Introduction to Probability Theory 1

Clayton Greenberg

CoLi, CS, MMCI, LSV, CRC 1102 (IDeaL) B4

October 22, 2014

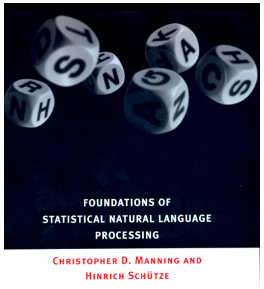
Key concepts

- rules of probability
- exponents
- logarithms
- surprisal
- chain rule
- Bayes' rule
- random variables
- expectation
- variance
- entropy
- mutual information
- relative entropy
- machine learning tasks
- supervision
- normal distributions
- linear regression

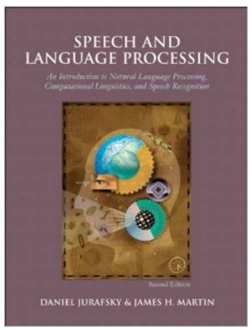
Schedule

22.10.2014	Calculate the probability of a given parse
23.10.2014	Solve the medical test Bayes' Rule problem
27.10.2014	Create a code for simplified Polynesian
29.10.2014	Identify types of machine learning problems
31.10.2014	Find a regression line for 2D data

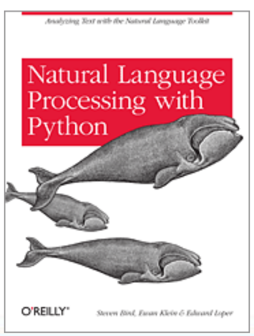
Textbook recommendations



Christopher D. Manning and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.



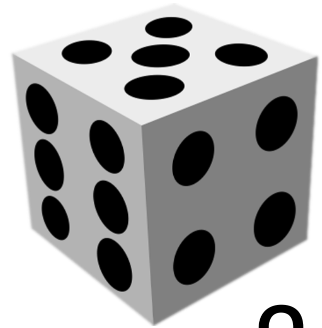
Dan Jurafsky and James H. Martin. *Speech & language processing*. 2nd edition. Prentice Hall, 2008.



Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. O'Reilly Media, Inc., 2009.

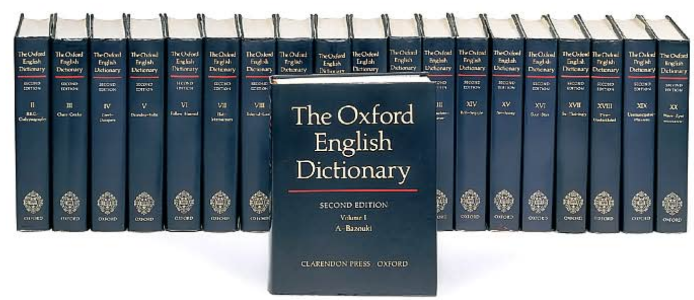
Probabilistic outcomes

Ω = {H, T}



Ω = {1, 2, 3, 4, 5, 6}

Ω = Z*



Ω = Vocabulary

Probabilistic events

- An event **A** is a set of outcomes.
- **A** has "occurred" or "taken place" if one of its member outcomes is observed.
- **Ω** is the certain event.
- ∅ is the impossible event.
- There are $2^{|\Omega|}$ events for a $|\Omega|$ outcome process.

Probabilistic events example

Process: roll a *fair*, three-sided die

$\mathbf{\Omega} = \{1, 2, 3\}$

Events = $\mathcal{P}(\mathbf{\Omega})$ =

$\{ \varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{1,3\}, \mathbf{\Omega} \}$

Event **A**: "roll a 2": $\{2\}$

Event **B**: "at least 2": $\{2,3\}$

Event **C**: "not a 2": $\{1,3\}$

Three definitions of probability

Formal: $p : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ such that

$$(1)\ p(\mathbf{A}) \in [0,1] \text{ for all } \mathbf{A} \in \mathcal{P}(\Omega)$$

$$(2)\ p(\Omega) = 1$$

Simple case: p(**A**) = |**A**| / |**Ω**|

Informal: probability = what you want / what is possible

Probability is a property of events.

Experimental values

To experimentally estimate probability:

1. Run the process many times, T.
2. Count how many times the event **A** occurs, N.
3. p(**A**) ≈ N / T = p̂(**A**)

Suppose you flip a coin 1000 times and get heads 651 times.
Then, p̂(H) = 0.651.
If the coin is fair, p = 0.5.

Axioms of probability

1. probabilities are non-negative real numbers
2. $p(\mathbf{\Omega}) = 1$
3. $\mathbf{A} \cap \mathbf{B} = \varnothing$ implies $p(\mathbf{A} \cup \mathbf{B}) = p(\mathbf{A}) + p(\mathbf{B})$

From these you can derive:

- $p(\varnothing) = 0$
- $\mathbf{A} \subseteq \mathbf{B}$ implies that $p(\mathbf{A}) \leq p(\mathbf{B})$

Two events together

Joint probability: p(**A** and **B**) or p(**A**, **B**)

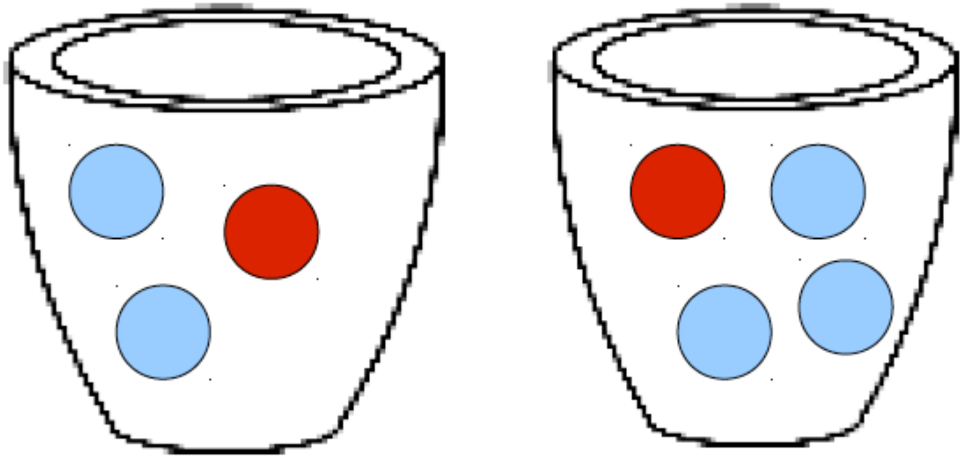
Independent: p(**A**, **B**) = p(**A**)*p(**B**)

P(**A** or **B**) = p(**A**) + p(**B**) – p(**A**, **B**)

mutually exclusive: p(**A**, **B**) = 0

Marginal probability

- p(**A**) = sum of probabilities of mutually exclusive outcomes in **A**.
- In math, $p(\mathbf{A}) = \sum_{\mathbf{B}} p(\mathbf{A}, \mathbf{B})$



Logarithms review

$\log_2(8) = 3$ or $2^3 = 8$

Logarithms are exponents

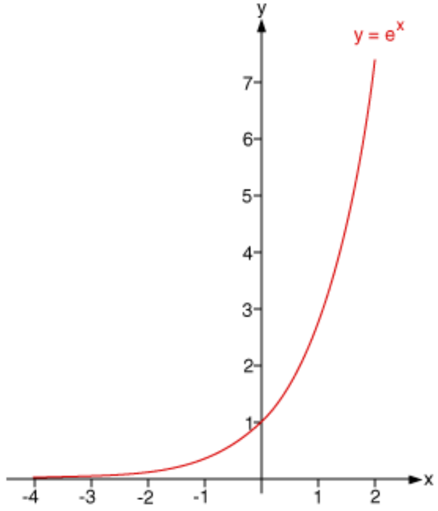
Surprisal(**A**) = $-\log(p(\mathbf{A}))$

usually, the base is 2

Surprisal of heads on a fair coin: $-\log_2(1/2) = 1$ bit

Properties of exponents

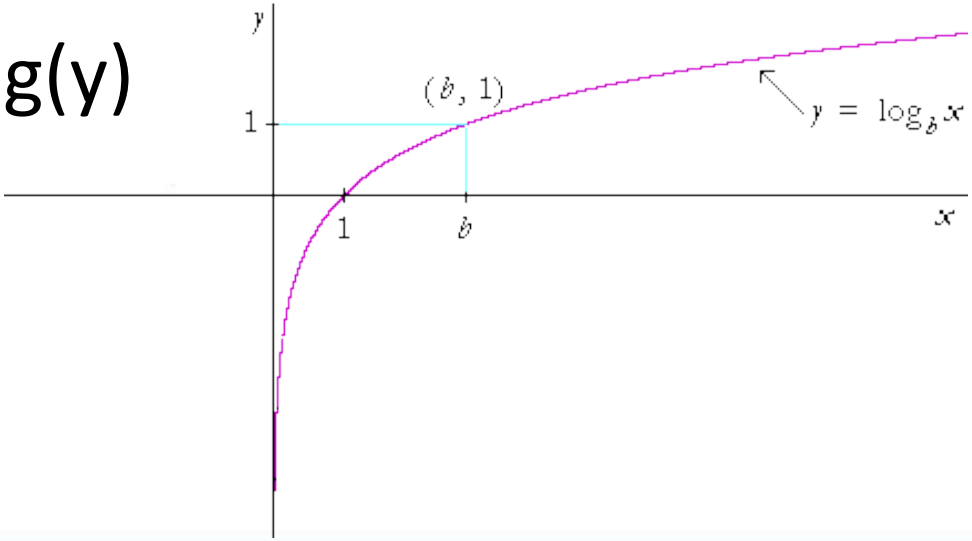
- $x^0 = 1$
- $1^x = 1$
- $0^x = 0$, for all $x \neq 0$
- 0^0 is undefined
- $x^{-y} = 1 / x^y$
- $x^{½} = \sqrt{x}$
- $x^a * x^b = x^{a+b}$
- $x^a / x^b = x^{a-b}$
- $(x^a)^b = x^{a*b}$



Properties of logarithms

- $\log_x(1) = 0$
- $\log(x)$ is undefined,

  for $x \leq 0$
- $\log_y(x) = \log(x)/\log(y)$
- $-\log(x) = \log(1/x)$
- $b^{\log_b(x)} = x$
- $\log(a*b) = \log(a) + \log(b)$
- $\log(a/b) = \log(a) - \log(b)$
- $\log(a^b) = b*\log(a)$



Review of grammar symbols

S → NP VP

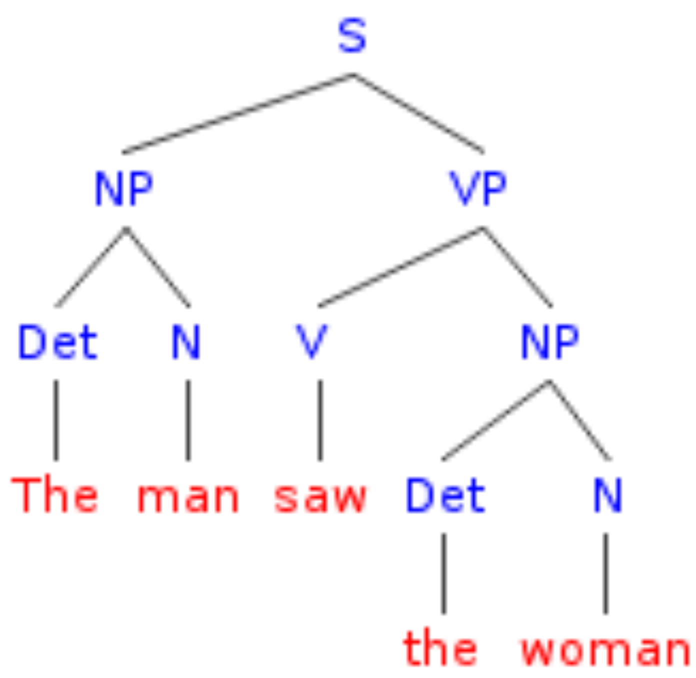
NP → Det N

NP → NP PP

PP → P NP

VP → V NP

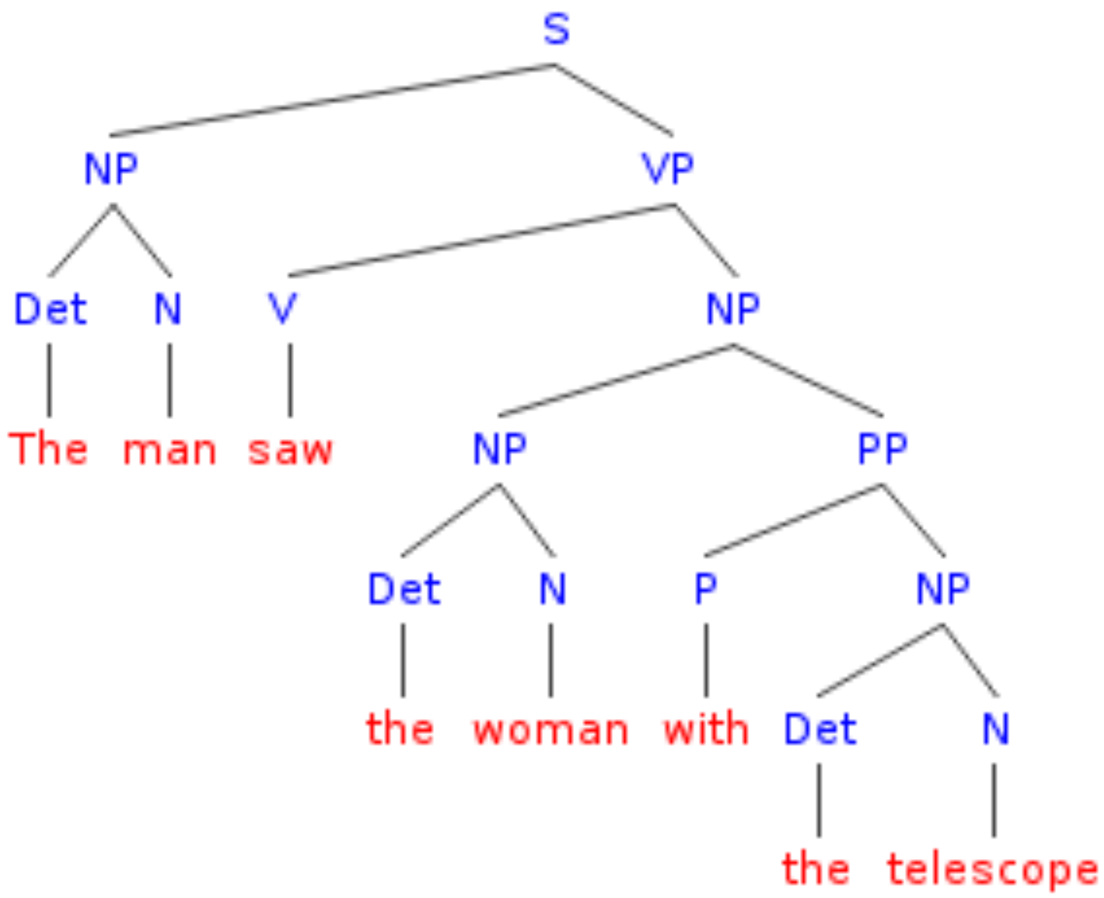
VP → VP PP



Part of speech tag reference

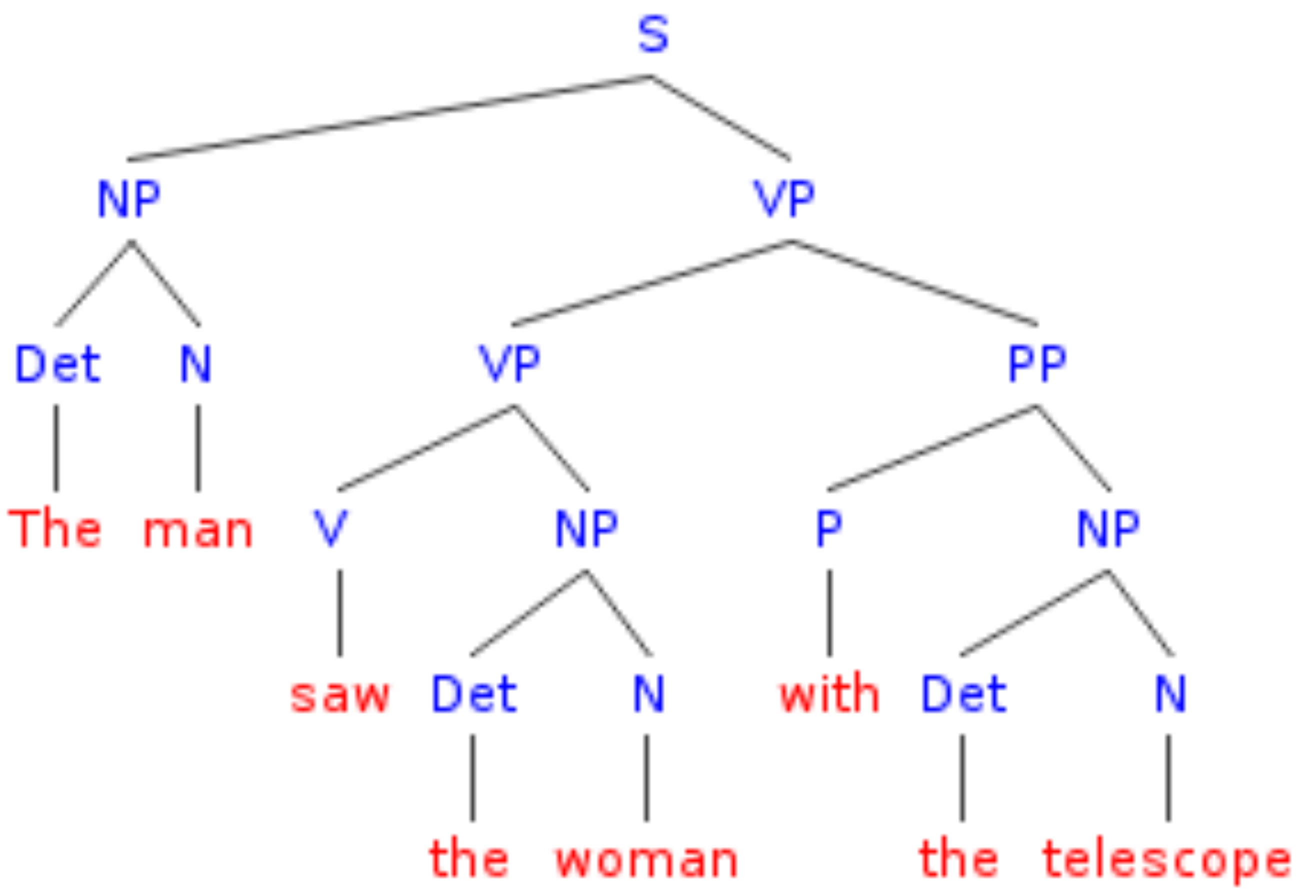
Tag	Description	Example	Tag	Description	Example
CC	coordinating conjunction	*and, but, or*	**SYM**	symbol	+,%, &
CD	cardinal number	*one, two, three*	**TO**	"to"	*to*
DT	determiner	*a, the*	**UH**	interjection	*ah, oops*
EX	existential "there"	*there*	**VB**	verb, base form	*eat*
FW	foreign word	*mea culpa*	**VBD**	verb, preterite (past tense)	*ate*
IN	preposition or subordinating conjunction	*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, snow*	**WP$**	possessive wh-	*whose*
NNS	noun, plural	*llamas*	**WRB**	wh-adverb	*how, where*
NNP	proper noun, singular	*IBM*	**$**	dollar sign	*$*
NNPS	proper noun, plural	*Carolinas*	**#**	pound sign	*#*
PDT	predeterminer	*all, both*	**"**	left quote	' or "
POS	possessive ending	*'s*	**"**	right quote	' or "
PRP	personal pronoun	*I, you, he*	**(**	left parenthesis	[, (, {, <
PRP$	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*			

Structural ambiguity 1



UNIVERSITÄT DES SAARLANDES

Structural ambiguity 2



Derivation

S → NP VP

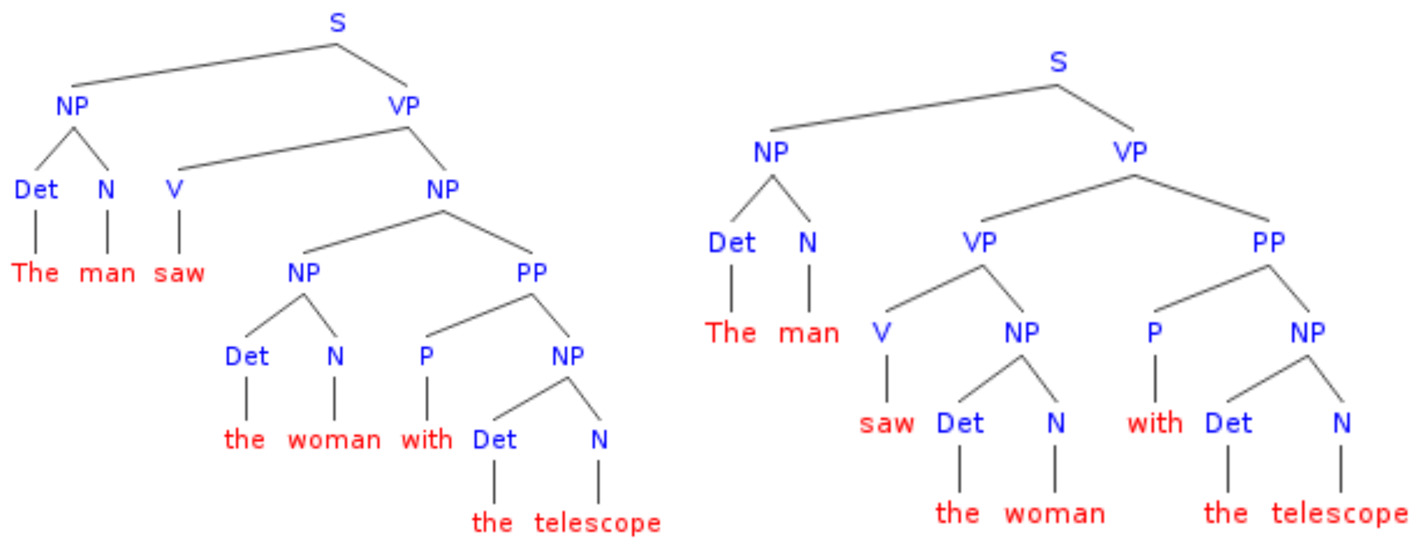
NP → Det N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



Probability of grammaticality

S → NP VP (1.0)	S → NP VP (1.0)
NP → Det N (0.8)	NP → Det N (0.8)
VP → V NP (0.7)	VP → VP PP (0.3)
NP → NP PP (0.2)	VP → V NP (0.7)
NP → Det N (0.8)	NP → Det N (0.8)
PP → P NP (1.0)	PP → P NP (1.0)
Product: 0.0896	Product: 0.1344

A classic sentence

The letter was delivered.

The letter written to John was delivered.

The letter sent to John was delivered.

The letter sent to John fell on the floor.

The letter sent to John fell.

The horse raced past the barn fell.

A simple (wrong) grammar

S → NP VP (1.0)

NP → Det N' (0.8)

NP → NP PP (0.2)

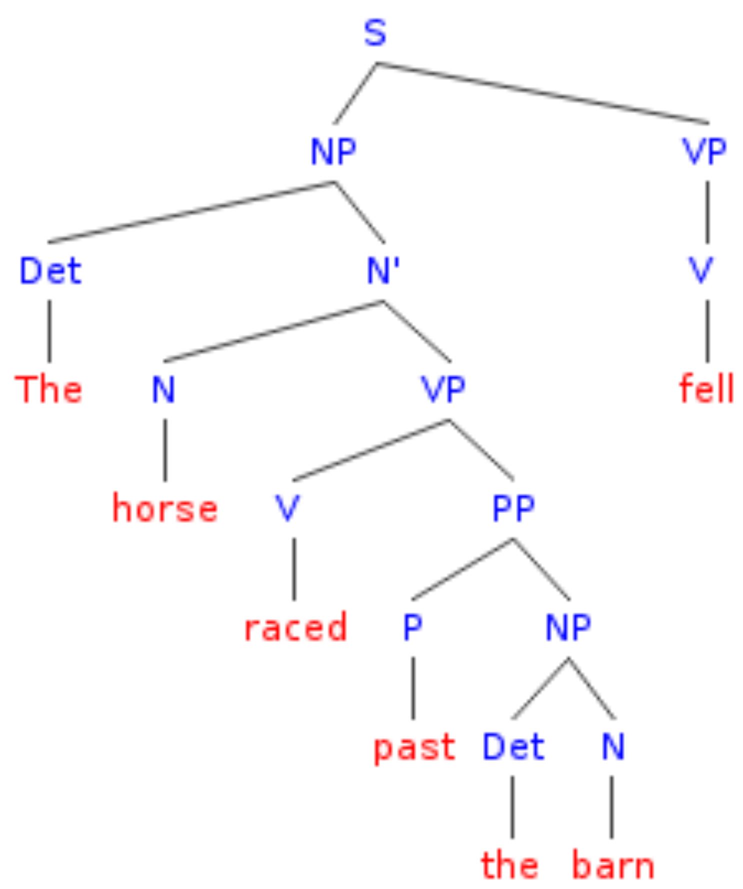
N' → N (0.9)

N' → N VP (0.1)

PP → P NP (1.0)

VP → VP PP (0.4)

VP → V (0.6)



Exercises

1. Memorize:
    1. probability = what you want / what is possible
    2. “and” = * (times) [if independent]
    3. “or” = + (plus) [if mutually exclusive]
    4. logarithms = exponents
    5. surprisal = the negative logarithm of probability

2. Calculate the probability of the parse on slide 23:
   “The horse raced past the barn fell.”