

Natural Language Processing

Spring 2017

Professor Liang Huang

Doesn't Google know everything?

What animal does a cat eat?

What animal does a cat eat

About 10,700,000 results (0.24 seconds)

[Advanced search](#)

[What wild animal would eat a domestic cat or dog? - Yahoo!7 Answers](#)

Aug 21, 2008 ... I went outside to get my kitty in a few minutes ago and couldn't ... I hope it's not what you think, I really **do**. But there are wild **animals** ...

[au.answers.yahoo.com](#) › [Pets](#) › [Cats](#) - [Cached](#) - [Similar](#)

[What If my cat eats rawhide?](#) - 2 answers - Aug 14, 2010

[Why does my cat eat grass?](#) - 15 answers - Nov 2, 2008

[More results from au.answers.yahoo.com](#) »

[WikiAnswers - What animal eats cat](#)

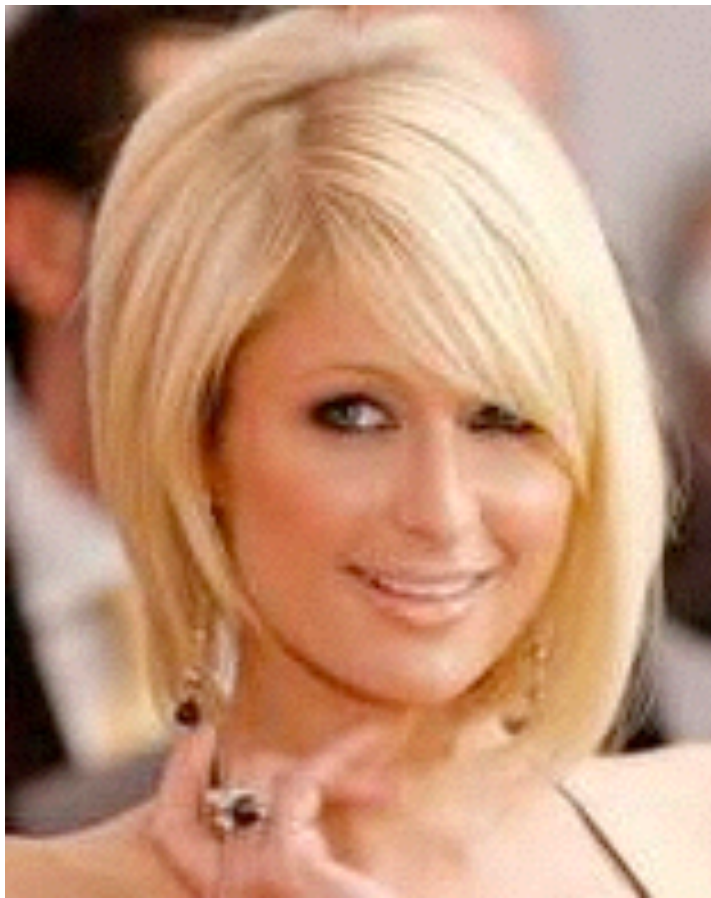
Wild **Animals** question: **What animal eats cat?** Mountain lions **do**, as will coyotes. I've heard of bobcats eating cats too, but i'm not sure it's true.

[wiki.answers.com/Q/What_animal_eats_cat](#) - [Cached](#) - [Similar](#)

Retrieved August 2010

Even Key Word Queries

- Paris Hilton -- not easy to book! (vs. Boston Hilton)



Ambiguity

Where can I spot a snow leopard?

Where can I spot a snow leopard

About 27,400,000 results (0.19 seconds)



[Apple - Mac OS X Snow Leopard - The world's most advanced OS](#)

To advance Mac OS X **Leopard**, Apple engineers went deep into the code to streamline, secure, and add new core technologies.

[Buy Mac OS X Snow Leopard now. - Compatibility - Desktop](#)

[www.apple.com/macosx/ - Cached - Similar](#)

[Snow Leopard - Wikipedia, the free encyclopedia](#)

The **snow leopard** (*Uncia uncia*) is a moderately large cat native to the ... their body with small **spots** of the same color on their heads and larger **spots** on ...

[Taxonomy - Etymology - Distribution - Ecology and behaviour](#)

[en.wikipedia.org/wiki/Snow_Leopard - Cached - Similar](#)

[Mac OS X Snow Leopard - Wikipedia, the free encyclopedia](#)

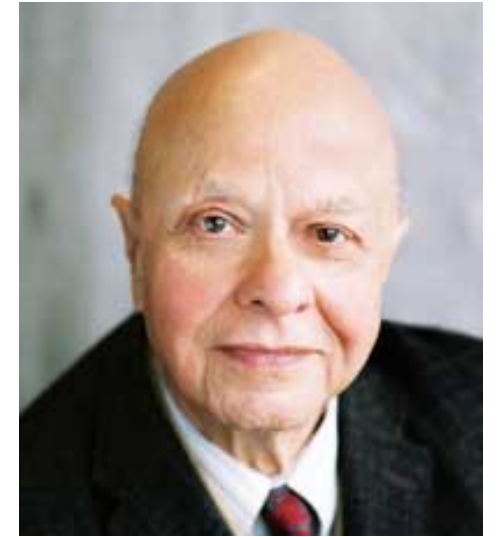
Mac OS X **Snow Leopard** (version 10.6) is the seventh and current major release of Mac OS X, Apple's desktop and server certified Unix operating system. ...

[en.wikipedia.org/wiki/Mac_OS_X_Snow_Leopard - Cached - Similar](#)

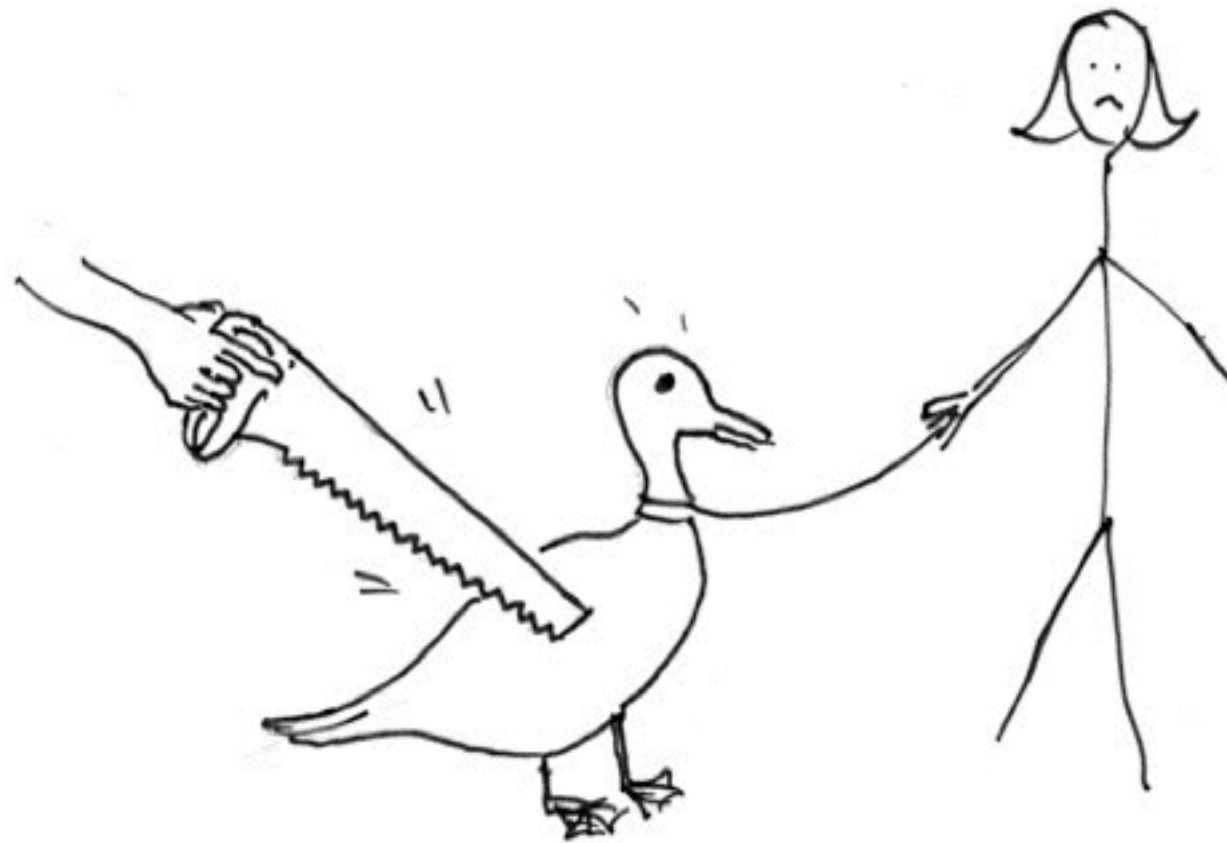
More about Ambiguities

- to middle school kids: what does this sentence mean?

I saw her duck.



Aravind Joshi

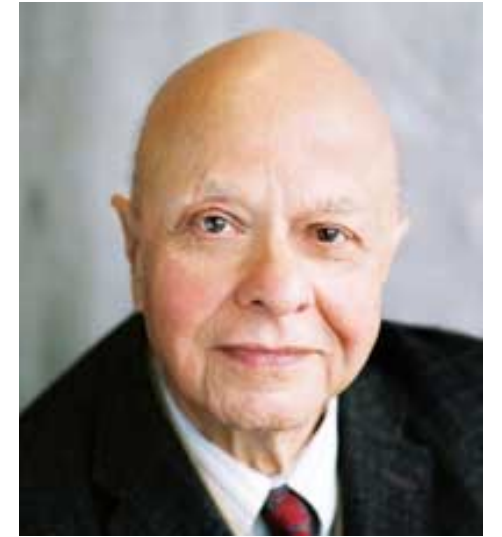


lexical ambiguity
(word-sense)

More about Ambiguities

- to middle school kids: what does this sentence mean?

I eat sushi with tuna.



Aravind Joshi

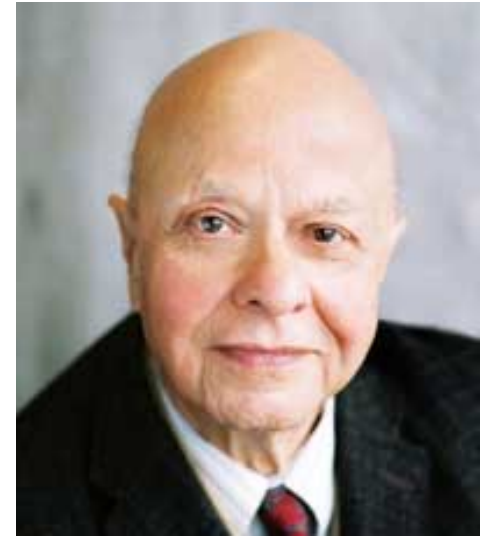


structural ambiguity
(PP-attachment)

More about Ambiguities

- to middle school kids: what does this sentence mean?

I eat sushi with tuna.



Aravind Joshi

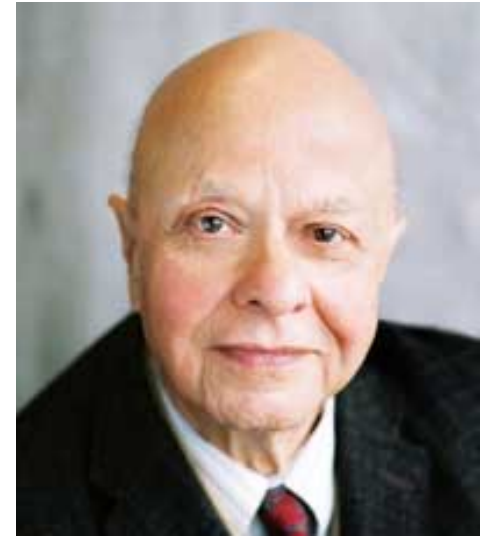


lexical ambiguity
(word-sense)

More about Ambiguities

- to middle school kids: what does this sentence mean?

Everybody loves somebody.



Aravind Joshi



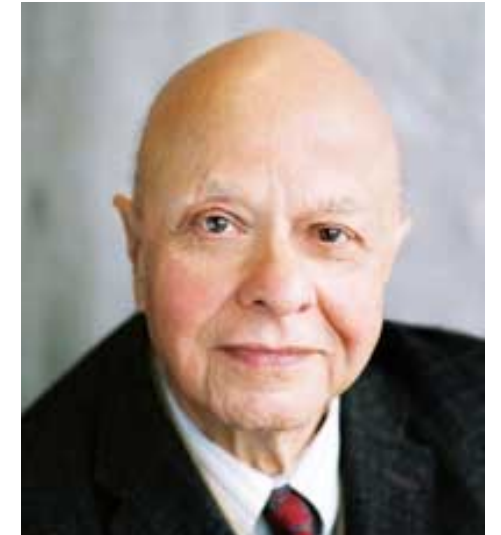
???

structural ambiguity
(quantifier scope)

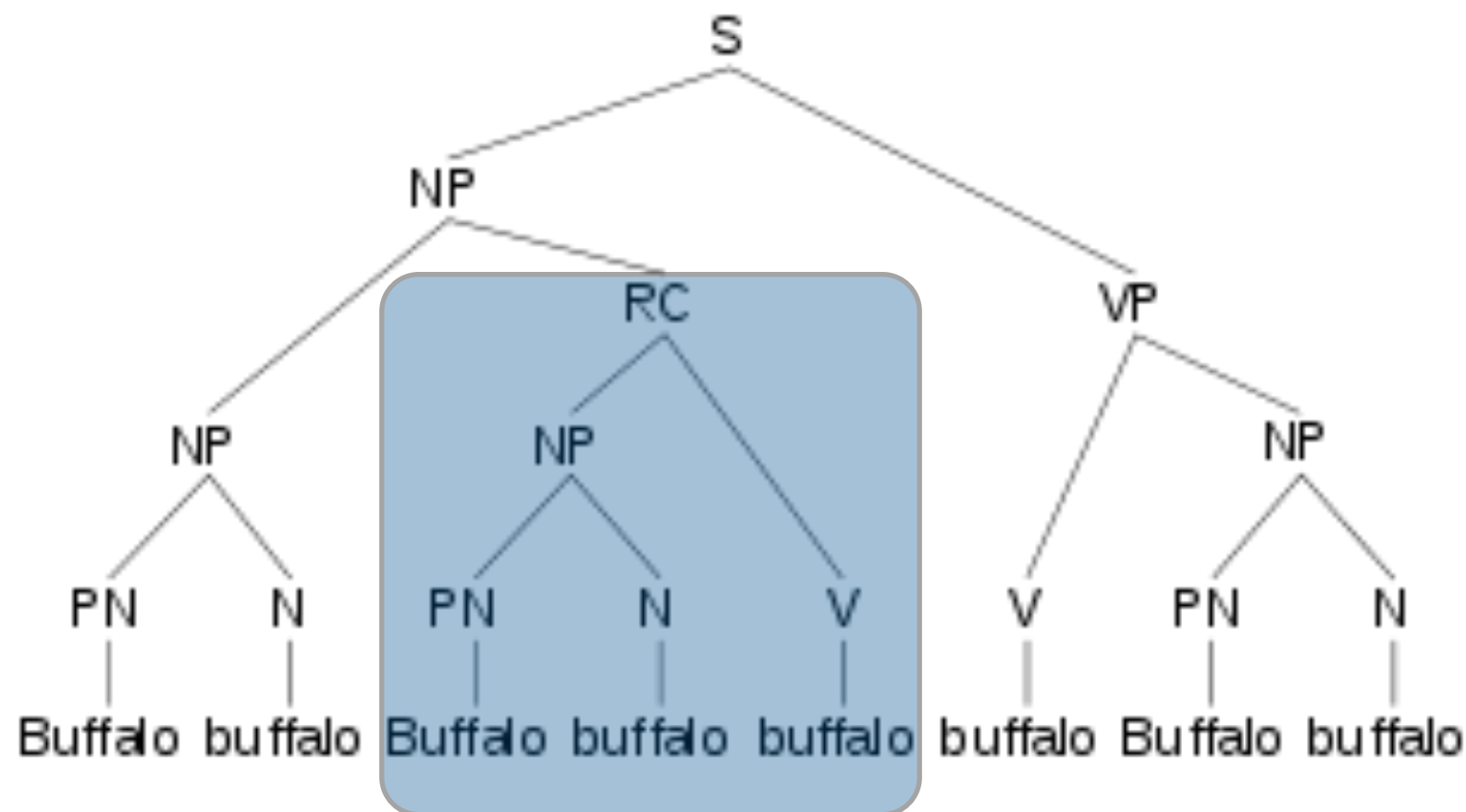
More about Ambiguities

- to middle school kids: what does this sentence mean?

Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo



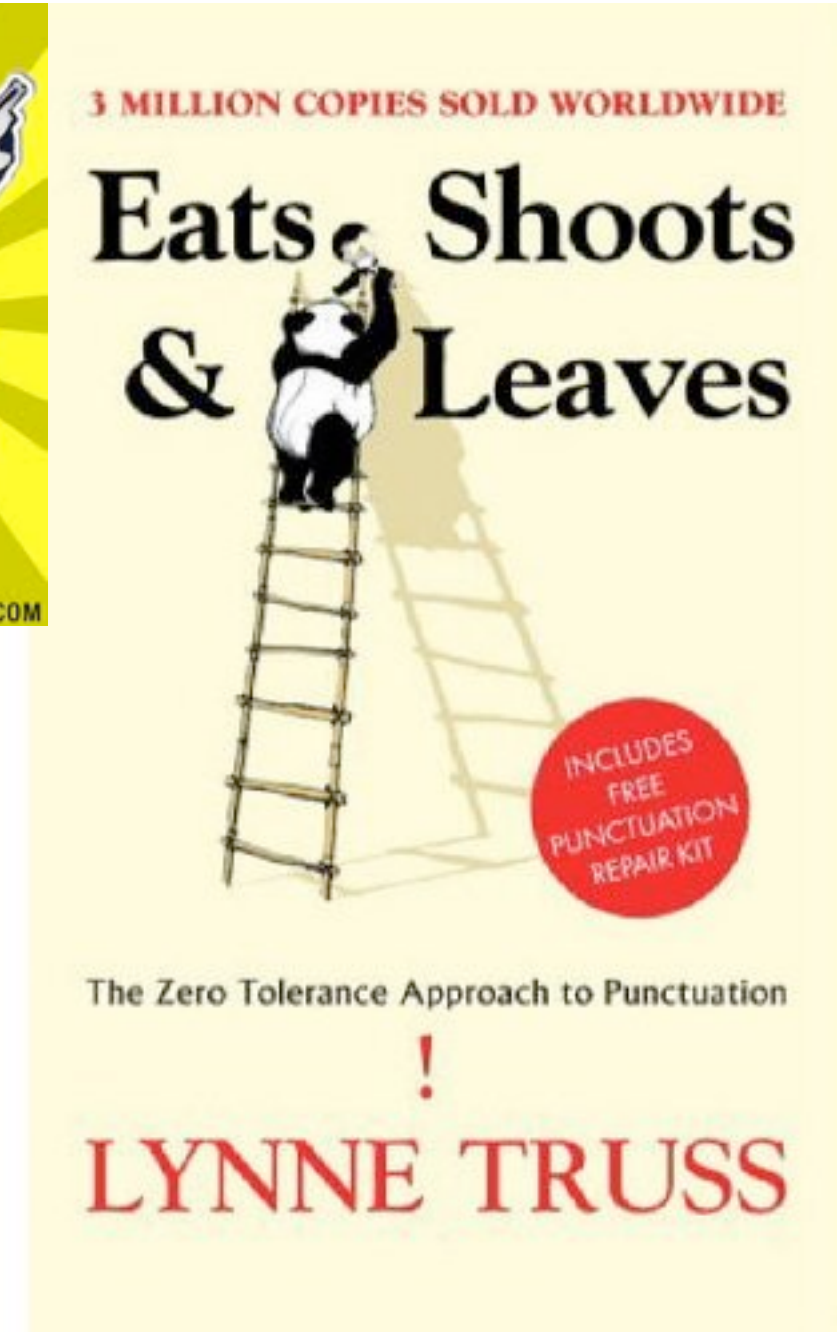
Aravind Joshi



Dogs dogs dog dog dogs.
Police police police police police

Prosody and Ambiguity

- a panda
 - eats [shoots]_N and [leaves]_N
 - eats [shoots]_V and [leaves]_V
- prosody marks this ambiguity by
 - prominence on *eats*
 - break between *eats* and *shoots*



Ambiguities in Translation



zi zhu zhong duan
自 助 终 端

self help terminal device

Ambiguities in Translation



Google translate: carefully slide



If you are stolen...

一旦失窃要报警，切莫姑息又养奸

If you are stolen, call the police at once.

Google translate: Once the theft to the police



上海市公安局城市轨道交通分局

Urban Mass Transportation Branch Shanghai Public Security Bureau



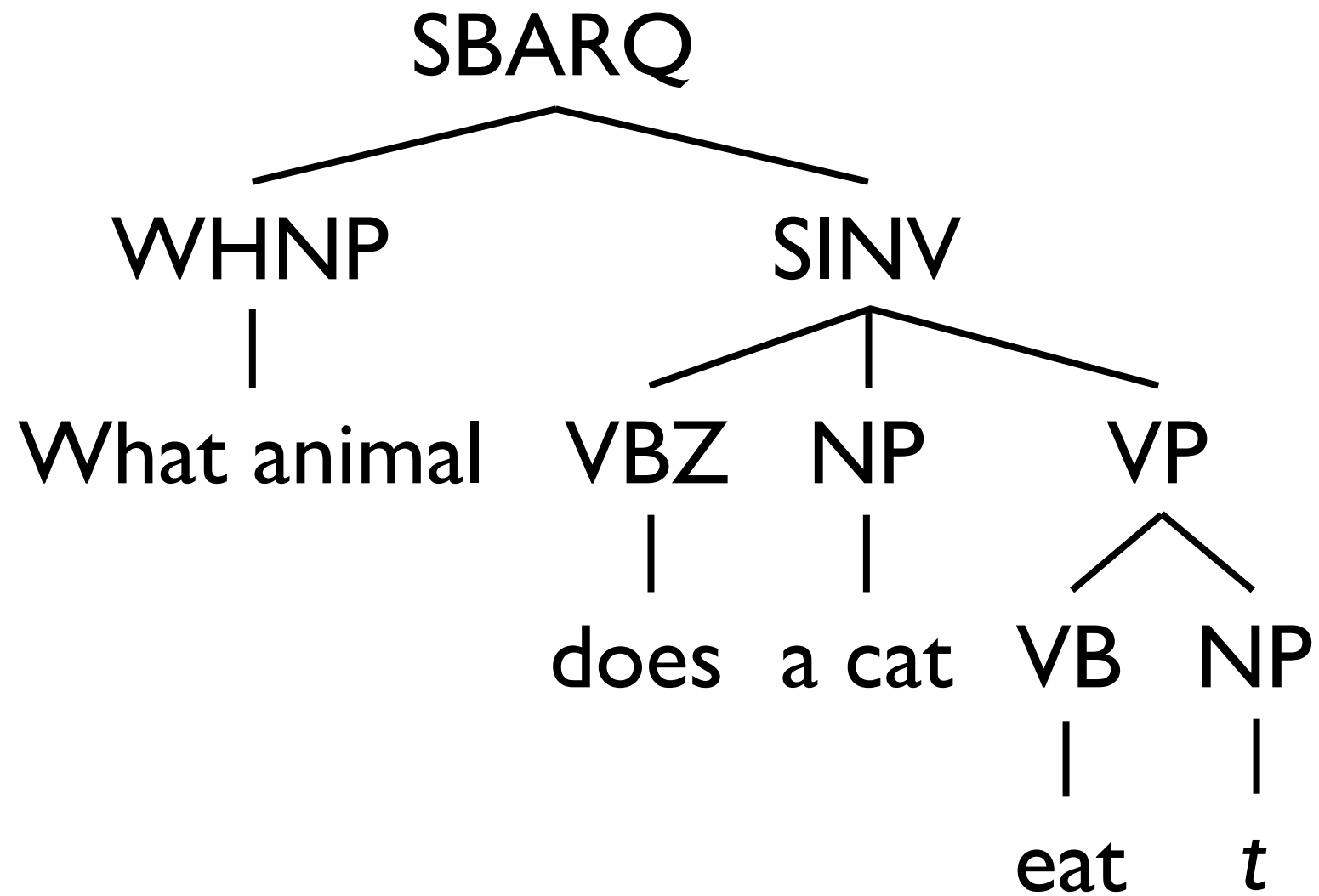
ENGLISH FUNNY.com

or even...



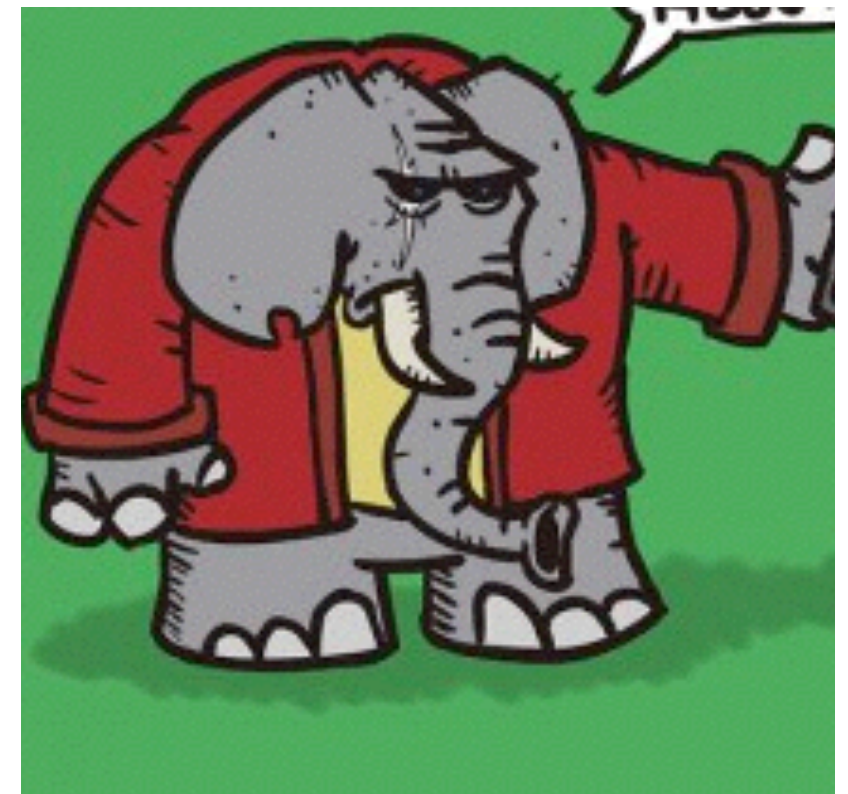
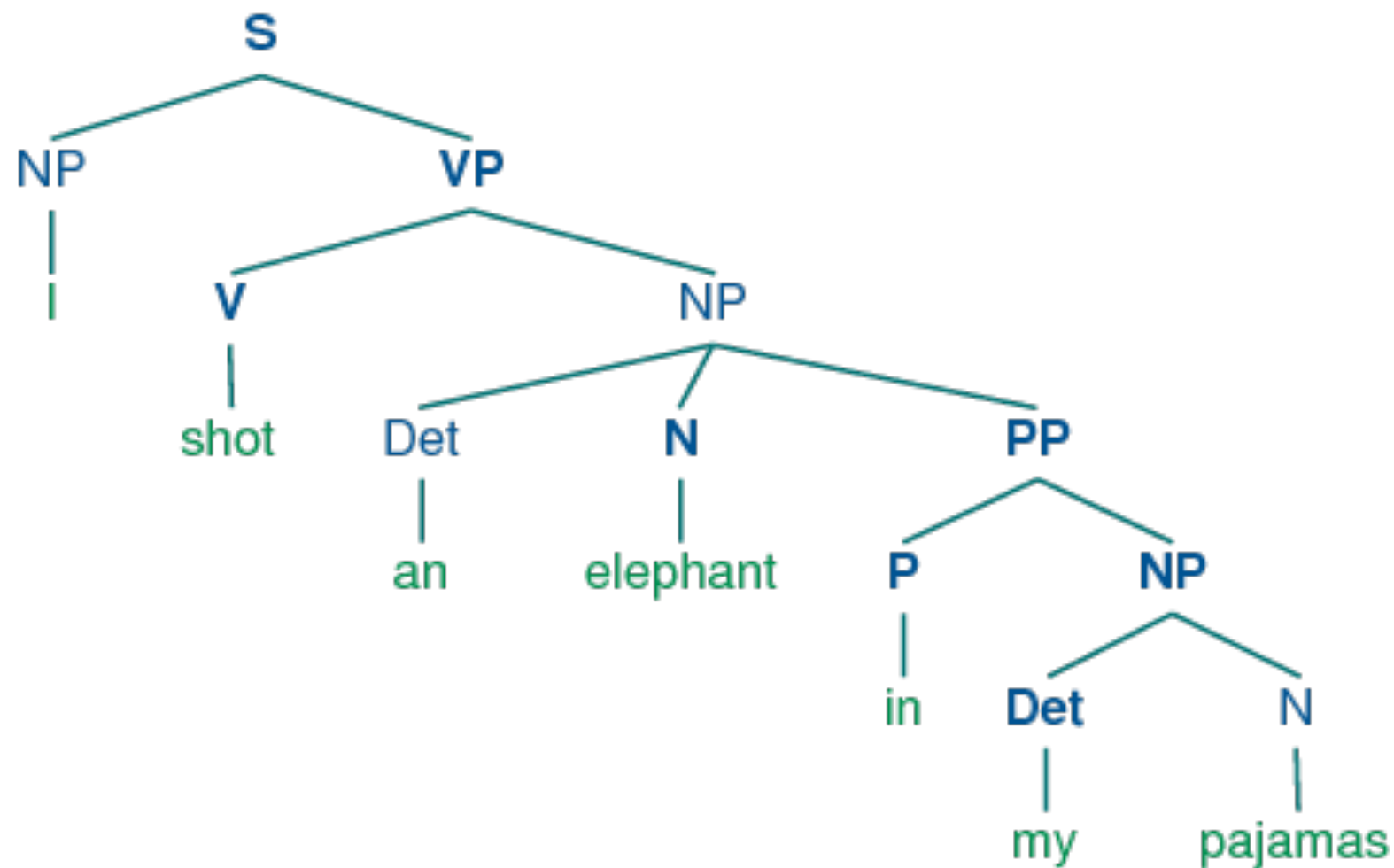
clear evidence that NLP is used in real life!

Grammar



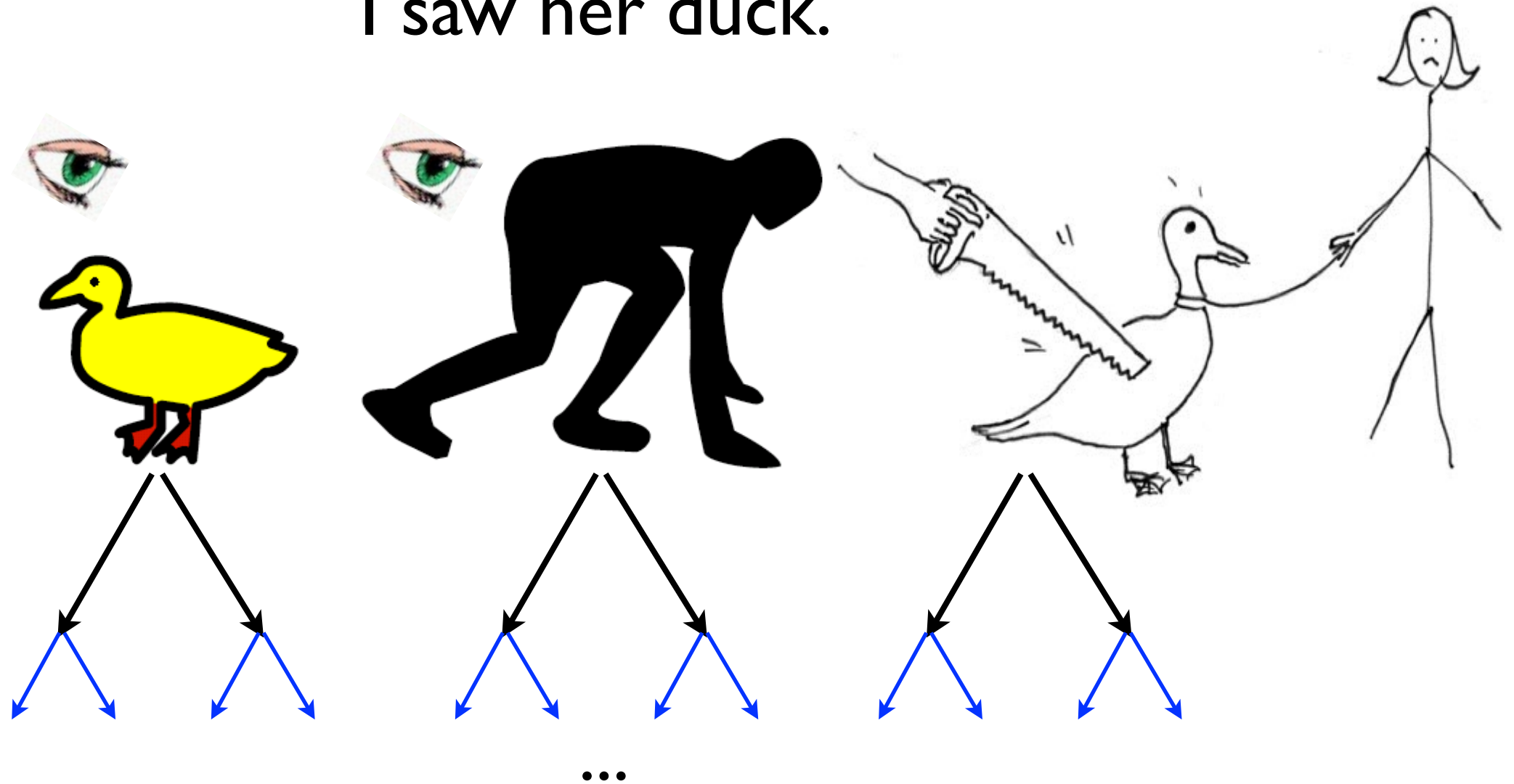
PP Attachment Ambiguity

One morning in Africa,
I shot an elephant in my pajamas;
how he got into my pajamas I'll never know.



Ambiguity Explosion

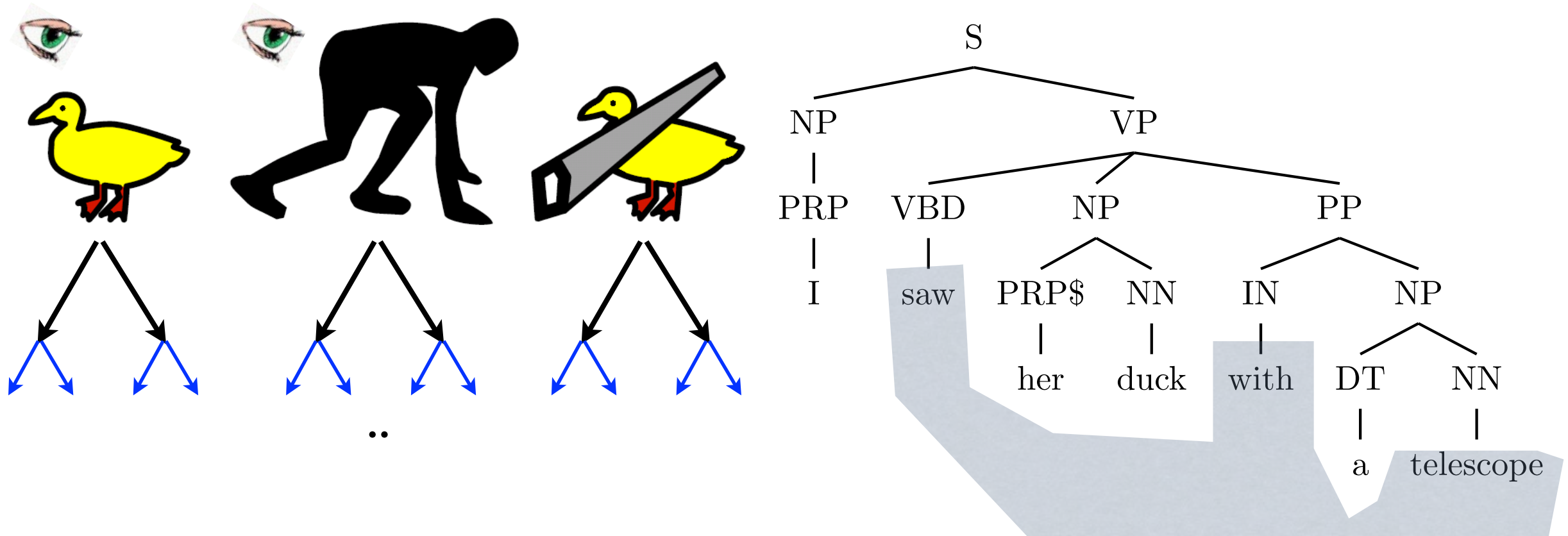
I saw her duck.



- how about...
 - I saw her duck with a telescope.
 - I saw her duck with a telescope in the garden...

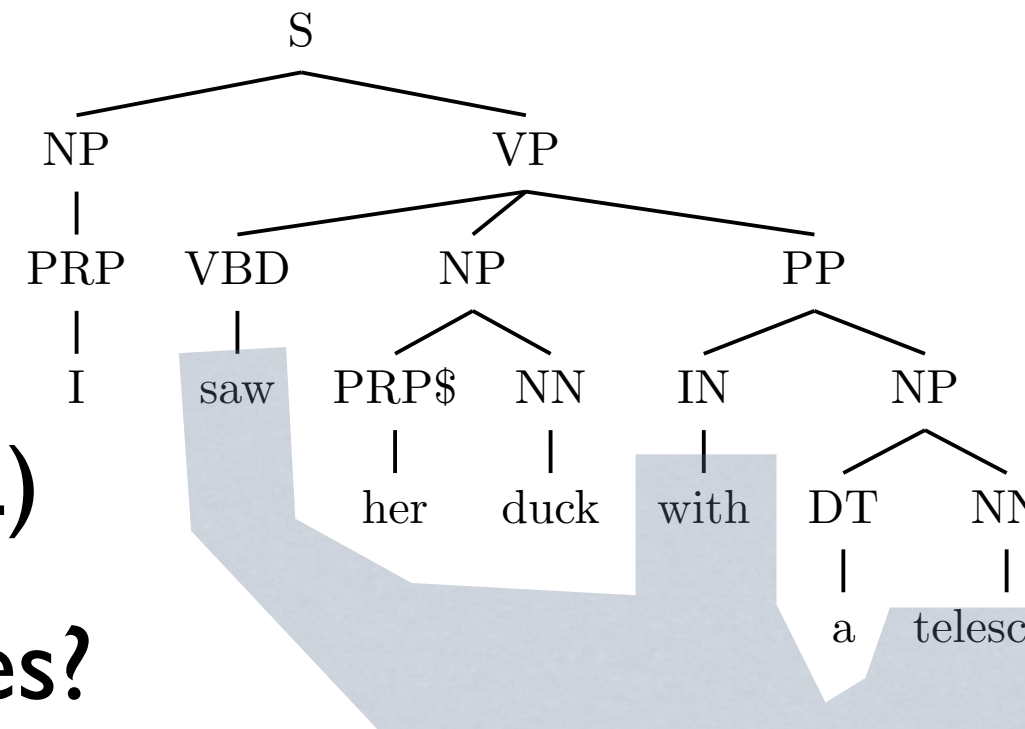
Ambiguity Explosion

- exponential explosion of the search space
- Q1: how to *represent* ambiguities (compactly)?
- Q2: how to *search* over this space (efficiently)?
- Q3: how to *rank* different hypotheses?



Answers...

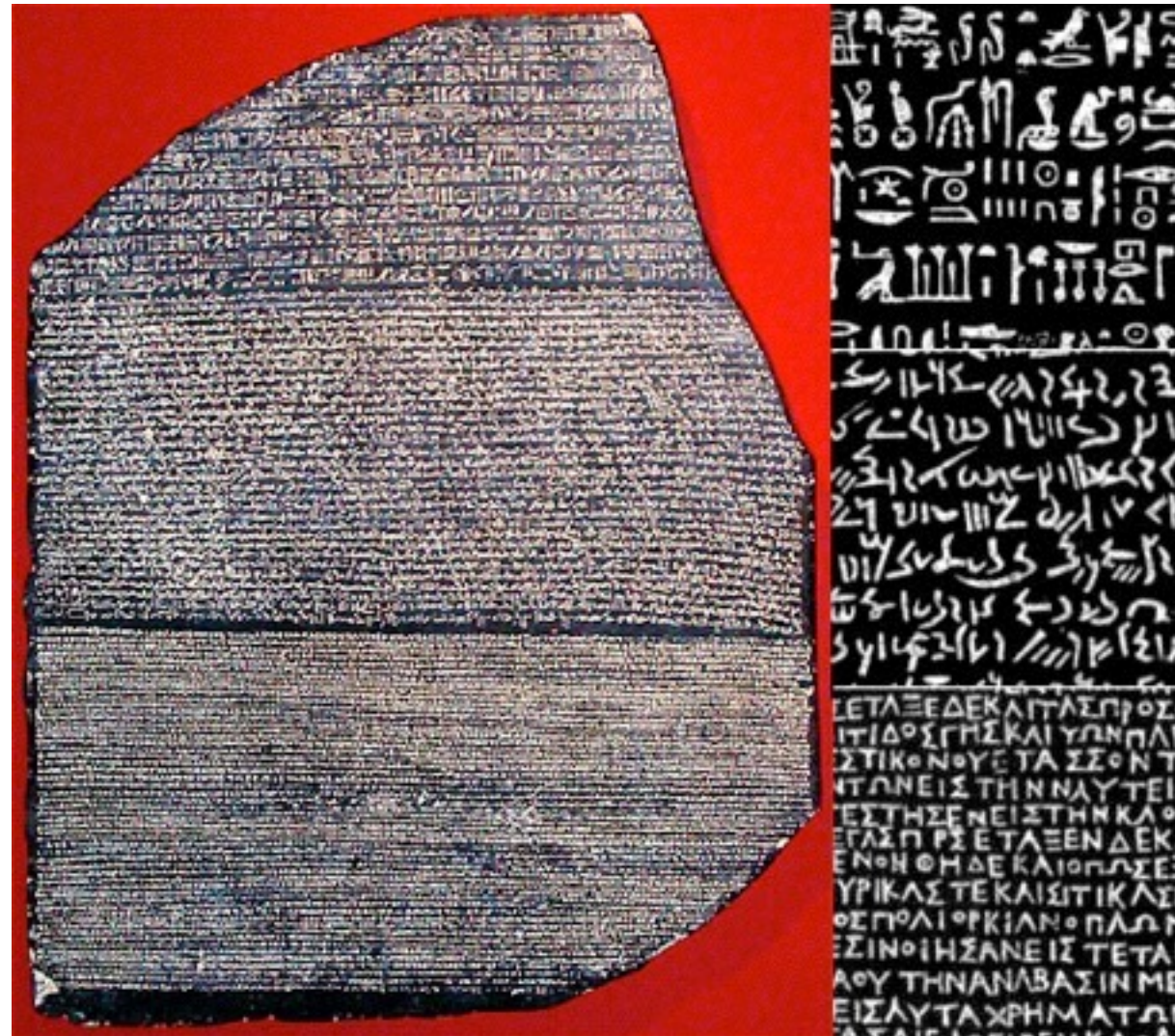
- Q1: how to *represent* ambiguities?
 - context-free grammar (unit 2)
 - finite-state automata (unit 1)
- Q2: how to *search* in this space?
 - dynamic programming (units 1&2)
- Q3: how to *rank* these hypotheses?
 - weighted grammar (units 1-3)
 - weights *learned* from data
 - (saw, with, telescope) seen more often in texts



Why Learning?

- learning is better than hand-written rules, because:
 - less work; easily adapts to new languages/domains
 - Powerset (now bing.com): 15 years for English grammar!
 - now they are writing their Chinese grammar...
 - and languages constantly change!
 - learning *can* work, and often works better!
 - machine translation: used to be dominated by rule-based
 - now statistical methods are better: google vs. systran
 - google learns from the web, and translates 40+ langs

Example - Rosetta Stone



- the most famous (tri-)parallel text
- machines can do the same job! (if given parallel text)
- UN/EU/Ca proceedings, News, tech manuals, ...

A sci-fi example

(Knight, 1997)

Your assignment: translate this Centauri
sentence into Arcturan

farok crrrok hihok yorok klok kantok ok-yurp

farok crrrok hihok yorok klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghirok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghirok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

farok crrrok hihok yorok klok kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghrok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghrok klok .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

farok **crrrok** hihok yorok klok kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok klok .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

farok crrrok **hihok** yorok klok kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok klok .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

farok crrrok hihok **yorok** klok kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok klok .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

farok crrrok hihok yorok **clock** kantok ok-yurp

<p>1c. ok-voon ororok sprok . 1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok . 3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok clock . 10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok . 5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .</p>

farok **crrrok** hihok yorok klok kantok ok-yurp

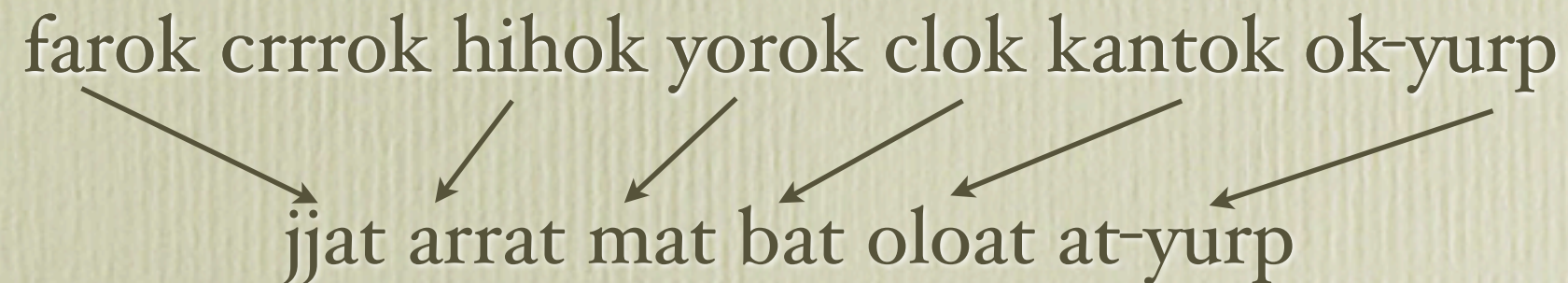
<p>1c. ok-voon ororok sprok . 1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok . 3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok klok . 10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok . 5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .</p>

A sci-fi example

(Knight, 1997)

Your assignment: translate this Centauri sentence into Arcturan

farok crrrok hihok yorok klok kantok ok-yurp
jjat arrat mat bat oloat at-yurp



Are these Arcturan words in Arcturan order?

Clients do not sell pharmaceuticals in Europe .

<p>1e. Garcia and associates . 1s. Garcia y asociados .</p>	<p>7e. the clients and the associates are enemies . 7s. los clients y los asociados son enemigos .</p>
<p>2e. Carlos Garcia has three associates . 2s. Carlos Garcia tiene tres asociados .</p>	<p>8e. the company has three groups . 8s. la empresa tiene tres grupos .</p>
<p>3e. his associates are not strong . 3s. sus asociados no son fuertes .</p>	<p>9e. its groups are in Europe . 9s. sus grupos estan en Europa .</p>
<p>4e. Garcia has a company also . 4s. Garcia tambien tiene una empresa .</p>	<p>10e. the modern groups sell strong pharmaceuticals . 10s. los grupos modernos venden medicinas fuertes .</p>
<p>5e. its clients are angry . 5s. sus clientes estan enfadados .</p>	<p>11e. the groups do not sell zenzanine . 11s. los grupos no venden zanzanina .</p>
<p>6e. the associates are also angry . 6s. los asociados tambien estan enfadados .</p>	<p>12e. the small groups are not modern . 12s. los grupos pequenos no son modernos .</p>

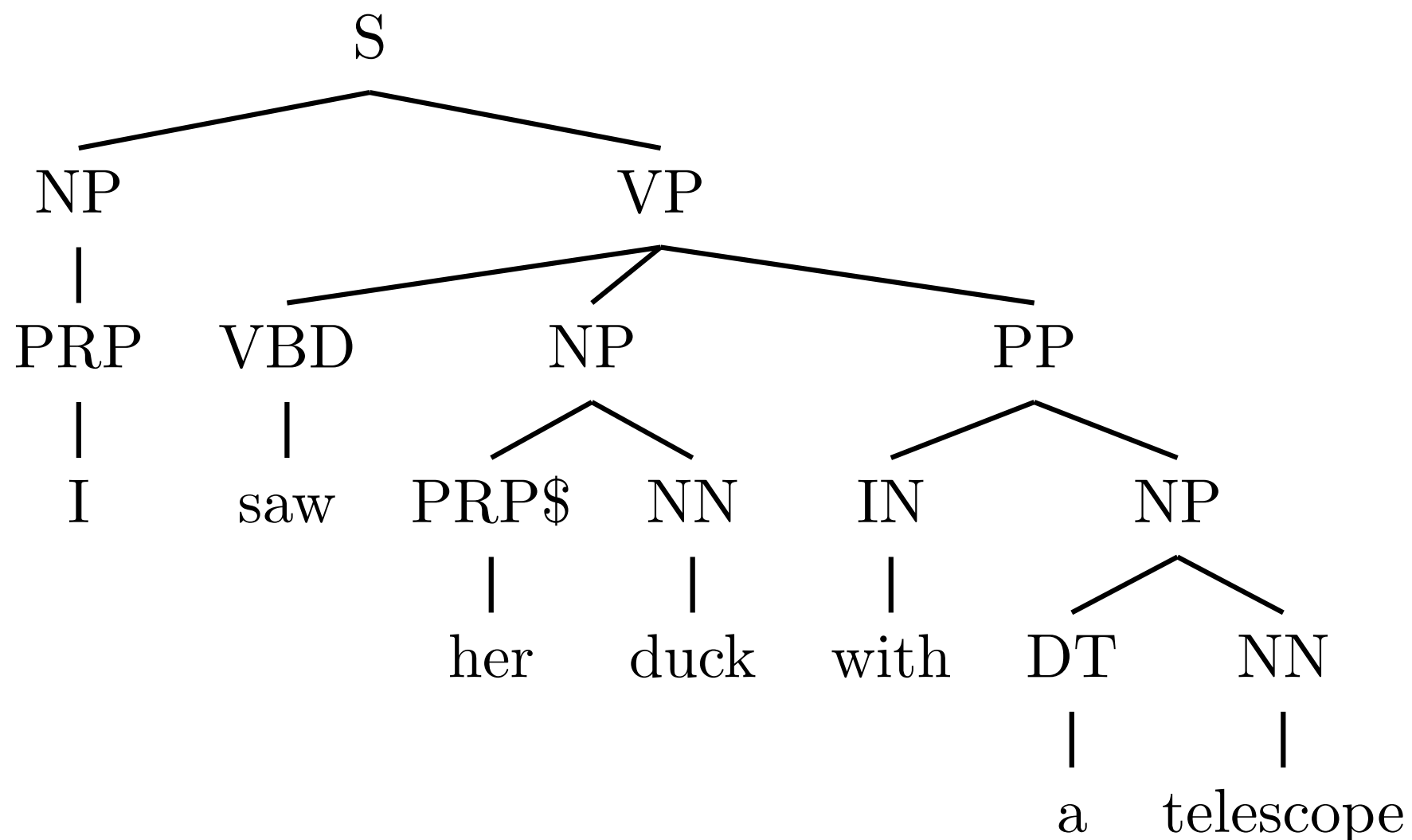
Take Home Message

- languages are *beyond* just bags of words!
 - ambiguity is everywhere, and NLP is all about that
- we'll teach machines how to read and translate...
 - and how to *learn* to read and translate from data
- have fun in this class! :)



Basic Linguistic Structures

- parse tree; grammar rules like $S \rightarrow NP VP$; $NP \rightarrow PRP$
- nonterminals like S, NP, VP, ...
- preterminals (part-of-speech tags): PRP, VBD, IN



Part-of-Speech Tags

- Penn Treebank Part-of-Speech Tags

CC	the cat and the dog	PRPS	his mat
CD	three cats	RB	The very fat cat
DT	the cat	RBR	The more corpulent cat
EX	there was a cat on the mat	RBS	The most corpulent cat
FW	le chat	RP	The cat sat down
IN	on the mat	SYM	50% of cats sit on mats.
JJ	the fat cat	TO	The cat walked to the mat
JJR	the fatter cat	UH	Hey , the cat sat on the mat
JJS	the fattest cat	VB	The cat wants to sit on the mat
LS	a. the cat	VBD	The cat sat on the mat
MD	the cat might sit on the mat	VBG	The cat is sitting on the mat
NN	the cat	VBN	The cat has sat on the mat
NNS	the cats	VBP	I sit on the mat sometimes
NNP	Felix sat on the mat.	VBZ	The cat sits on the mat
NNPS	Were there any mats in <i>Cats</i> ?	WDT	Which mat does the cat sit on?
PDT	all the cats	WP	Who sits on the mat?
POS	the cat's mat	WPS	Whose mat does he sit on?
PRP	he sat on the mat	WRB	Where does the cat sit?

Nonterminal Labels

ADJP	the very fat cat
ADVP	The cat sat on the mat very happily .
CONJP	The cat as well as the dog sat on the mat.
FRAG	Who sat on the mat? The cat .
INTJ	Oh no , the cat sat on the mat!
LST	The following sat on the mat: a. the cat b. the dog
NAC/NML	Secretary of Defense Dick Cheney sat on the mat.
NP	the fat cat
NX	the fat cat and skinny dog
PP	on the mat
PRN	The cat (not the dog) sat on the mat.
PRT	The cat sat down .
QP	one hundred and one cats sat on the mat
RRC	the mat the cat sat on
S	The cat sat on the mat.
SBAR	The dog wondered why the cat sat on the mat .
SBARQ	Did the cat sit on the mat?
SINV	There sat the cat on the mat.
SQ	Did the cat sit on the mat?
UCP	the fat and growing cat
VP	The cat sat on the mat .
WHADJP	How fat was the cat?
WHADVP	How long did the cat sit?
WHNP	Which mat did the cat sit on?
WHPP	On which mat did the cat sit?
X	The fatter the cat, the bigger the mat.