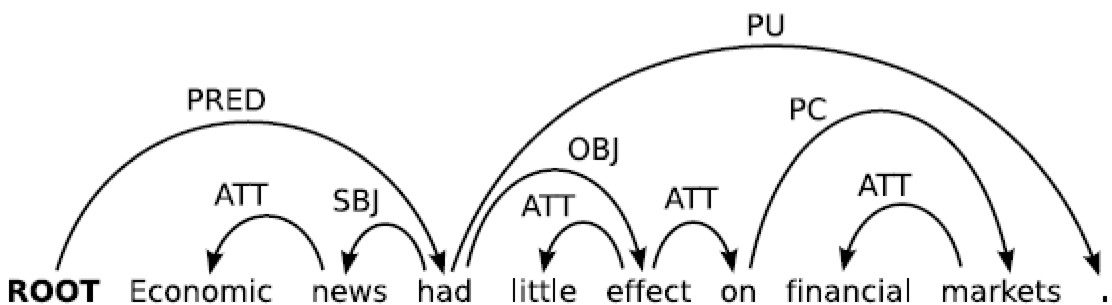
Sentence → NP VP

→ noun [verb noun]

Sample Parse



Parse tree



Dependency parse

kubler-mcdonald-nivre-2009_dependency-parsing

Tagged Corpus

आयकर\NC.0.sg.dir.0 आयुक्त\NC.0.sg.obl.gen (\PU अपील्स\NC.fem.sg.dir.0)\RDS के\PP.0.0.gen आदेशों\NC.mas.pl.obl.abl से\PP.0.0.abl पीडित\JJ.0.0.dir निर्धारिती\NC.0.sg.dir.0 ,\PU अपीलीय\JJ.0.0.dir न्यायाधिकरण\NC.mas.sg.obl.gen के\PP.0.0.gen समक्ष\NST.dir.0 अगली\NST.dir.0 अपील\NC.fem.sg.dir.0 कर\VAUX.0.0.0.0.0.0.nfn.0 सकता\VAUX.mas.sg.3.prs.pft.dcl.fin.n है\VAUX.0.sg.3.prs.pft.dcl.fin.n

Tagged Corpus

- Difficult to update for new usage structures

- Tags = Intermediate levels of analysis
  - Based on a theory
  - Does the theory have sufficient explanatory power?
  - Poor inter-annotator agreement

- Syntactic Analysis
  - Attempt to map to semantics based on syntax

VOL. LIX. No. 236.] [October, 1950

MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

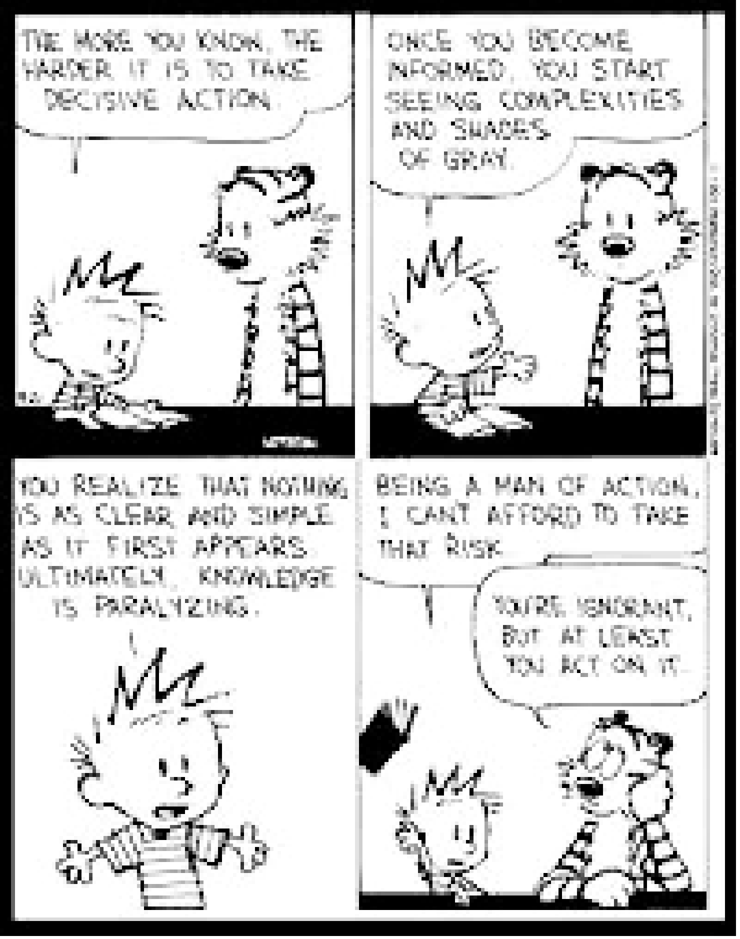
1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms

Instead of a programme to
simulate the adult mind,
why not rather try to produce one
which simulates the child's?

If this were then subjected to an
appropriate course of education one
would obtain the adult brain.

- Alan Turing, 1950



Machine Learning :

Unsupervised Discovery

vs

Knowledge-based Supervision

Shannon Entropy

- Predict the next word/letter, given *(n-1)* previous letters or words : Fn = entropy = $\text{SUM}_i (p_i \log p_i)$
- probabilities p_i (of n-grams) from corpus:
  - F_0 (only alphabet) = $\log_2 27$ = 4.76 bits per letter
  - F_1 (1-gram frequencies p_i) = 4.03 bits
  - F_2 (bigram frequencies) = 3.32 bits
  - F_3 (trigrams) = 3.1 bits
  - F_{word} = 2.62 bits
    (avg word entropy = 11.8 bits per 4.5 letter word)

Claude E. Shannon. "Prediction and Entropy of Printed English", 1951.

Shannon generation: English

- 1. Zero-order
  - XFOML RXKHR JFF JU J ZLPWCFWKCY JFFJEYVKCQSGXYD QI'AAMKBZAACIBZLHJQD
- 2. First-order (unigram frequencies as English)
  - OCR0 HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENH'ITPA OOBTTVA NAH BRL
- 3. Second-order (bigram).
  - ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE
- 4. Third-order (trigram)
  - IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

Shannon generation: English

- 5. Word models: First-Order
  - REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE
- 6. Word Model: Second-Order (bigram)
  - THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED T

Claude E. Shannon. *A Mathematical Theory of Communication*, 1948.

PARTS OF SPEECH

Parts of speech

- What are the English parts of speech?
  - 8 parts of speech?
    - Noun (person, place or thing)
    - Verb (actions and processes)
    - Adjective (modify nouns)
    - Adverb (modify verbs)
    - Preposition (on, in, by, to, with)
    - Determiners (a, an, the, what, which, that)
    - Conjunctions (and, but, or)
    - Particle (off, up)

English parts of speech

- Brown corpus: 87 POS tags
- Penn Treebank: ~45 POS tags
  - Derived from the Brown tagset
  - Most common in NLP
  - Many of the examples we'll show us this one

- British National Corpus (C5 tagset): 61 tags
- C6 tagset: 148
- C7 tagset: 146
- C8 tagset: 171

English POS Subcategories

- Adjective (modify nouns)
  - Basic (JJ): red, tall
  - Comparative (JJR): redder, taller
  - Superlative (JJS): reddest, tallest
- Adverb (modify verbs)
  - Basic (RB): quickly
  - Comparative (RBR): quicker
  - Superlative (RBS): quickest
- Preposition (IN): on, in, by, to, with
- Determiner:
  - Basic (DT) a, an, the
  - WH-determiner (WDT): which, that
- Coordinating Conjunction (CC): and, but, or,
- Particle (RP): off (took off), up (put up)

Hindi Parts of Speech - Base

- 1. Noun (N)
- 2. Pronoun (P)
- 3. Demonstrative (D)
- 4. Nominal Modifier (J)
- 5. Verb (V)
- 6. Adverb (A)
- 7. Postposition (PP)
- 8. Particle (C)
- 9. Numeral (NUM)
- 10. Reduplication (RDP)
- 11. Residual (RD)
- 12. Unknown (UNK)
- 13. Punctuation (PU)

Hindi Parts of Speech - Details

- Noun (N)
  - Common(NC) Gender, Number, Case, Distributive, Honorificity
  - Proper(NP) Gender, Number, Case, Honorificity
  - Verbal(NV) Case ex: जाने\NV के\PP लिए\PP
  - Spatio-temporal (NST) Case, Distributive, Emphatic, Dimension ex: आज, समक्ष

- Nominal Modifier (J)
  - Adjective (JJ) Gender, Number, Case, Distributive
  - Quantifier (JQ) Gender, Number, Case, Numeral, Distributive
  - Intensifier (JINT) Gender, Number, Case

POS Tagset: Hindi, Version 0.3, Oct 1, 2009 2

Hindi Parts of Speech - Details

Particle (C)

- Co-ordinating (CCD)
- Subordinating (CSB)
- Interjection (CIN)
- (Dis)Agreement (CAGR)
- Emphatic (CEMP)
- Topic (CTOP)
- Delimitive (CDLIM)
- Honorific (CHON)
- Dedative (CDED)
- Exclusive (CEXCL)
- Interrogative (CINT)
- Dubitative (CDUB)
- Similative (CSIM) Gender, Number
- Others (CX) Gender, Number, Case

POS Tagset: Hindi, Version 0.3, Oct 1, 2009 2

POS categories

"parts-of-speech" : not sharply defined
some may be more **prototypical**:

prototypical		non-prototypical
noun:	*cat, dog*	*equipment* (plural form?)
verb:	*go, tell*	*must* (**musted, *to must*)
adj:	*big, old,*	*asleep* (**an asleep dog*)

Syntax-Semantics Continuum

- What is a noun?
  - Parts of speech categories – are they purely syntactic?
- What about deictics : *you*, *the vase there*
- Some grammatical categories (e.g. plural-singular, mass-count, tense) – correlated with meaning?
- What is language about, if not about meaning

[pinker 94]: language instinct

Closed vs. Open Class

- ***Closed class*** categories are composed of a small, fixed set of grammatical function words for a given language.
  - Pronouns, Prepositions, Modals, Determiners, Particles, Conjunctions
- Open class categories have large number of words and new ones are easily invented.
  - Nouns (Googler, futon, iPad), Verbs (Google, futoning), Adjectives (geeky), Abverb (chompingly)

Part of speech tagging

- Annotate each word in a sentence with a part-of-speech marker
- Lowest level of syntactic analysis

John saw the saw and decided to take it to the table.

NNP VBD DT NN CC VBD TO VB PRP IN DT NN

Ambiguity in POS Tagging

I like candy.

VBP: (verb, non-3rd person, singular, present)

Time flies like an arrow.

IN: (preposition)

Syntactic (POS) and semantic role of “like”

Ambiguity in POS Tagging

I bought it at the shop around the corner.

IN: (preposition)

I never got around to getting the car.

RP: (particle… on, off)

The cost of a new Prius is around $25K.

RB:(adverb)

Role of “around” ?

Ambiguity in POS tagging

- Brown corpus analysis
  - Though only 11.5% of word types are ambiguous
  - 40% of tokens are ambiguous
  - Because most frequently used words are ambiguous
- Pick up the most common POS tag → Accuracy of 90%

Phrase structure

Syntax

Syntax: Study of how words may be assembled into sentences, or
how sentences may be broken down into smaller parts (hierarchy)

1. Break down sentence into relevant parts (**constituents**)

2. Assign **grammatical category** to constituents
   [e.g. “noun phrase”, “coordinator”]

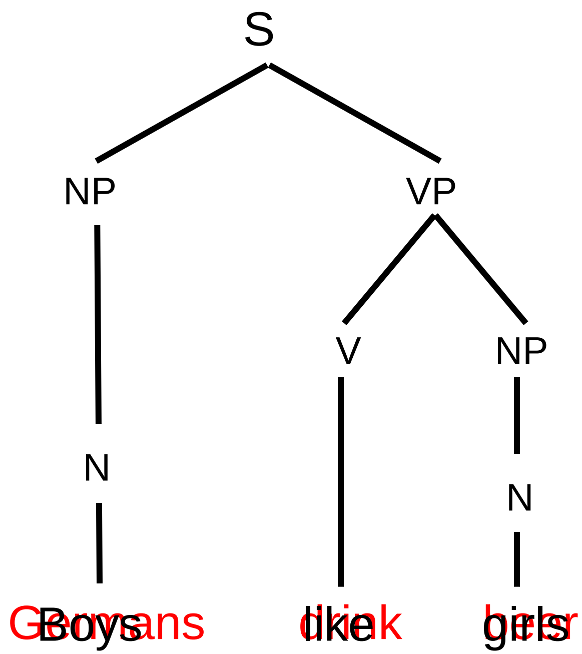
Syntactic Analysis

Sentence:	Germans drink beer	
Constituents:	[Germans]	[drink beer]
Category:	NP	VP

Verb phrase:	drink beer	
Constituents:	[drink]	[beer]
Category:	V	NP

Constituents may be from the lexicon (terminal) or may be phrases (non-terminal)

Syntactic Analysis



Phrase structure rules

S → NP VP
NP → N
VP → V NP
NP → det N

Lexicon

N → german[s], boy[s], girl[s], beer
V → like, drink

Hierarchy in Grammar

discourse	more than a single sentence
sentence	may be single **clause**, or **coordination** of multiple clauses
clause	**predicate** with **subject** [English: S P]
phrase	
word	lexical unit
morpheme	Smallest meaning-bearing unit

Clauses and Sentences

Single-clause
Sentence: Germans drink beer

Coordination
Sentence: The snake killed the rat
and swallowed it

Subordinate
Clause: No one doubts that the rat was killed

Hierarchy in Grammar

discourse	
sentence	$[_S$ Germans drink beer$]$
clause	S → NP VP
phrase	$[_S$ $[_{NP}$ Germans$]$ $[_{VP}$ drink beer$]$ $]$
	NP → N
word	$[_S$ $[_{NP}$ $[_N$ Germans$]]$ $[_{VP}$ $[_V$ drink $[_{NP}[_N$ beer$]]$ $]$ $]$
	NP → N VP → V NP
morpheme	$[_S$ $[_{NP}$ $[_N$ $[_{pl}$ German $[$-s$]]]]$ $[_{VP}$ $[_V$ $[_{pl}$ drink $[$-ø$]]]$ $[_{NP}[_N$ beer$]]$ $]$ $]$

Grammatical Function vs Grammatical Category

	Germans	like beer
function:	subject	predicate
category:	NP	VP

function: relation with other parts (subject of a clause)

category: grammatically similar expressions

Grammatical Function vs Grammatical Category

Germans is the subject of the clause

Germans like beer

Subject : w.r.t. a clause (not just subject)

Noun Phrase: is a category - may have different functions

Grammatical Function vs Grammatical Category

Same function, different categories:

[His guilt] was obvious. [NP]

[That he was guilty] was obvious.

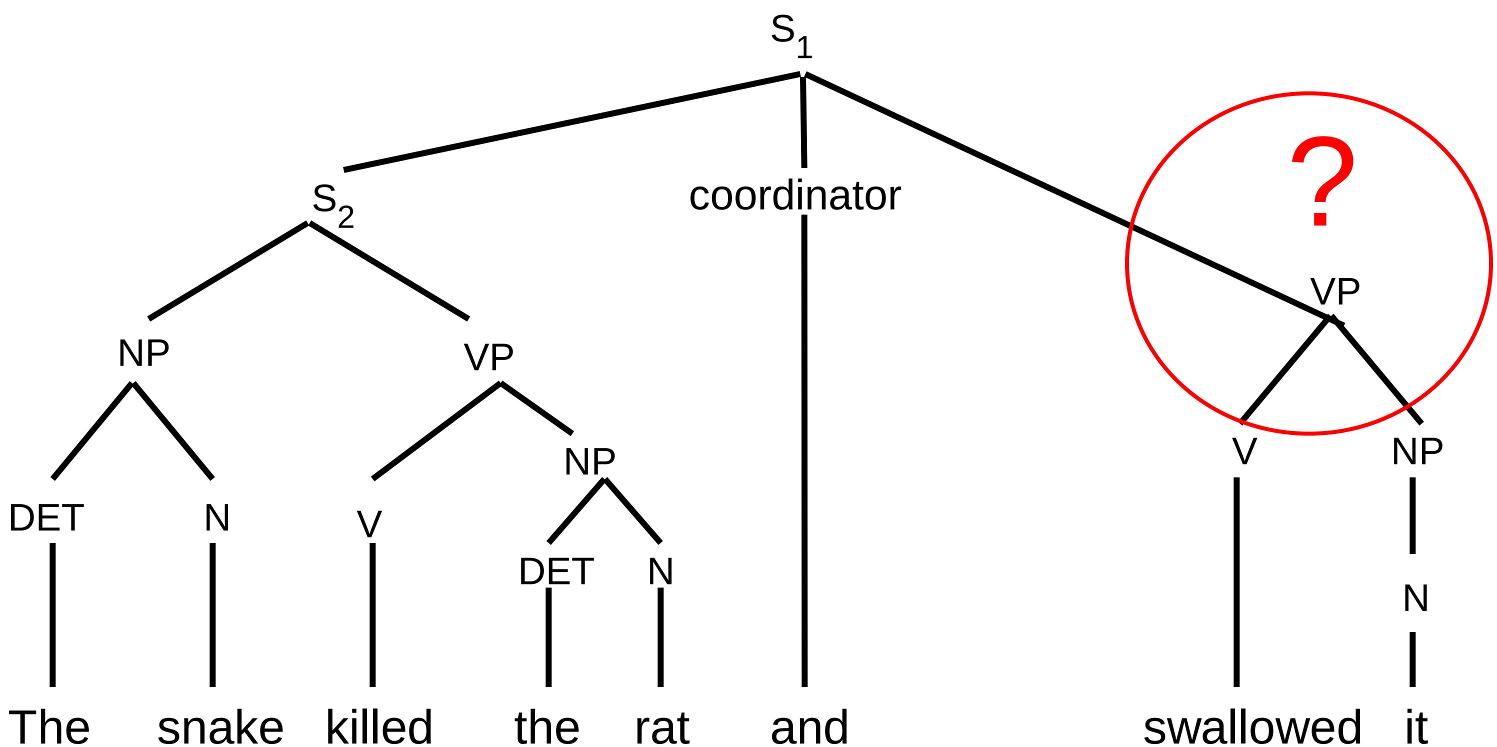
[Subordinate clause, with own subj/pred]

Same category, different functions:

[Some customers] complained. [subject]

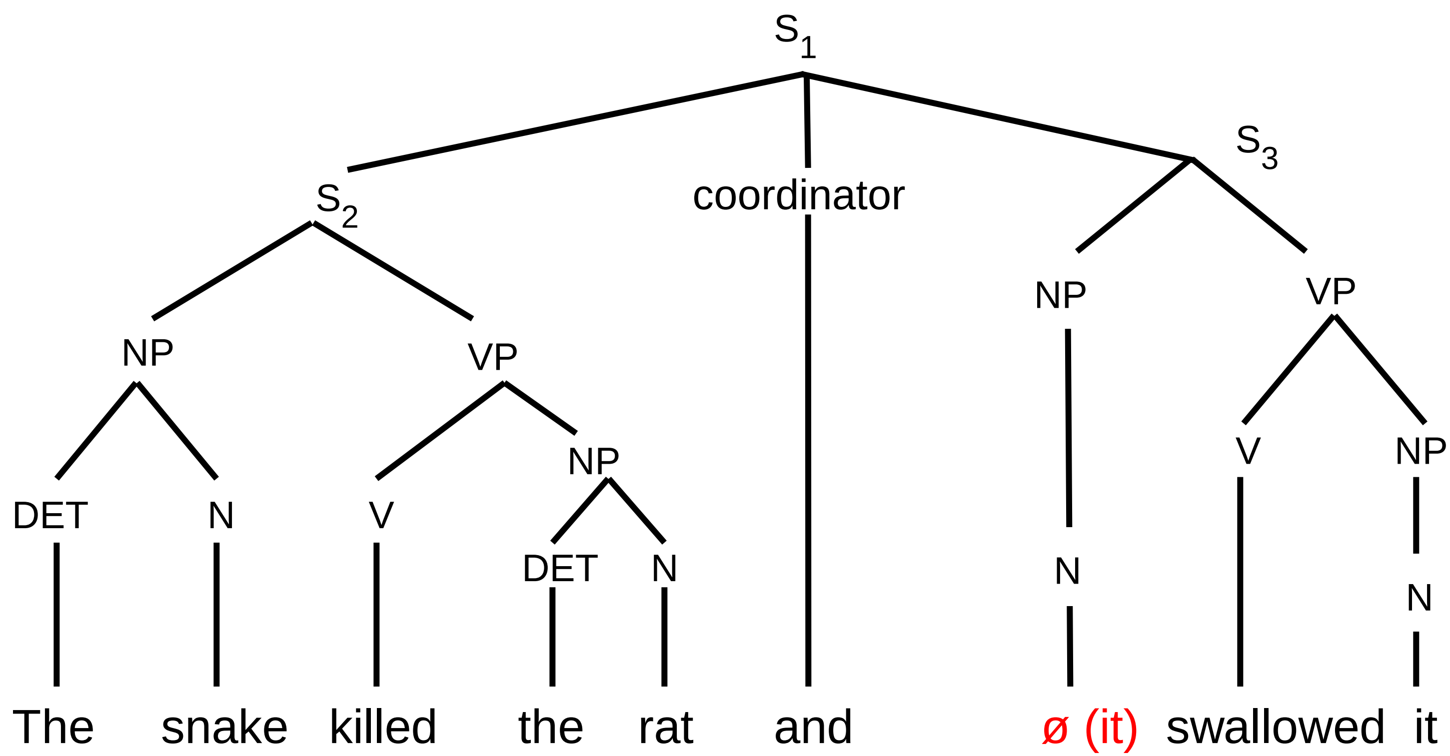
Kim insulted [some customers] [object]

Missing Elements?



[haegeman wekker 03] modern course in english syntax

Missing Elements : ?Ellipsis?



ellipsis

[haegeman wekker 03] modern course in english syntax

Bare argument ellipsis (BAE)

A: I hear Harriet's been drinking again.
B: *Yeah, scotch, probably*

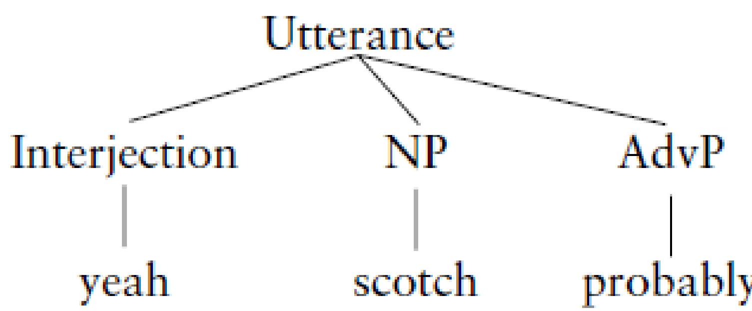
Generative Grammar analysis (ellipsis):

B: *Yeah, [Harriet has been drinking] scotch probably*

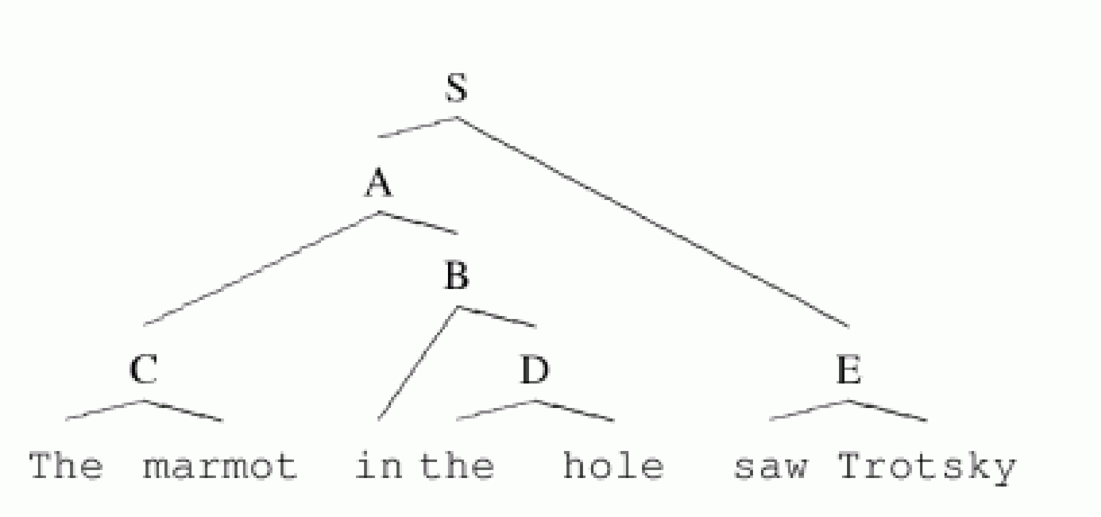
$[_{\text{ADVP}}$ *Yeah*] $[_{\text{NP}}$ e] $[_{\text{VP}}$ e *scotch*]] $[_{\text{ADVP}}$ *probably*]

Culicover / Jackendoff 02*:*

Accept fragment as is
use semantics / pragmatics
to judge grammaticality



Language and general cognition



Language as occlusion: Minsky, Society of Mind

Dependencies

Hierarchy in Grammar

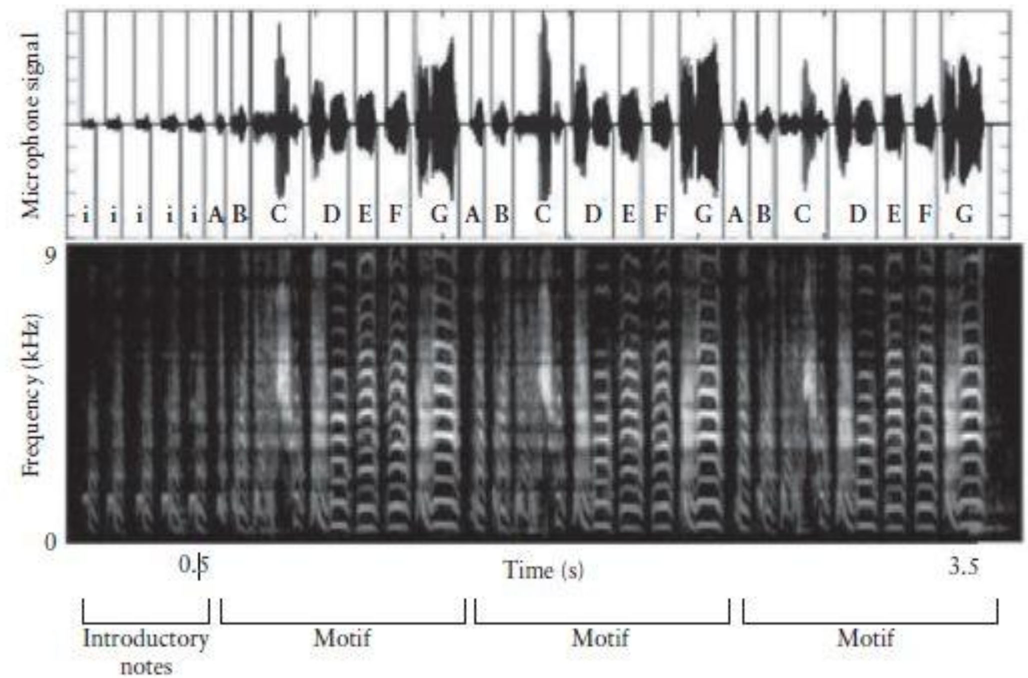
discourse	
sentence	[The snake killed the rat and swallowed it]
clause	[[The snake killed the rat] and [ø swallowed it]]
phrase	[[[The snake] [killed [the rat]] and [[ø] [swallowed [it]]
word	[[[[The] [snake]] [[killed] [[the] [rat]]]] [and] [[[ø][swallowed] [[it]]]]]
morpheme	

Lumpers vs **Splitters** in Syntax

- Some grammarians tend to **lump** different grammatical category into one super-category
- Others tend to **split** a category, making fine distinctions based on grammaticality data
- Also true for phrase structure rules
- But "there is no way to stop splitting"
  → Occam's Razor

Croft 04: Radical Construction Grammar

Zebra finch song

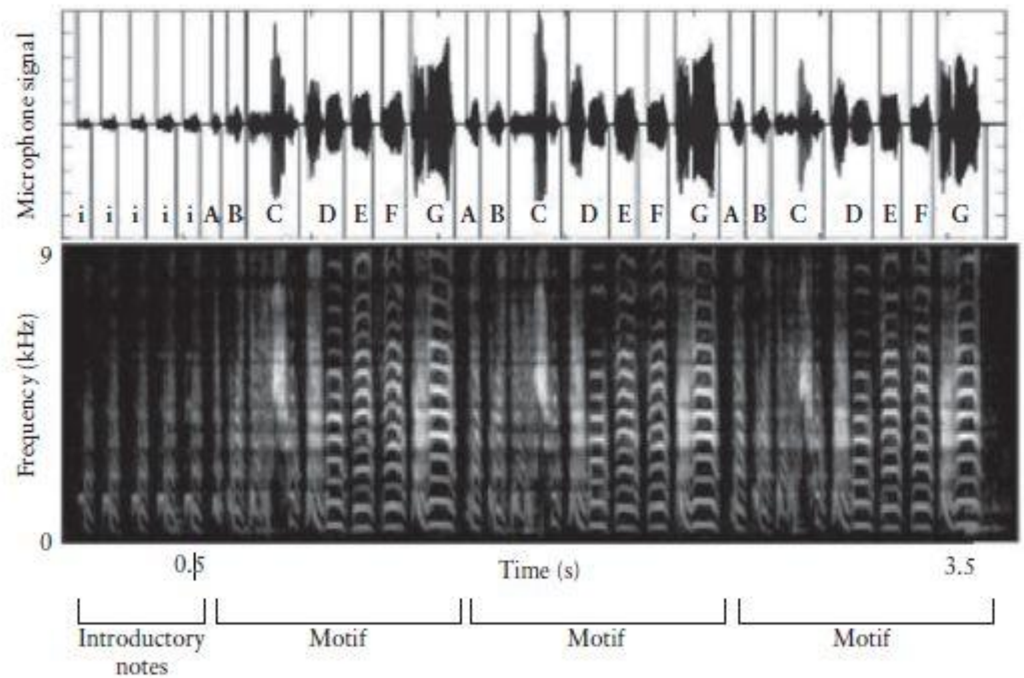


www.youtube.com : zebra finch song

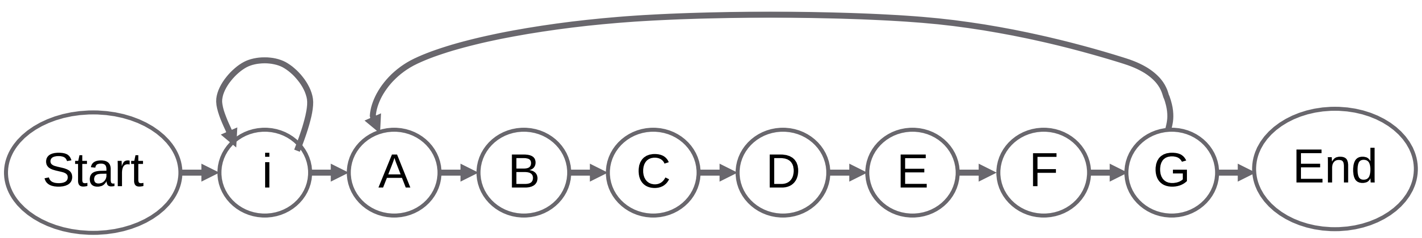
initial notes - "i" - repeated a few times

motif of syllables - ABCDEFG - repeated variable # of times.

Regular Grammar?



www.youtube.com : zebra finch song



[hurford 12] origins of grammar

STATISTICAL NATURAL LANGUAGE PARSING

POS-Tagging

POS Tagging Approaches

- **Rule-Based**: Human crafted rules based on lexical and other linguistic knowledge (e.g. ENGTWOL 95)
- **Stochastic**: Trained on human annotated corpora like the Penn Treebank
  - **Statistical models**: Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM), Conditional Random Field (CRF), log-linear models, support vector machines
  - **Rule learning**: Transformation Based Learning (TBL)

- Many English POS-taggers are publicly available
- Hindi / Bangla POS tagger:
  - http://nltr.org/snltr-software/

Deciding on a POS tagset

NOUN	The DOG barked.	WE saw YOU.
VERB	The dog BARKED.	It IS impossible.
ADJECTIVE	He's very OLD.	I've got a NEW car.
DETERMINATIVE	THE dog barked.	I need SOME nails.
ADVERB	She spoke CLEARLY.	He's VERY old.
PREPOSITION	It's IN the car.	I gave it TO Sam.
COORDINATOR	I got up AND left.	It's cheap BUT strong.
SUBORDINATOR	It's odd THAT they were late.	I wonder WHETHER it's still there.
INTERJECTOR	OH, HELLO, WOW, OUCH	

from [huddleston-pullum 05] *Student's intro to English Grammar*

Coordinator / subordinator: markers for coordinate / subordinate clauses
POS distinctions based on analysis of syntax and semantics

POS Tagset

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	*and, but, or*	SYM	Symbol	+,%, &
CD	Cardinal number	*one, two, three*	TO	"to"	*to*
DT	Determiner	*a, the*	UH	Interjection	*ah, oops*
EX	Existential 'there'	*there*	VB	Verb, base form	*eat*
FW	Foreign word	*mea culpa*	VBD	Verb, past tense	*ate*
IN	Preposition/sub-conj	*of, in, by*	VBG	Verb, gerund	*eating*
JJ	Adjective	*yellow*	VBN	Verb, past participle	*eaten*
JJR	Adj., comparative	*bigger*	VBP	Verb, non-3sg pres	*eat*
JJS	Adj., superlative	*wildest*	VBZ	Verb, 3sg pres	*eats*
LS	List item marker	*1, 2, One*	WDT	Wh-determiner	*which, that*
MD	Modal	*can, should*	WP	Wh-pronoun	*what, who*
NN	Noun, sing. or mass	*llama*	WP$	Possessive wh-	*whose*
NNS	Noun, plural	*llamas*	WRB	Wh-adverb	*how, where*
NNP	Proper noun, singular	*IBM*	$	Dollar sign	$
NNPS	Proper noun, plural	*Carolinas*	#	Pound sign	#
PDT	Predeterminer	*all, both*	"	Left quote	(' or ")
POS	Possessive ending	*'s*	"	Right quote	(' or ")
PP	Personal pronoun	*I, you, he*	(	Left parenthesis	([, (, {, <)
PP$	Possessive pronoun	*your, one's*	)	Right parenthesis	( ],), }, >)
RB	Adverb	*quickly, never*	,	Comma	,
RBR	Adverb, comparative	*faster*	.	Sentence-final punc	(. ! ?)
RBS	Adverb, superlative	*fastest*	:	Mid-sentence punc	(: ; ... – -)
RP	Particle	*up, off*			

Figure 8.6 Penn Treebank Part-of-Speech Tags (Including Punctuation)

Penn Treebank
[Marcus etal 93]

Figure: jurafsky-martin ch.8 (2000)

Rule-based POS: Attributes/Features

Word	POS	Additional POS features
smaller	ADJ	COMPARATIVE
entire	ADJ	ABSOLUTE ATTRIBUTIVE
fast	ADV	SUPERLATIVE
that	DET	CENTRAL DEMONSTRATIVE SG
all	DET	PREDETERMINER SG/PL QUANTIFIER
dog's	N	GENITIVE SG
furniture	N	NOMINATIVE SG NOINDEFDETERMINER
one-third	NUM	SG
she	PRON	PERSONAL FEMININE NOMINATIVE SG3
show	V	IMPERATIVE VFIN
show	V	PRESENT -SG3 VFIN
show	N	NOMINATIVE SG
shown	PCP2	SVOO SVO SV
occurred	PCP2	SV
occurred	V	PAST VFIN SV

Attributes (Hindi)

आयकर\NC.0.sg.dir.0 आयुक्त\NC.0.sg.obl.gen (\PU
अपील्स\NC.fem.sg.dir.0)\RDS के\PP.0.0.gen
आदेशों\NC.mas.pl.obl.abl से\PP.0.0.abl पीडित\JJ.0.0.dir
निर्धारिती\NC.0.sg.dir.0 ,\PU अपीलीय\JJ.0.0.dir
न्यायाधिकरण\NC.mas.sg.obl.gen के\PP.0.0.gen
समक्ष\NST.dir.0 अगली\NST.dir.0
अपील\NC.fem.sg.dir.0 कर\VAUX.0.0.0.0.0.0.nfn.0
सकता\VAUX.mas.sg.3.prs.pft.dcl.fin.n
है\VAUX.0.sg.3.prs.pft.dcl.fin.n

Rule-based POS: Lexicon lookup

Pavlov	PAVLOV N NOM SG PROPER	
had	HAVE V PAST VFIN SVO	
	HAVE PCP2 SVO	
shown	SHOW PCP2 SVOO SVO SV	(past participle)
that	ADV	
	PRON DEM SG	
	DET CENTRAL DEM SG	
	CS	(complementizer / subordinator)
salivation	N NOM SG	

Rule-based POS: Apply Rules

- Apply constraints to eliminate choices
- ENGTWOL: 1100 rules, e.g.

ADVERBIAL-THAT RULE
Given input: "that"
if
(+1 A/ADV/QUANT); /* *if next word is adj, adverb, or quantifier* */
(+2 SENT-LIM); /* *and following which is a sentence boundary,* */
(NOT -1 SVOC/A); /* *and the previous word is not a verb like* */
/* *'consider' which allows adjs as object complements* */
then eliminate non-ADV tags
else eliminate ADV tag

Stochastic POS-tagging

- Markovian assumption : tag depends on limited set of previous tags

- HMM:

  maximize P(word|tag) * P(tag| previous n tags)

- Maximize the probability for whole sentence, not single word

$$S = \arg\max_{t_1 \ldots t_n} \prod_{i=1,n} P(w_i \mid t_i) P(t_i \mid t_{i-1})$$

Stochastic POS-tagging

- Secretariat/NNP is/VBZ expected/VBN to/TO race/VB tomorrow/NN
- People/NNS continue/VBP to/TO inquire/VB the/DT reason/NN for/IN the/DT race/NN for/IN outer/JJ space/NN
- *to race* vs. *the race*

Stochastic POS-tagging

- *to/TO race* *the/DT race*

- P(VB|TO) P(*race*|VB)
- P(NN|TO) P(*race*|NN)

- P(NN|TO) = .021 P(race|NN) = .00041
- P(VB|TO) = .34 P(race|VB) = .00003

- P(VB|TO)P(race|VB) = .00001
- P(NN|TO)P(race|NN) = .000007

Weakly-supervised POS-tagging

Small training data

Automatic Part-of-Speech Tagging for Bengali: An Approach for Morphologically Rich Languages in a Poor Resource Scenario

Sandipan Dandapat, Sudeshna Sarkar, Anupam Basu
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
India 721302
{sandipan,sudeshna,anupam.basu}@cse.iitkgp.ernet.in

Abstract

This paper describes our work on building Part-of-Speech (POS) tagger for Bengali. We have use Hidden Markov Model (HMM) and Maximum Entropy (ME) based stochastic taggers. Bengali is a morphologically rich language and our taggers make use of morphological and contextual information of the words. Since only a small labeled training set is available (45,000 words), simple stochastic approach does not yield very good results. In this work, we have studied the effect of using a morphological analyzer to improve the performance of the tagger.

ditional information may be coded into the HMM model to achieve higher accuracy (Cutting et al., 1992). The semi-supervised model described in Cutting et al. (1992), makes use of both labeled training text and some amount of unlabeled text. Incorporating a diverse set of overlapping features in a HMM-based tagger is difficult and complicates the smoothing typically used for such taggers. In contrast, methods based on Maximum Entropy (Ratnaparkhi, 1996), Conditional Random Field (Shrivastav, 2006) etc. can deal with diverse, overlapping features.

1.1 Previous Work on Indian Language POS Tagging

Although some work has been done on POS tagging of different Indian languages, the systems

Weakly-supervised POS-tagging

- HMM models:

  maximize over sentence P(word|tag) * P(tag| previous n tags)

$$S = \arg\max_{t_1 \ldots t_n} \prod_{i=1,n} P(w_i \mid t_i) P(t_i \mid t_{i-1})$$

- Maximum Entropy: estimate probabilities based on constraints (derived from training data)

$$p(t_1 \ldots t_n \mid w_1 \ldots w_n) = \prod_{i=1,n} p(t_i \mid h_i)$$

Weakly-supervised POS-tagging

- Morphologically rich languages: Can constrain based on morphology

Unsupervised POS-tagging

Method	Accuracy		
	10K	20K	40K
HMM-S	57.53	70.61	77.29
HMM-S+suf	75.12	79.76	83.85
HMM-S+MA	82.39	84.06	86.64
HMM-S+suf+MA	84.73	87.35	88.75
HMM-SS	63.40	70.67	77.16
HMM-SS+suf	75.08	79.31	83.76
HMM-SS+MA	83.04	84.47	86.41
HMM-SS+suf+MA	84.41	87.16	87.95
ME	74.37	79.50	84.56
ME+suf	77.38	82.63	86.78
ME+MA	82.34	84.97	87.38
ME+suf+MA	84.13	87.07	88.41

POS categories - Unsupervised

$$\begin{bmatrix} ball \\ block \\ box \\ circle \\ square \end{bmatrix} \quad \begin{bmatrix} in \\ inside \\ into \end{bmatrix} \quad \begin{bmatrix} chases \\ pushes \\ corners \\ the \end{bmatrix} \quad \begin{bmatrix} big \\ large \\ little \\ the \end{bmatrix}$$

[mukerjee nayak 12] based on ADIOS
[solan rupin edelman 05]

STATISTICAL NATURAL LANGUAGE PARSING

Unsupervised POS and Syntax: Grounded Models

Grounded Language

- grounded **lexicon**:
  relation between sounds and sensorimotor patterns

- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Minimal Commitment

- minimize prior knowledge in agent:
  - preference: minimize description lengths
    → inventory of machine learning algorithms
  - no knowledge of grammar – no POS tags, no syntactic structure
  - no knowledge of domain
- **bootstrapping** stage:
  - semantic schemas come first
  - language regularities later

POS categories – can we discover them?

$$\begin{bmatrix} ball \\ block \\ box \\ circle \\ square \end{bmatrix} \begin{bmatrix} in \\ inside \\ into \end{bmatrix} \begin{bmatrix} chases \\ pushes \\ corners \\ the \end{bmatrix} \quad \begin{bmatrix} big \\ large \\ little \\ the \end{bmatrix}$$

[nayak mukerjee COLING-12] based on
ADIOS [solan rupin edelman 05]

Minimal Commitment Acquisition

Previous Work: Unsupervised Semantics

- single word or phrase learning   [no grammar]
    - Hand-coded propositional (T/F) semantics
        - [plunkett etal 92]   [siskind, 94/03] (phrases)
        - [regier 96] (prepositions)
        - [steels 03] [roy/reiter 05]   [caza/knott 12]
    - Supervised Learning of semantics
        - [kate/mooney 06] : set of predicates are known
        - [yu/ballard 07] : semantics = scene-region

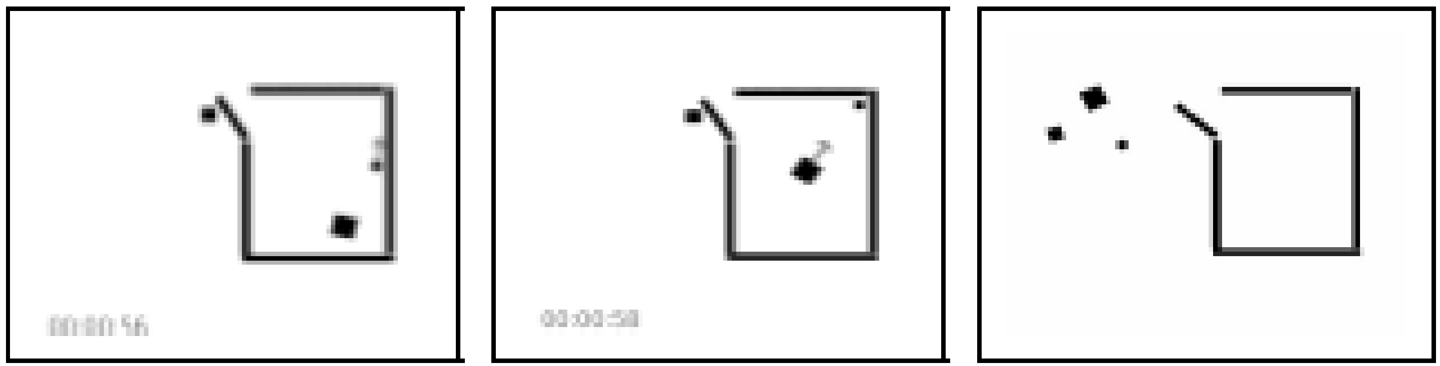
- Unsupervised Semantic Acquisition :
  "right" granularity for concepts; dynamic predicates

Previous Work: Grammar

- Grammar learning:
  - Grammatical categories:
    - [redington etal 98] (RNN)
    - [wang / mintz 07] (frequent frame)
  - Grammar induction : Structure is known
    - No semantics:
      - [marino etal 07] [solan edelman 05]
    - Propositional semantics
      - [dominey /boucher 05]
      - [kwiatkowski zettlemoyer 10] (SVM)
      - [kim/mooney 12] (altered visual input)

Language Acquisition : Domains

- Perceptual input



[heider/simmel 1944] [hard/tversky 2003]

- Discovery Targets:
  - semantics: objects, 2-agent actions, relations
  - lexicon : nominal, transitive verbs, preposition
  - lexical categories: N VT P Adj
  - constructions: PP VP S
  - sense extension (metaphor)   [nayak/mukerjee (AAAI-12)]

Language Acquisition : Domain 2

- Perceptual input

- Discovery Targets:
  - semantics: object categories, motion categories

[mukerjee / joshi RANLP 11]

Language Acquisition : Domain 2

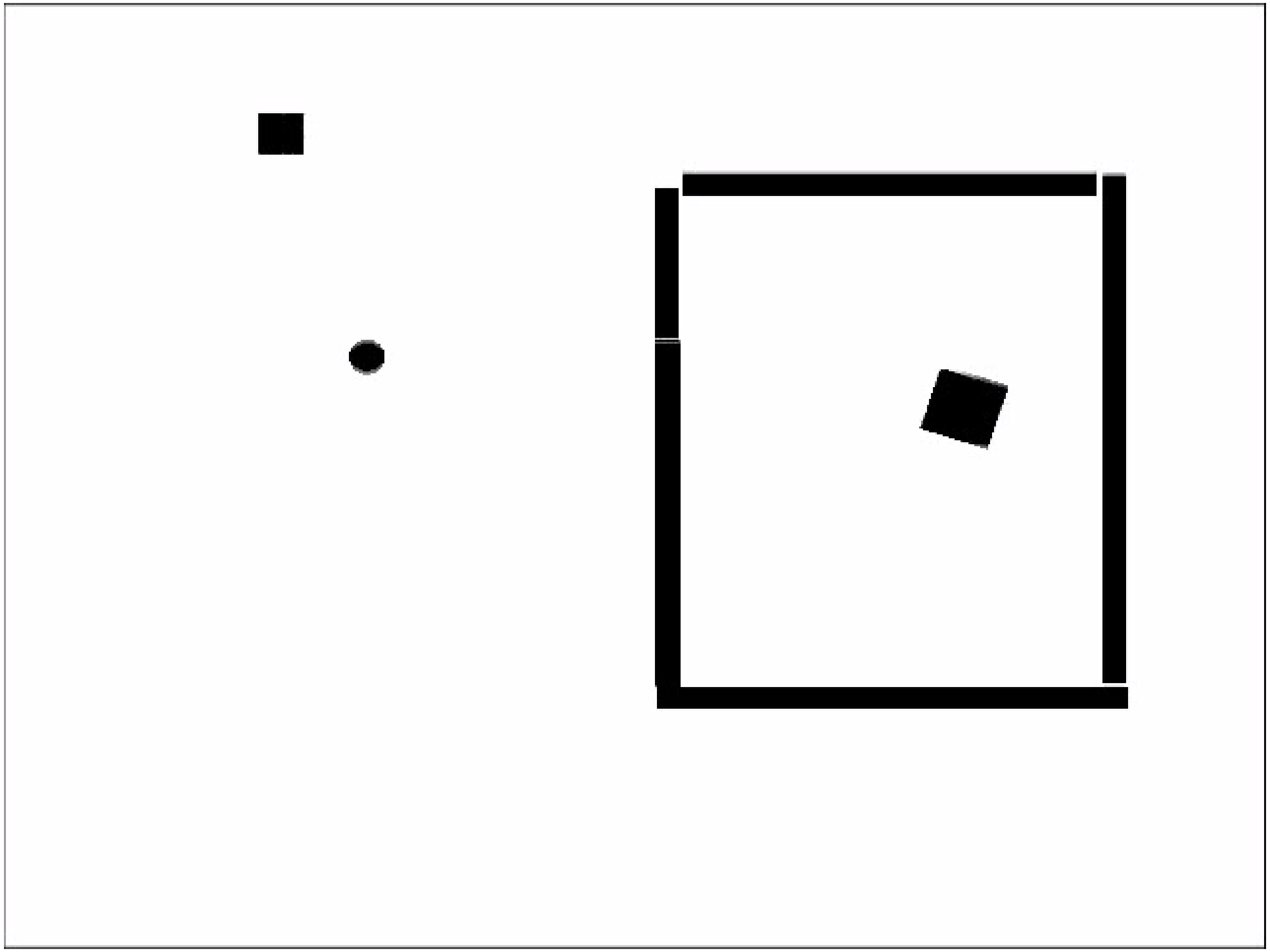
- object categories

- Discovery Targets:

[mukerjee / joshi RANLP 11]

  - semantics: object categories, motion categories
  - lexicon : word boundaries, nominals, intransitive verbs
  - construction: intransitive VP

Video Fragment



Linguistic input

- input = description commentaries transcribed into text
    - 48 descriptions in English / 10 : Hindi

- Unconstrained description by different subjects:

•the little square hit the big square
•they're hitting each other
•the big square hit the little square
•circle and square in [unitelligible stammer]
•the two squares stopped fighting

•छोटा बक्सा बडा बक्सा मे कुछ बातचीत होती है
little box big box between some talk happens

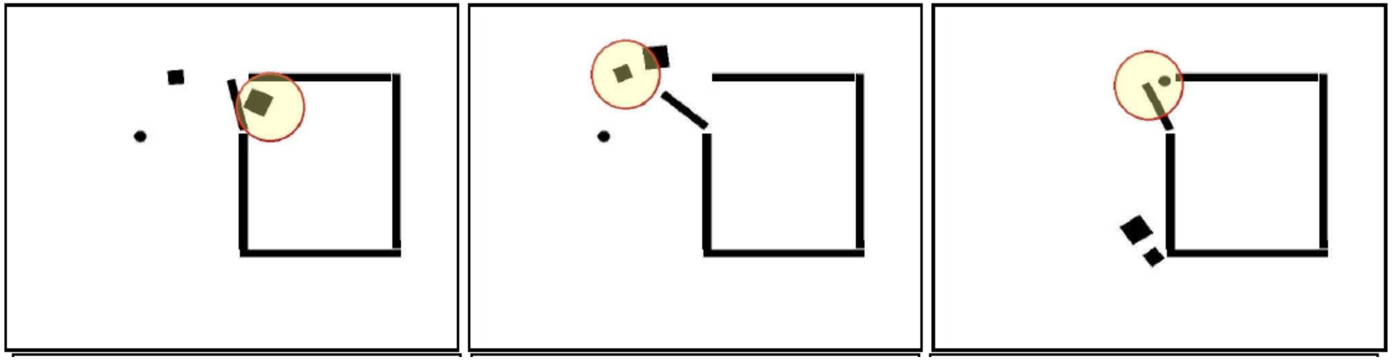
Discovering Language

- Perceptual structure discovery:
  - Given perceptual space *W* discover set of structures Γ that partition it into patterns relevant to agents goals.
  - Elements $\gamma \in \Gamma$ constitute a hierarchy; structures learned earlier are used for more complex patterns
- Linguistic Structure Discovery
  - Given set of sentences formed from words w $\in L$, discover set of subsequences Λ that result in a more compact description of the structure
  - Elements $\lambda \in \Lambda$ constitute a hierarchy, leaf nodes (POS) are subsets of L

Semantics First:
Objects / Nominals

Language Grounding: Entity/Object

- object = coherent salient region in perceptual space
  - object view schema [white maruti 800 from camera 1]
  - object schema [white maruti 800]
  - object category schema [car]
- bottom-up dynamic attention



[singh maji mukerjee 06]

Language – Meaning Associaction

- Relative Association (bayesian)

$$P(\gamma_j|\lambda_i) = \frac{P(\lambda_i|\gamma_j)P(\gamma_j)}{P(\lambda_i)} \propto \frac{P(\lambda_i|\gamma_j)}{P(\lambda_i)}$$

- Mutual association (contribution to M.I.)

$$P(\lambda_i,\gamma_j)\log\frac{P(\lambda_i,\gamma_j)}{P(\lambda_i)P(\gamma_j)}$$

$$I(\Gamma,\Lambda) = \sum_i\sum_j P(\lambda_i,\gamma_j)\log\frac{P(\lambda_i,\gamma_j)}{P(\lambda_i)P(\gamma_j)}$$

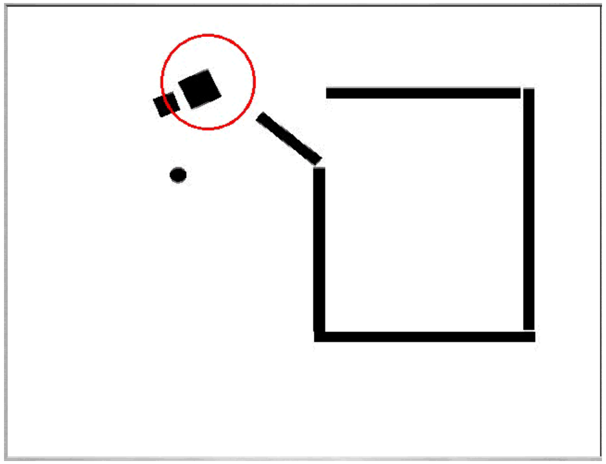
Language Grounding: Nominals

[BS]			[SS]			[C]		
word(s)	A_{ij}^{rel}	A_{ij}^{mut}	word(s)	A_{ij}^{rel}	A_{ij}^{mut}	word(s)	A_{ij}^{rel}	A_{ij}^{mut}
square	0.70	1.41	little	0.66	0.79	circle	0.79	2.11
big	0.89	1.11	small	0.72	0.63	square	0.41	1.54
box	0.69	0.78	square	0.46	1.12	little	0.68	1.22
the big	0.87	0.71	small square	0.93	0.53	the little	0.71	0.81
big square	0.94	0.75	little square	0.89	0.46	little circle	0.91	0.60
large square	0.86	0.15	the little	0.70	0.54	the big	0.48	0.61

Perceptual Discovery : Actions : Verbs

Perceptual Discovery: 2-agent actions

- Consider every pair of objects A,B
  A : attended to object (tr)
  B : other object (landmark, lm).



- 2 features suffice:

relative-velocity and relative position

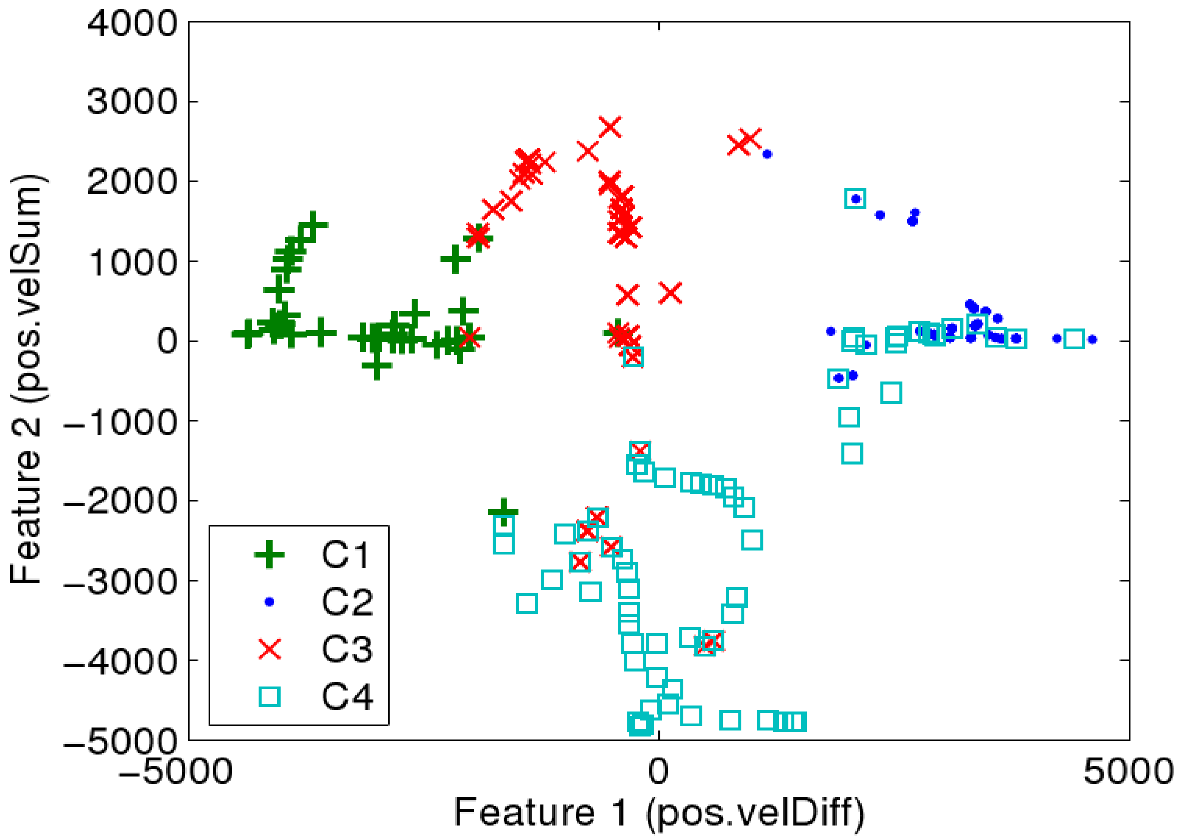
$$pos{\cdot}velDiff : (\vec{x}_B - \vec{x}_A) \cdot (\vec{v}_B - \vec{v}_A)$$

relative pose and the sum of the velocities

$$pos{\cdot}velSum : (\vec{x}_B - \vec{x}_A) \cdot (\vec{v}_B + \vec{v}_A)$$

Perceptual Discovery: 2-agent actions

- Static time-shots of feature space trajectories



Emergent Clusters

- Human Labels *(CC, MA, Chase)* → Ground Truth
- Label Vs Cluster assigned

	C_1	C_2	C_3	C_4	Total	%	TCA
CC	399	6	10	29	444	90	
MA	16	311	5	48	380	82	84
Chase	21	59	149	154	383	79	

Number of Clusters from MNG = 4 when *Edge Aging* = 30 (0.9 prob)

CC: Come-Closer (C1), MA: Move Away (C2), C_3 & C_4 : Chase

Chase sub-categories:

Chase_*RO-chases-LO*: C_3 →

Chase_*LO-chases-RO*: C_4 →

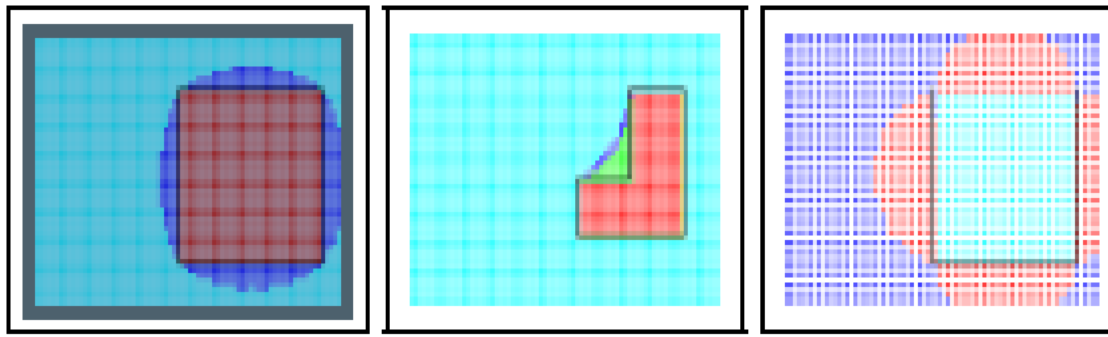
Learning verbs

Discovering Containment Relations : Prepositions

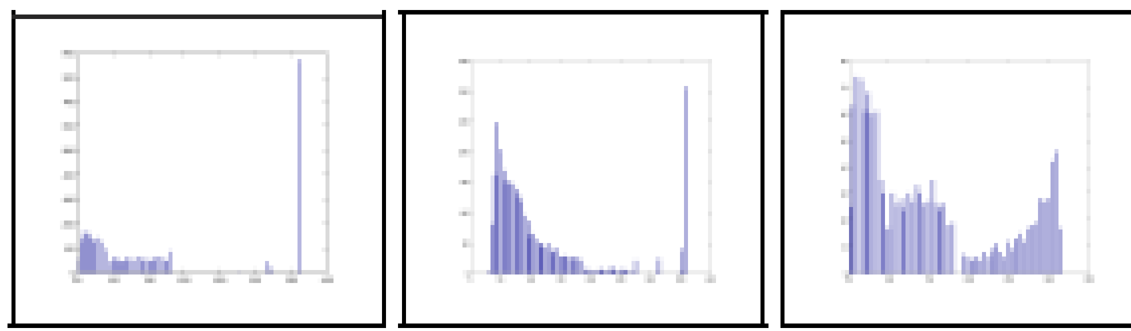
Clustering spatial relations

Feature Commitment:
Visual angle subtended
at trajector by landmark

Meanshift clusters
on subtended
visual angle for
diff shapes



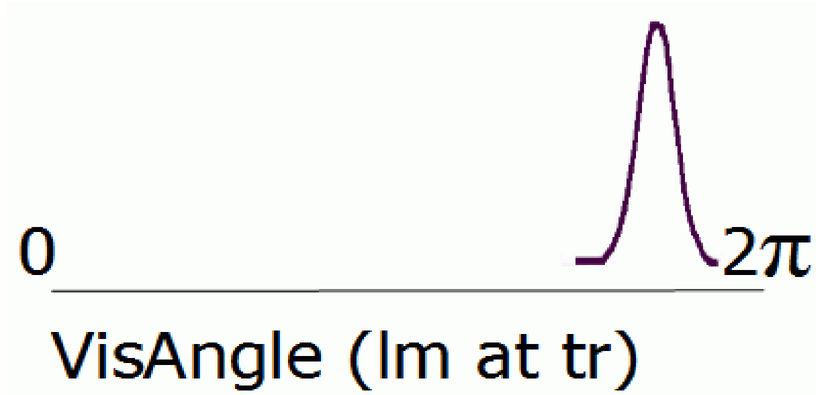
Histogram of visual
subtended angle
for the 3 shapes



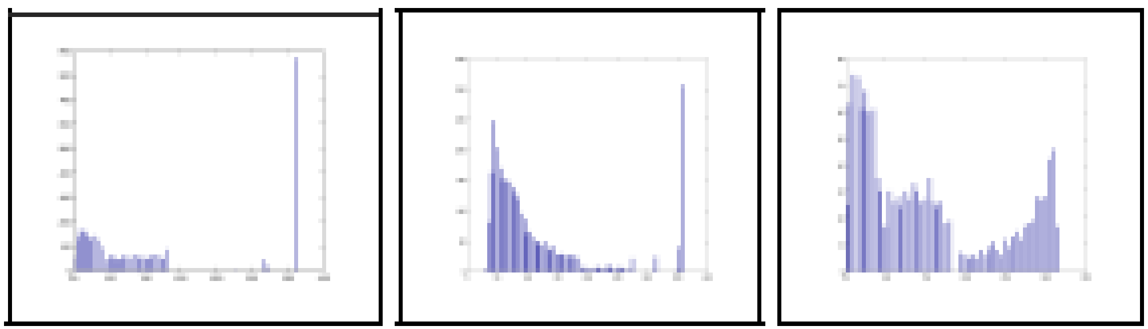
[Sarkar/Mukerjee 07; Nayak/Mukerjee 12]

Clustering spatial relations

IN cluster
(emergent)

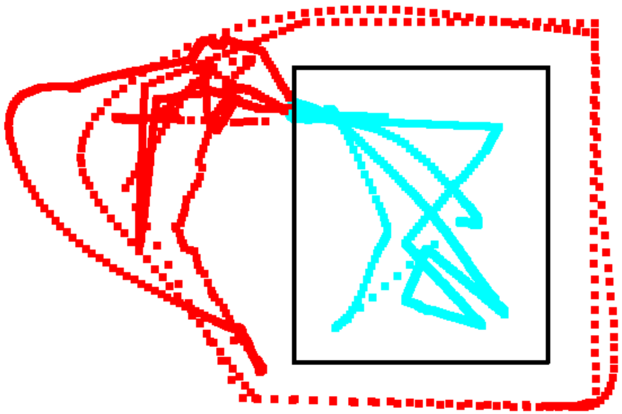


Histogram of visual subtended angle for the 3 shapes



[Sarkar/Mukerjee 07; Nayak/Mukerjee 12]

Words for motions ending in / out



IN	A_{ij}^{rel}	A_{ij}^{mut}	INTO	A_{ij}^{rel}	A_{ij}^{mut}	OUT OF	A_{ij}^{rel}	A_{ij}^{mut}
inside	0.79	11.78	into	0.82	6.98	out	0.65	5.71
into	0.90	9.43	inside	0.53	1.03	leaves	1.00	4.16
in	0.61	4.16	enters	1.00	4.85	exits	1.00	3.46

Syntax discovery and Semantic Association

Syntax Discovery

- Syntactic discovery:
  - Given input text, attempt to find graph that results in minimizing the description length
  - Relational Graph RDS: patterns as nodes; edges as transitions
  - Attempt to edit RDS to detect significant patterns
  - Equivalence classes emerge at the nodes

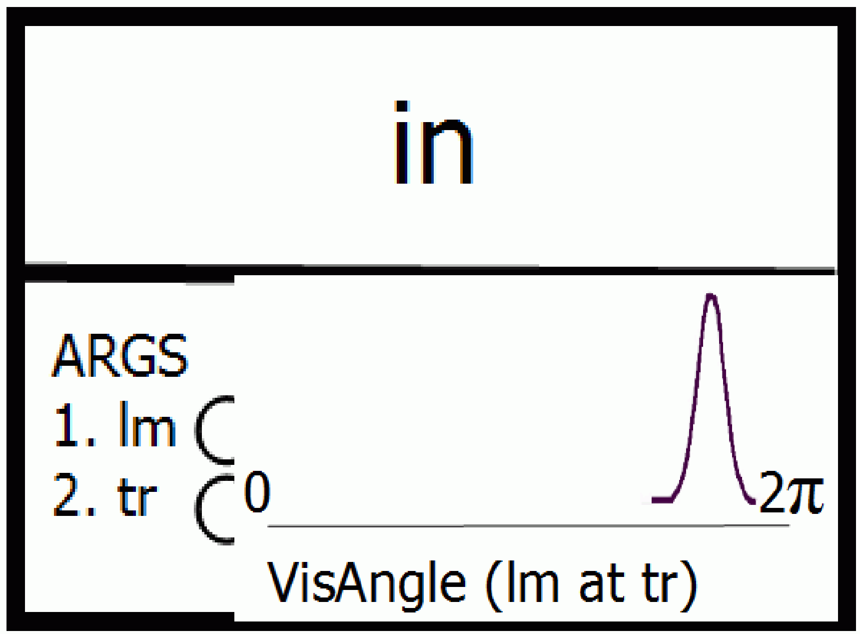
$$\left[\left[\begin{array}{c} in \\ \\ inside \\ \\ into \end{array}\right] \rightarrow the\right] \rightarrow box$$

ADIOS [solan / edelman 05]

Computing the Image Schema

Our reflective baby
has discovered:
"in" = label corresponding to
this image schema

Hence: symbol for [IN] is



(note: this is an early, very basic,
low-confidence characterization

Language Structures : Verbs

$$1.\ \begin{bmatrix} \mathit{the} \rightarrow \begin{bmatrix} \mathit{big} \\ \mathit{large} \end{bmatrix} \rightarrow \mathit{square} \\ \mathit{the} \rightarrow \mathit{square} \end{bmatrix} \rightarrow \begin{bmatrix} \mathit{scares} \\ \mathit{approaches} \\ \mathit{chases} \end{bmatrix} \rightarrow \begin{bmatrix} \mathit{the} \rightarrow \begin{bmatrix} \mathit{small} \\ \mathit{little} \end{bmatrix} \end{bmatrix}$$

$$2.\ \begin{bmatrix} \mathit{the} \rightarrow \begin{bmatrix} \mathit{ball} \\ \mathit{box} \\ \mathit{door} \\ \mathit{square} \end{bmatrix} \\ \mathit{circle} \\ \mathit{it} \end{bmatrix} \rightarrow \begin{bmatrix} \mathit{moved} \\ \mathit{moves} \\ \mathit{runs} \end{bmatrix}$$

ADIOS [solan / edelman 05]

Hindi Acquisition: Word learning

[BS]			[SS]			[C]			[IN]		
word(s)	A^{rel}_{ij}	A^{m}_{ij}	word(s)	A^{rel}_{ij}	A^{m}_{ij}	word(s)	A^{rel}_{ij}	A^{m}_{ij}	word(s)	A^{rel}_{ij}	A^{m}_{ij}
बक्सा baksA/box	.77	.37	बक्सा baksA/box	.62	.44	गोला golA/ball	.83	.54	अन्दर andar/in	.80	1.30
बडा(badA/ big) बक्सा	.85	.18	छोटा(chota/ small) बक्सा	.90	.25	बक्से के(ke/–)	.63	.27	बाहर (bA-har/out)	.78	.73

Incipient Syntax

$$\left[\begin{array}{c}\text{डब्बे(dabbA/box)}\\ \text{बक्से(bakse/box)}\end{array}\right] \rightarrow \begin{array}{c}\text{के}\\ \text{(ke/–)}\end{array} \rightarrow \left[\begin{array}{ccc}\begin{array}{c}\text{बाहर}\\ \text{(bAhar/out)}\end{array} & \left[\begin{array}{c}\text{बाहर(bAhar/out)}\\ \text{आ(aa/come)}\\ \text{भाग(bhAg/run)}\end{array}\right] & \begin{array}{c}\text{जाता}\\ \text{(jAtA/goes)}\end{array}\end{array}\right]$$