Part-of-Speech Tagging

based on Jimmy Lin
The iSchool
University of Maryland

Source: Calvin and Hobbs

Outline

- What are parts of speech (POS)?
- What is POS tagging?
- Methods for automatic POS tagging
  - Rule-based POS tagging
  - Transformation-based learning for POS tagging
- Along the way…
  - Evaluation
  - Supervised machine learning

Parts of Speech

- "Equivalence class" of linguistic entities
  - "Categories" or "types" of words
- Study dates back to the ancient Greeks
  - Dionysius Thrax of Alexandria (*c.* 100 BC)
  - 8 parts of speech: noun, verb, pronoun, preposition, adverb, conjunction, participle, article
  - الاسم، الفعل، الحرف
  - اسماء الاشارة الافعال الناقصة الاسماء الخمسة ان واخواتها
  - Remarkably enduring list!

How do we define POS?

- By meaning
  - Verbs are actions
  - Adjectives are properties
  - Nouns are things

Unreliable! Think back to the comic!

- By the syntactic environment
  - What occurs nearby (in the sentence?)?
  - What does it act as?
- By what morphological processes affect it
  - What affixes does it take (un-, -able, -tion, لم -س ،-ات ،___ال?
- Combination of the above

Parts of Speech

- Open class
  - Impossible to completely enumerate
  - New words continuously being invented, borrowed, etc.
- Closed class
  - Closed, fixed membership
  - Reasonably easy to enumerate
  - Generally, short function words that “structure” sentences

Open Class POS

- Four major open classes in English
  - Nouns
  - Verbs
  - Adjectives
  - Adverbs
- All languages have nouns and verbs... but may not have the other two

Nouns

- Open class
  - New inventions all the time: muggle, webinar, .
  - كمبيوتر، تويتر، ايفون، جهاز لوحي
- Semantics:
  - Generally, words for people, places, things (entities?)
  - But not always (bandwidth, energy, ...)
- Syntactic environment:
  - Occurring with determiners (the, ال:not all languages, though)
  - Pluralizable, possessivizable: Ali’s, towns, children, phenomena
- Other characteristics:
  - Mass vs. count nouns:

Verbs

- Open class
  - New inventions all the time: google, tweet, ..
  - شير، سيف، غرد، يفرمت.
- Semantics:
  - Generally, denote actions, processes, etc.
- Syntactic environment:
  - Intransitive, transitive, لازم ومتعدي بمفعول أو اكثر، ناقص
  - Alternations
- Other characteristics:
  - Main vs. auxiliary verbsتام او ناقص :
  - Gerunds (verbs behaving like nouns)
  - Participles (verbs behaving like adjectives)

Adjectives and Adverbs

- Adjectives
  - Generally modify nouns, e.g., *tall* girlصفات
- Adverbs
  - Sometimes modify verbs, e.g., sang *beautifully*
  - Sometimes modify adjectives, e.g., *extremely* hot

Closed Class POS

- Prepositions
  - In English, occurring before noun phrases
  - Specifying some type of relation (spatial, temporal, …)
  - Examples: *on* the shelf, *before* noon
  - حروف الجر مثلا
- Particles
  - Resembles a preposition, but used with a verb (“phrasal verbs”)
  - Examples: find *out*, turn *over*, go *on*

Particle vs. Prepositions

He came *by* the office in a hurry (by = preposition)
He came *by* his fortune honestly (by = particle)

We ran *up* the phone bill (up = particle)
We ran *up* the small hill (up = preposition)

He lived *down* the block (down = preposition)
He never lived *down* the nicknames (down = particle)

More Closed Class POS

- Determiners
  - Establish reference for a noun
  - Examples: *a*, *an*, *the* (articles), *that*, *this*, *many*, *such*, …
- Pronouns
  - Refer to person or entities: *he*, *she*, *it*
  - Possessive pronouns: *his*, *her*, *its*
  - Wh-pronouns: *what*, *who*

Note variations between languages: compare with Arabic

Closed Class POS: Conjunctions

- Coordinating conjunctions
  - Join two elements of “equal status”
  - Examples: cats *and* dogs, salad *or* soup
  - Is 'و ، مع، ثم،' a conjunction: role? Is it separate from next word?
- Subordinating conjunctions
  - Join two elements of “unequal status”
  - Examples: We’ll leave *after* you finish eating. *While* I was waiting in line, I saw my friend.
  - Complementizers are a special case: I think *that* you should finish your assignment

Digression

Language Variation

You already know that: Arabic and English:
Do you need definite articles?

Do you separate word parts: word vs sentence:

رايتهم

[Do you use compound names as one word:

بيرزيت، رام الله،

How many tenses: past, present, future, more:

Gender effects: extremes: from no to quite heavy:

Arabic, English, Russian, German,…

Much more!

So be careful when working with Arabic using foreign literature!

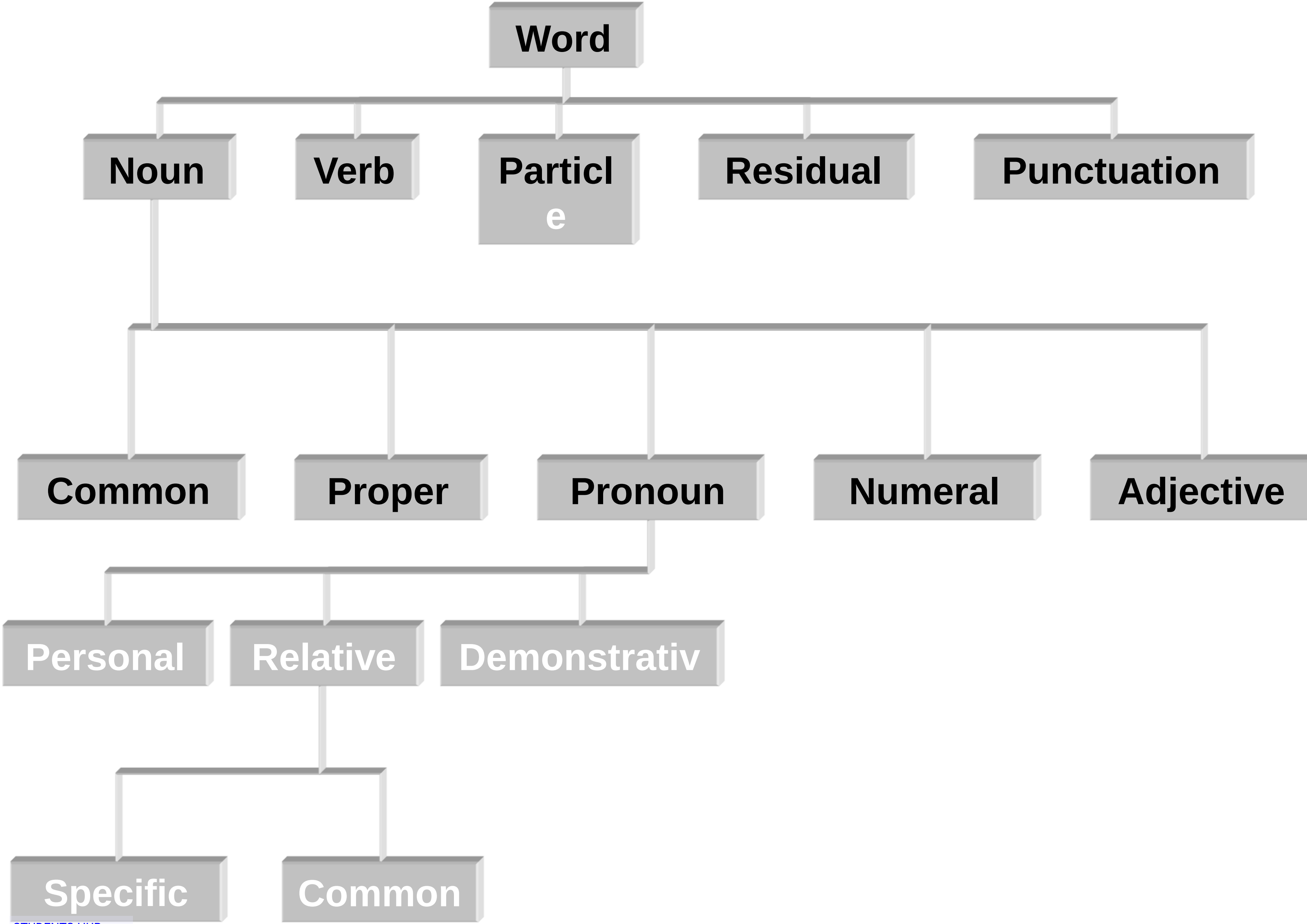
A must verb in sentence (En), not really (Ar)

POS Tagging: What's the task?

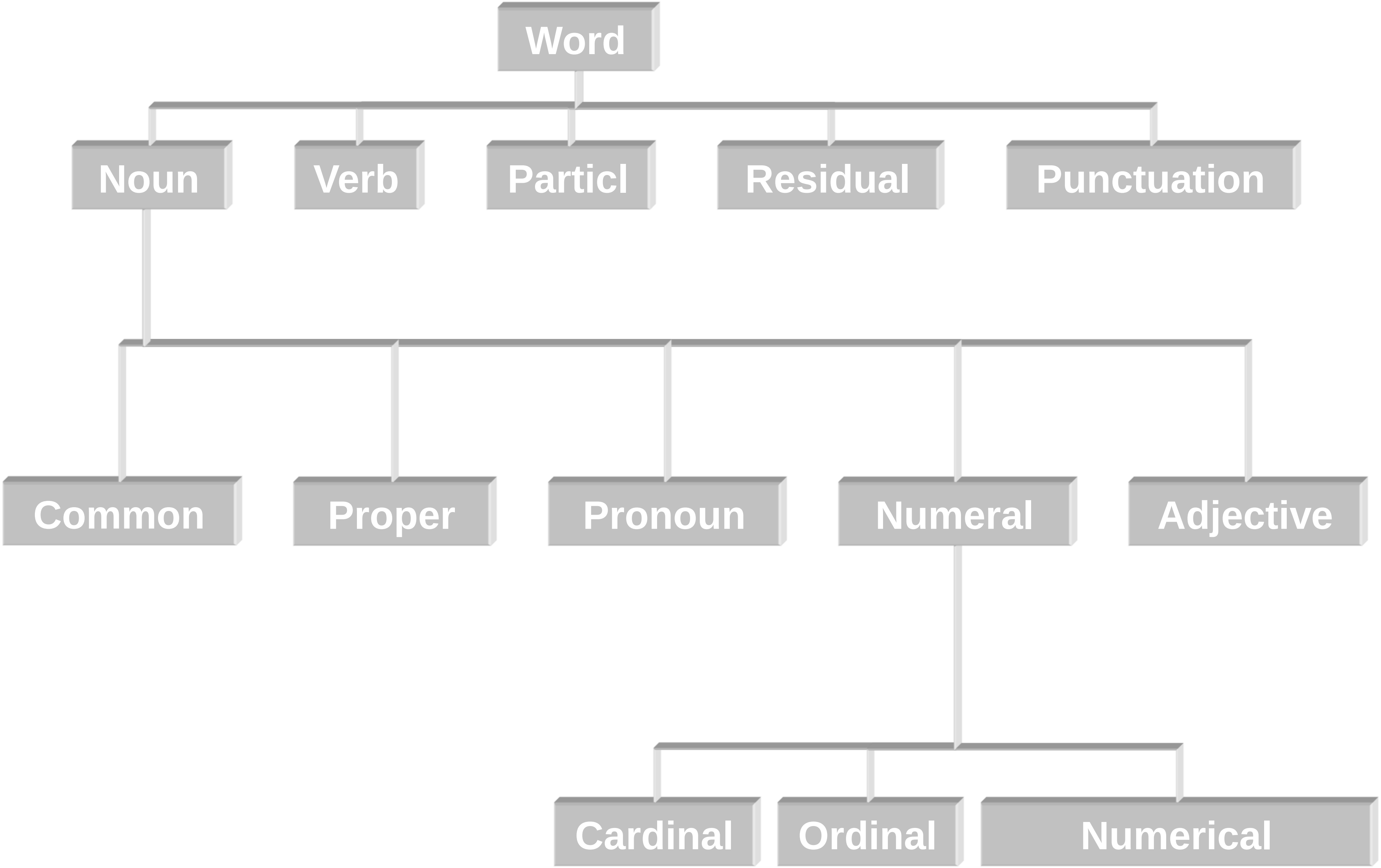
- Process of assigning part-of-speech (POS) **tags** to words
- But what tags are we going to assign?

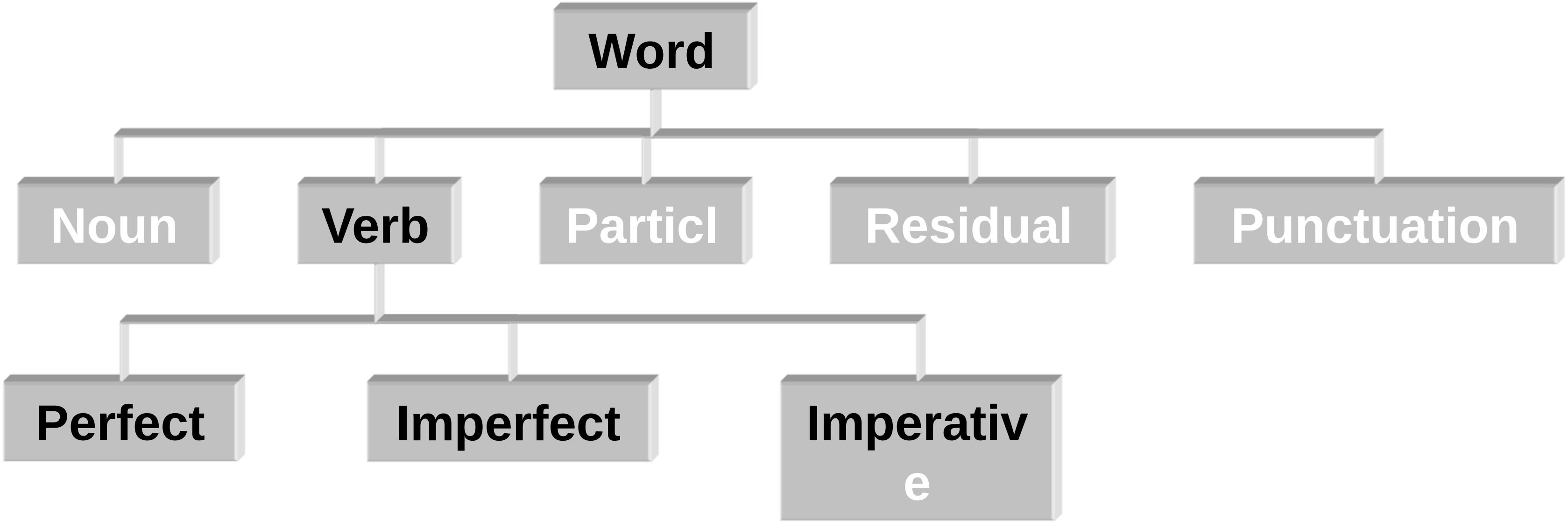
What's the tradeoff?

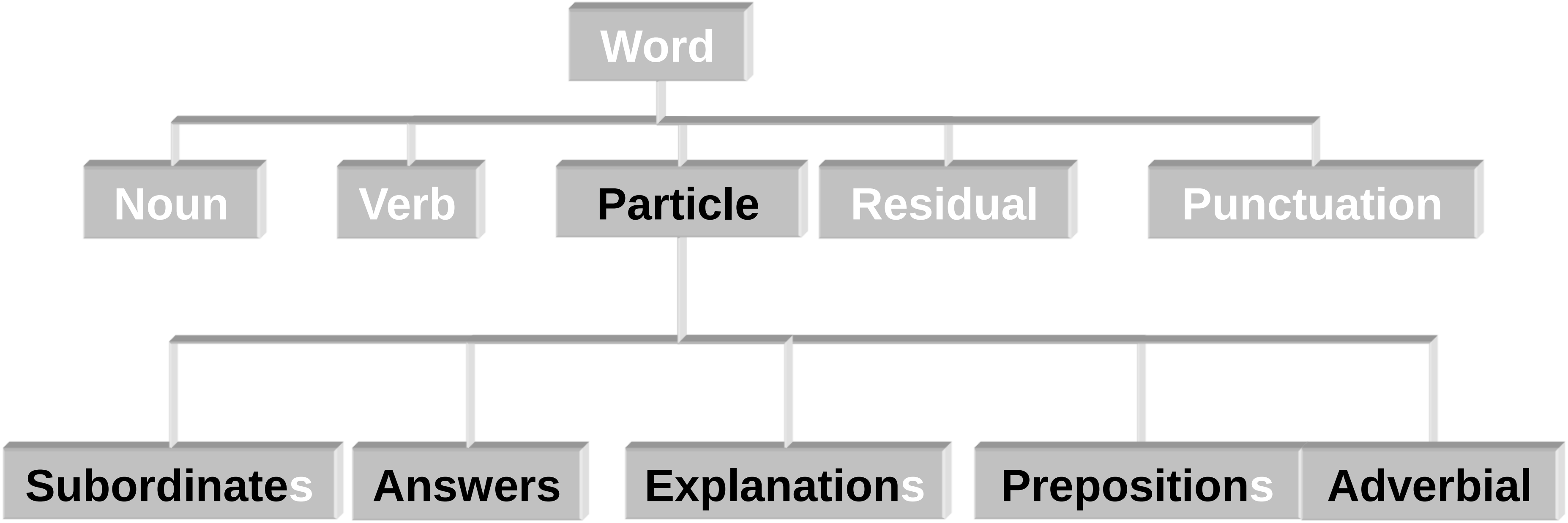
  - Coarse grained: noun, verb, adjective, adverb, … (small list, large content in each) اسم، اسم علم، اسم علم مؤنث، اسم من الخمسة و هكذا
  - Fine grained: {proper, common} noun (Larger list, less content in each)
  - Even finer-grained: {proper, common} noun ± animate
- Important issues to remember
  - Choice of tags encodes certain distinctions/non-distinctions (see coarse vs fine grained just mentioned)
  - Tagsets will differ across languages!
- For English, Penn Treebank is the most common tagset
- What about Arabic: Noun, verb and particle (Sh. Khoja Slides)

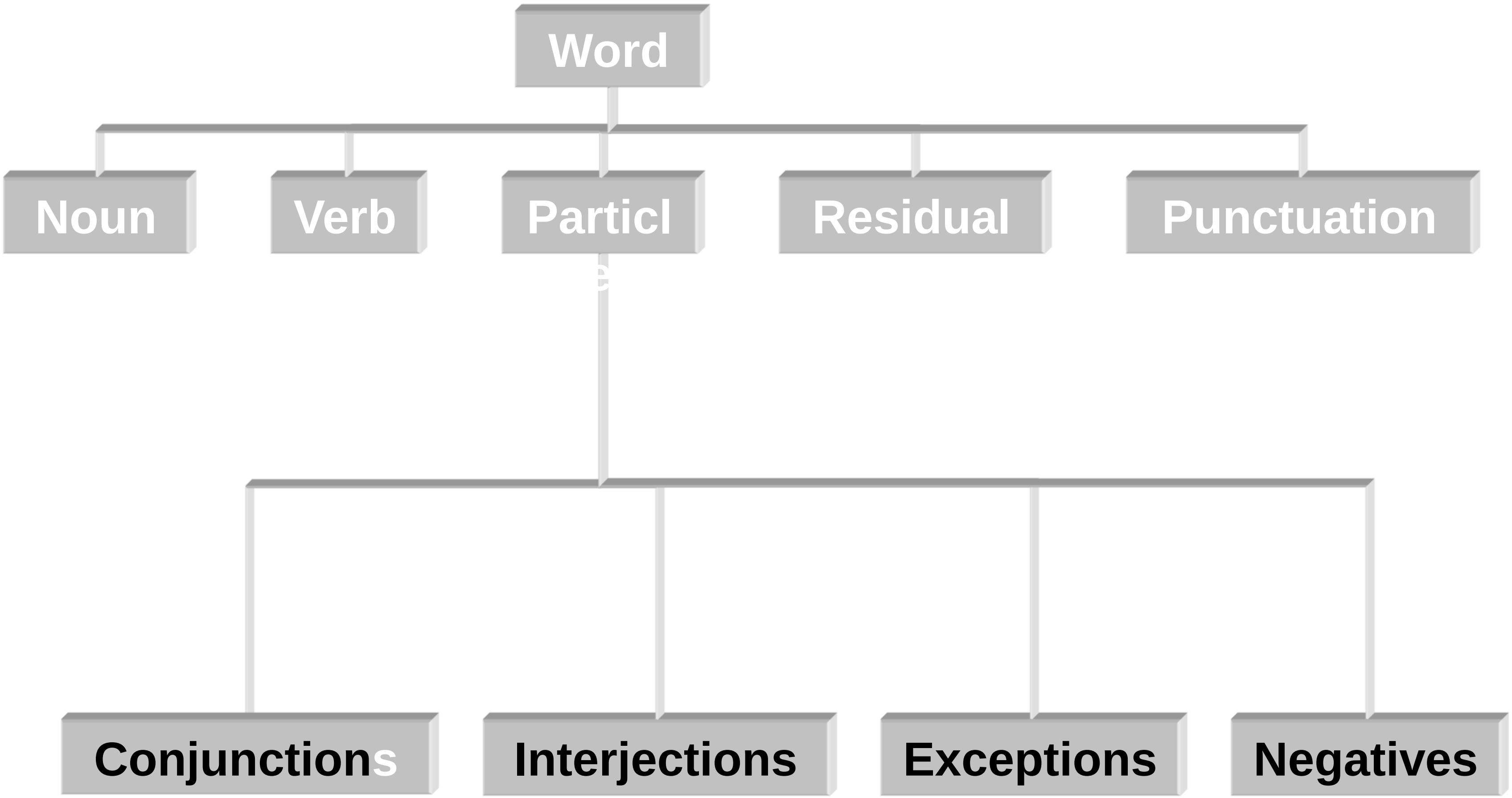


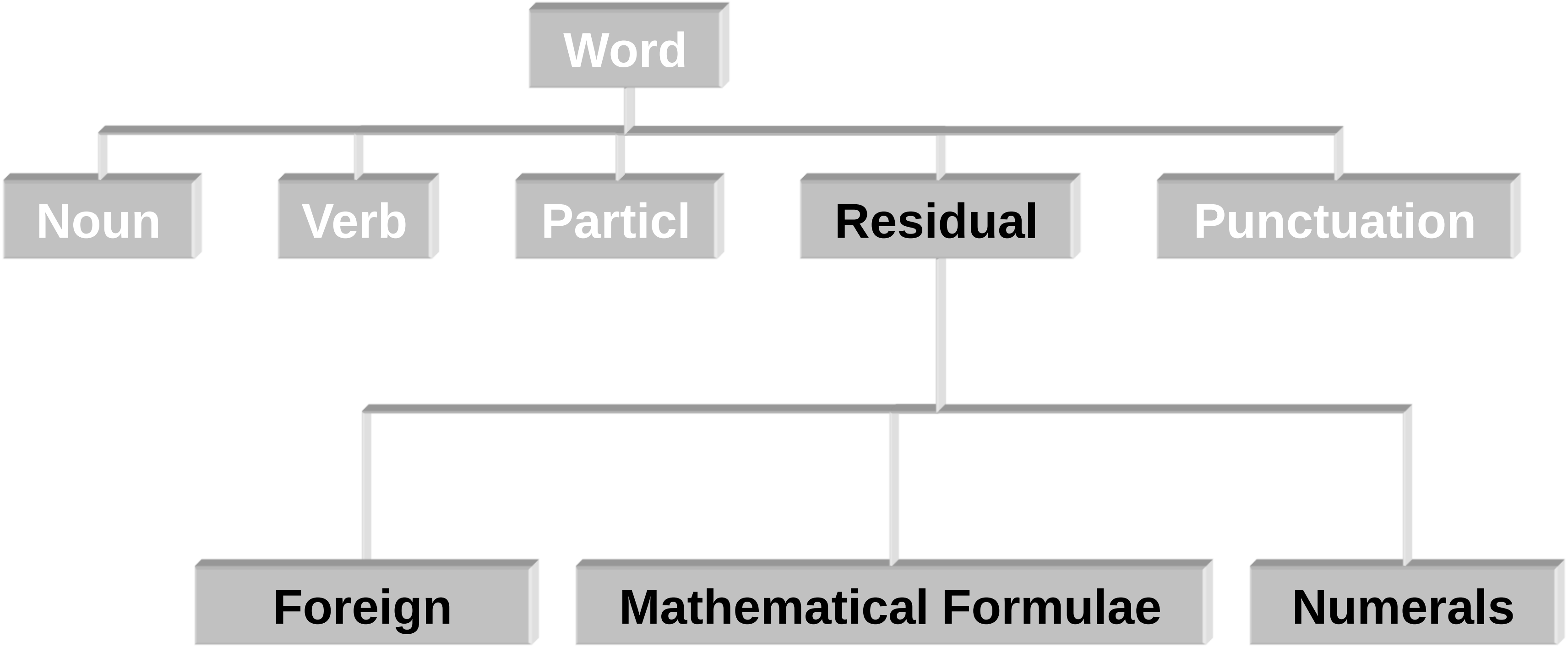
Word
Noun
Verb
Particl
e
Residual
Punctuation
Common
Proper
Pronoun
Numeral
Adjective
Personal
Relative
Demonstrativ
Specific
Common

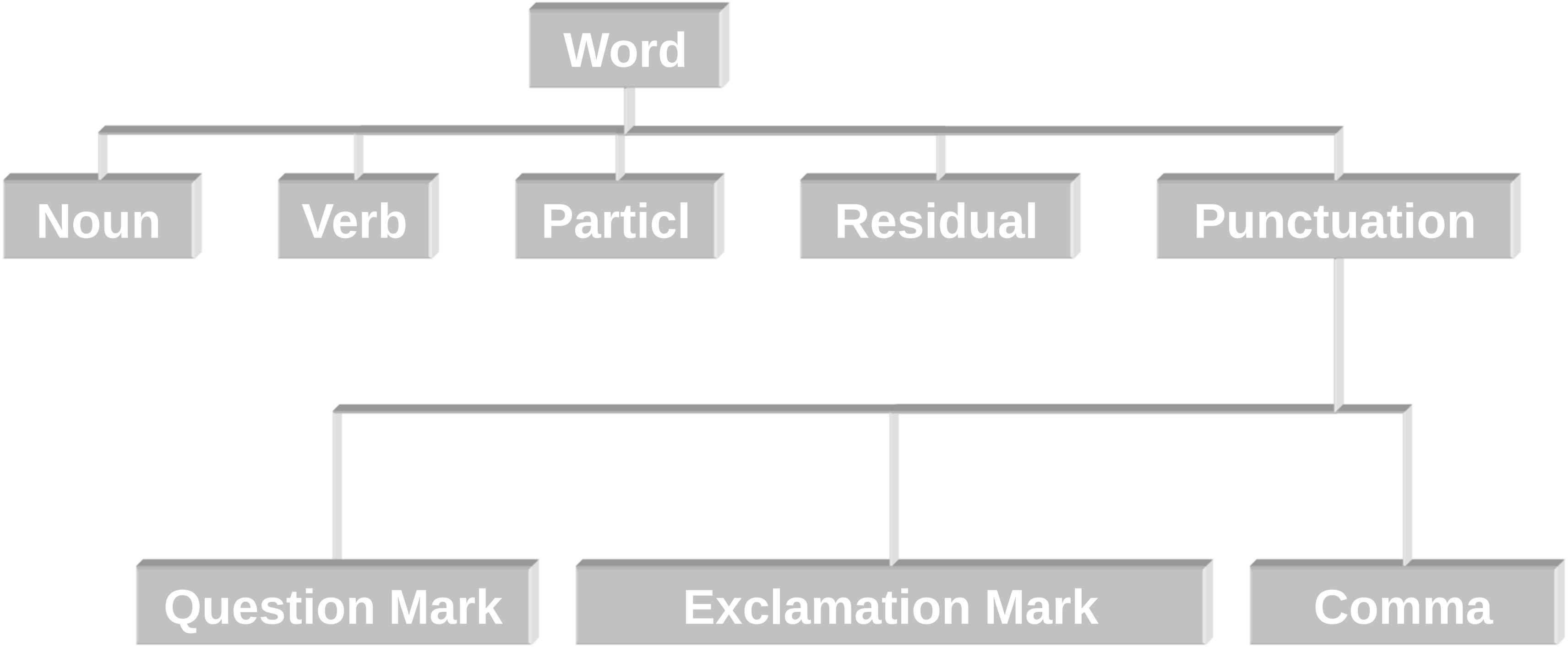

Word
Noun
Verb
Particl
Residual
Punctuation
Common
Proper
Pronoun
Numeral
Adjective
Cardinal
Ordinal
Numerical


Word
Noun
Verb
Particl
Residual
Punctuation
Perfect
Imperfect
Imperativ
e


Word
Noun
Verb
Particle
Residual
Punctuation
Subordinates
Answers
Explanations
Prepositions
Adverbial


Word
Noun
Verb
Particle
Residual
Punctuation
Conjunctions
Interjections
Exceptions
Negatives


Word
Noun
Verb
Particl
Residual
Punctuation
Foreign
Mathematical Formulae
Numerals


Word
Noun
Verb
Particl
Residual
Punctuation
Question Mark
Exclamation Mark
Comma

Penn Treebank Tagset: 45 Tags

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 ”
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*			

Penn Treebank Tagset: Choices

- Example:
  - The/DT grand/JJ jury/NN commmented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS ./.
- Distinctions and non-distinctions
  - Prepositions and subordinating conjunctions are tagged “IN” (“Although/IN I/PRP..”)
  - Except the preposition/complementizer “to” is tagged “TO”

Don’t think this is correct? Doesn’t make sense?

Often, must suspend linguistic intuition and defer to the annotation guidelines!

Why do POS tagging?

- One of the most basic NLP tasks
  - Nicely illustrates principles of statistical NLP
- Useful for higher-level analysis
  - Needed for syntactic analysis
  - Needed for semantic analysis
- Sample applications that require POS tagging
  - Machine translation
  - Information extraction
  - Lots more…

Why is it hard?

- Not only a lexical problem
  - Remember ambiguity?
- Better modeled as sequence labeling problem
  - Need to take into account **context**!

Try tagging...

- The back door
- On my back
- Win the voters back
- Promised to back the bill
- OR
- وضعت الكتاب على الطاولة
- على كانت المتفوقة من أخواتها رغم مرضها المزمن
- على كل منكم حل الوظيفة منفردا
- على الباغي تدور الدوائر
- القت الشرطة القبض على المجرم خلال ساعات

Try your hand at tagging...

- I thought that you...
- That day was nice
- You can go that far

Why is it hard?*

		87-tag Original Brown		45-tag Treebank Brown	
Unambiguous (1 tag)		**44,019**		**38,857**	
Ambiguous (2–7 tags)		**5,490**		**8844**	
Details:	2 tags	4,967		6,731	
	3 tags	411		1621	
	4 tags	91		357	
	5 tags	17		90	
	6 tags	2	(*well*, *beat*)	32	
	7 tags	2	(*still*, *down*)	6	(*well*, *set*, *round*, *open*, *fit*, *down*)
	8 tags			4	(*'s*, *half*, *back*, *a*)
	9 tags			3	(*that*, *more*, *in*)

Part-of-Speech Tagging

- How do you do it automatically?
- How well does it work? **This first**

Evolution of the Evaluation

- Evaluation by **argument**
- Evaluation by **inspection** of examples
- Evaluation by **demonstration**
- Evaluation by **improvised** demonstration
- Evaluation on **data** using a figure of merit
- Evaluation on **test data**
- Evaluation on **common** test data
- Evaluation on common, **unseen** test data

Evaluation Metric

- Binary condition (correct/incorrect):
  - Accuracy
- Set-based metrics (illustrated with document retrieval):

	Relevant	Not relevant
Retrieved	A	B
Not retrieved	C	D

Collection size = A+B+C+D
Relevant = A+C
Retrieved = A+B

  - Precision = A / (A+B)
  - Recall = A / (A+C)
  - Miss = C / (A+C)
  - False alarm (fallout) = B / (B+D)
  - F-measure: $$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$
  - So if we tested 1000 words and 100 out of 150 were tagged correctly as names, 40 were tagged as names which are not: compute the above for this name tagger!

Components of a Proper Evaluation

- Figures(s) of merit
- Baseline
- Upper bound
- Tests of statistical significance

Part-of-Speech Tagging

- How do you do it automatically? ⇐ **Now this**
- How well does it work?

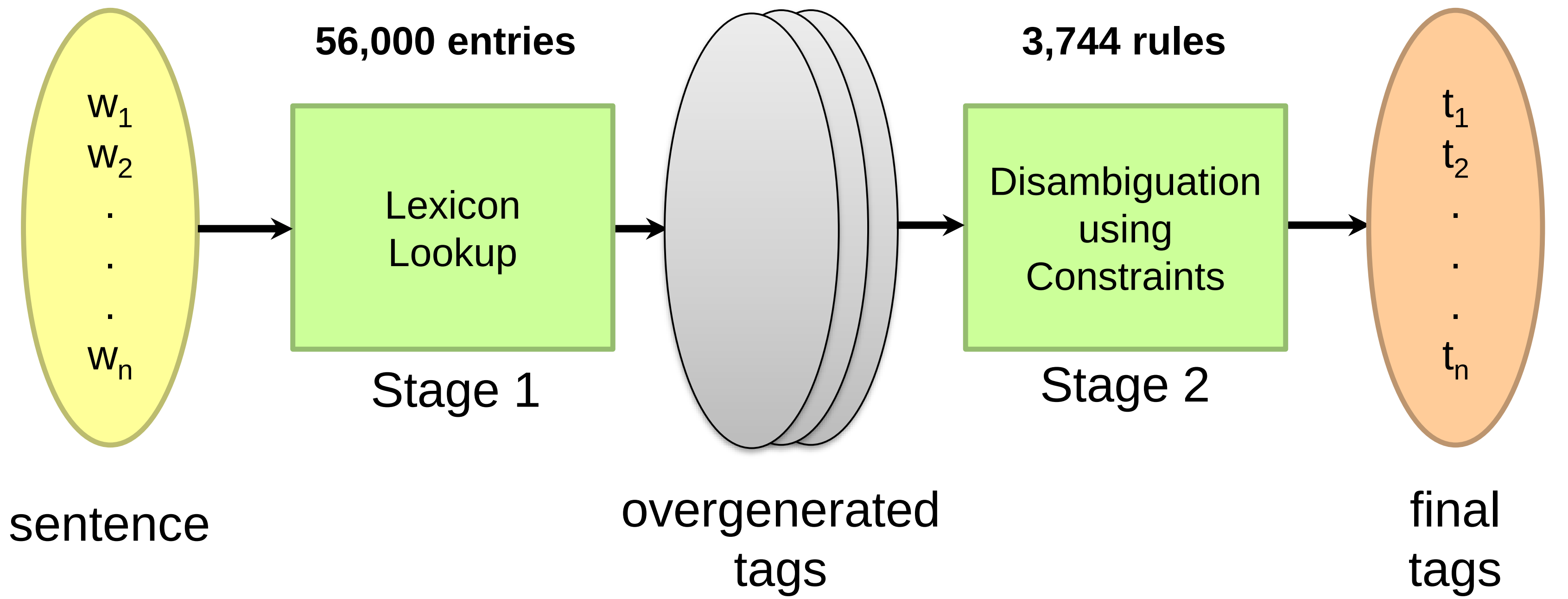
Automatic POS Tagging

- Rule-based POS tagging (now)
- Transformation-based learning for POS tagging (May be later)
- Hidden Markov Models (May be later)
- Maximum Entropy Models (you do it if needed)
- Conditional Random Fields (you do it if needed)

Rule-Based POS Tagging

- Dates back to the 1960's
- Combination of lexicon + hand crafted rules
  - Example: EngCG (English Constraint Grammar)

EngCG Architecture



EngCG: Sample Lexical Entries

Word	POS	Additional POS features
smaller	ADJ	COMPARATIVE
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	PRESENT -SG3 VFIN
show	N	NOMINATIVE SG
shown	PCP2	SVOO SVO SV
occurred	PCP2	SV
occurred	V	PAST VFIN SV

EngCG: Constraint Rule Application

Example Sentence: *Newman had originally practiced that ...*

Newman	**NEWMAN N NOM SG PROPER**
had	**HAVE <SVO> V PAST VFIN**
	HAVE <SVO> PCP2
originally	**ORIGINAL ADV**
practiced	PRACTICE <SVO> <SV> V PAST VFIN
	PRACTICE <SVO> <SV> PCP2
that	~~ADV~~
	PRON DEM SG
	DET CENTRAL DEM SG
	CS

overgenerated tags

```
ADVERBIAL-THAT Rule
Given input: that
if
    (+1 A/ADV/QUANT);
    (+2 SENT-LIM);
    (NOT -1 SVOC/A);
then eliminate non-ADV tags
else eliminate ADV tag
```

disambiguation constraint

I thought that you... (subordinating conjunction)
That day was nice. (determiner)
You can go that far. (adverb)

EngCG: Evaluation

- Accuracy ~96%*
- A lot of effort to write the rules and create the lexicon
  - Try debugging interaction between thousands of rules!
  - Recall discussion from the first lecture?
- Assume we had a corpus *annotated* with POS tags
  - Can we *learn* POS tagging automatically?

Supervised Machine Learning

- Start with annotated corpus
  - Desired input/output behavior
- Training phase:
  - Represent the training data in some manner
  - Apply learning algorithm to produce a system (tagger)
- Testing phase:
  - Apply system to unseen test data
  - Evaluate output

Three Laws of Machine Learning

- Thou shalt not mingle training data with test data
- Thou shalt not mingle training data with test data
- Thou shalt not mingle training data with test data

But what do you do if you need more test data?

Three Pillars of Statistical NLP

- Corpora (training data)
- Representations (features)
- Learning approach (models and algorithms)

Automatic POS Tagging

- Rule-based POS tagging (before)
- Transformation-based learning for POS tagging (now)
- Hidden Markov Models (May be later)
- Maximum Entropy Models (you do it if needed)
- Conditional Random Fields (you do it if needed)

**The problem isn't with rules per se…
but with manually writing rules!**

TBL Painting Algorithm

```
function TBL-Paint
(given: empty canvas with goal painting)

begin

  apply initial transformation to canvas

  repeat

    try all color transformation rules

    find transformation rule yielding most improvements

    apply color transformation rule to canvas

  until improvement below some threshold

end
```

TBL Painting Algorithm

```
function TBL-Paint
(given: empty ca

begin

  apply initial

  repeat

    try all colo

    find transfo                                 improvements

    apply color                                  vas

  until improvem

end
```

Now, substitute:

‘*tag*’ for ‘*color*’
‘*corpus*’ for ‘*canvas*’
‘*untagged*’ for ‘*empty*’
‘*tagging*’ for ‘*painting*’

TBL Painting Algorithm

```
function TBL-Paint
(given: empty canvas with goal painting)
begin
  apply initial transformation to canvas
  repeat
    try all color transformation rules
    find transformation rule yielding most improvements
    apply color transformation rule to canvas
  until improvement below some threshold
end
```

Impossible!

TBL Templates

Change tag **t1** to tag **t2** when:
  w-1 (w+1) is tagged **t3**
  w-2 (w+2) is tagged **t3**
  w-1 is tagged **t3** and w+1 is tagged **t4**
  w-1 is tagged **t3** and w+2 is tagged **t4**

Non-Lexicalized

Change tag **t1** to tag **t2** when:
  w-1 (w+1) is *foo*
  w-2 (w+2) is *bar*
  w is *foo* and w-1 is *bar*
  w is *foo*, w-2 is *bar* and w+1 is *baz*

Lexicalized

Only try instances of these (and their combinations)

TBL Example Rules

He/PRP is/VBZ as/IN tall/JJ as/IN her/PRP$

`Change from` **`IN`** `to` **`RB`** `if w+2 is` *`as`*

He/PRP is/VBZ as/RB tall/JJ as/IN her/PRP$

He/PRP is/VBZ expected/VBN to/TO race/NN today/NN

`Change from` **`NN`** `to` **`VB`** `if w-1 is tagged as` **`TO`**

He/PRP is/VBZ expected/VBN to/TO race/VB today/NN

TBL POS Tagging

- Rule-based, but data-driven
  - No manual knowledge engineering!
- Training on 600k words, testing on known words only
  - Lexicalized rules: learned 447 rules, 97.2% accuracy
  - Early rules do most of the work: 100 → 96.8%, 200 → 97.0%
  - Non-lexicalized rules: learned 378 rules, 97.0% accuracy
  - Little difference… why?
- How good is it?
  - Baseline: 93-94%
  - Upper bound: 96-97%

Source: Brill (Computational Linguistics, 1995)

Three Pillars of Statistical NLP

- Corpora (training data)
- Representations (features)
- Learning approach (models and algorithms)

Penn Treebank Tagset

- Why does everyone use it?
- What's the problem?
- How do we get around it?

What we covered today...

- What are parts of speech (POS)?
- What is POS tagging?
- Methods for automatic POS tagging
  - Rule-based POS tagging
  - Transformation-based learning for POS tagging
- Along the way…
  - Evaluation
  - Supervised machine learning

Information Extraction

- Identify phrases in language that refer to specific types of entities and relations in text.
- Named entity recognition is task of identifying names of people, places, organizations, etc. in text.

  people organizations places

  - Michael Dell is the CEO of Dell Computer Corporation and lives in Austin Texas.
- Extract pieces of information relevant to a specific application, e.g. used car ads:

  make model year mileage price

  - For sale, 2002 Toyota Prius, 20,000 mi, $15K or best offer. Available starting July 30, 2006.

Semantic Role Labeling

- For each clause, determine the semantic role played by each noun phrase that is an argument to the verb.

  agent   patient   source   destination   instrument

  - John drove Mary from Austin to Dallas in his Toyota Prius.
  - The hammer broke the window.
- Also referred to a “case role analysis,” “thematic analysis,” and “shallow semantic parsing”

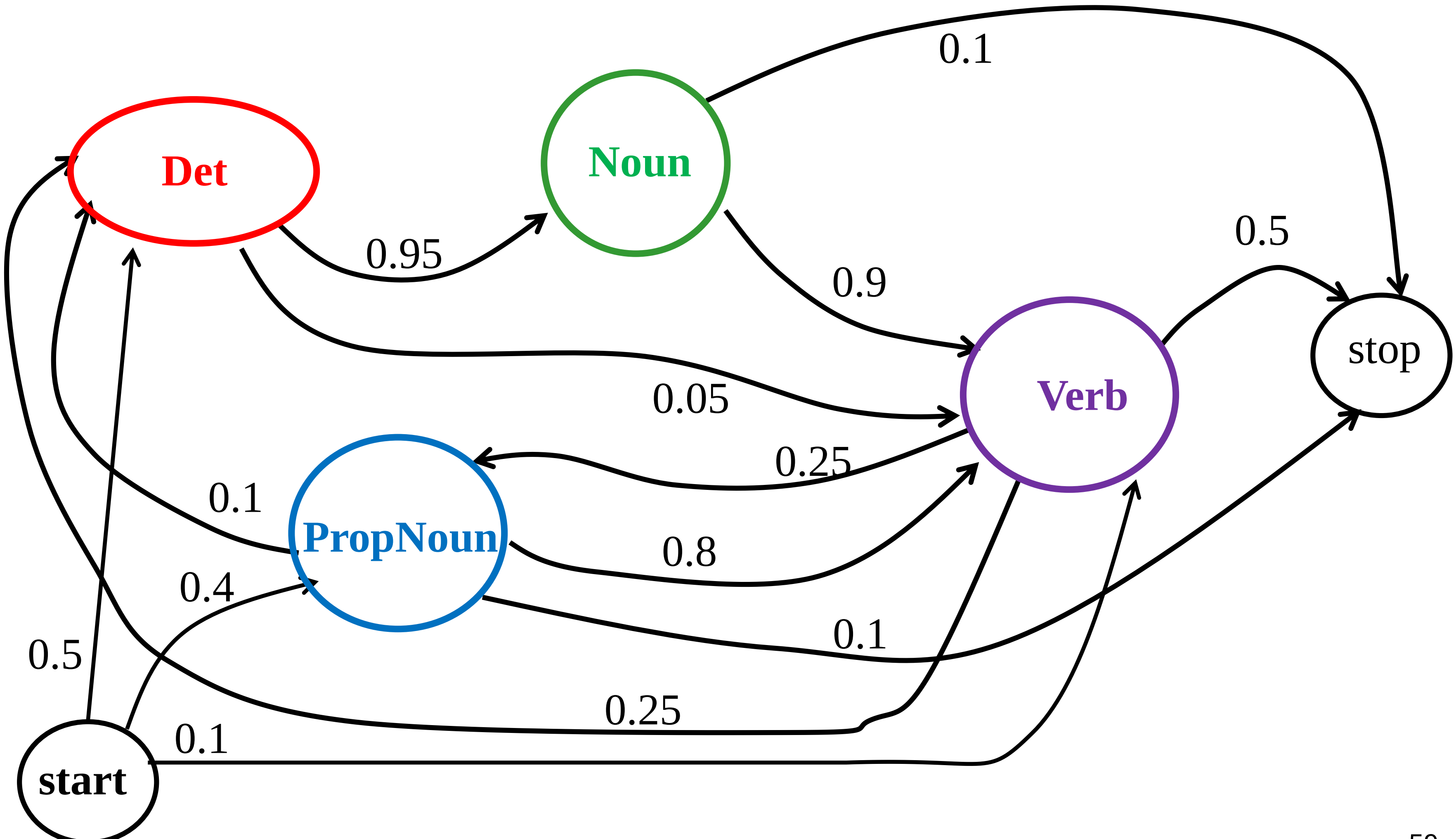
Probabilistic Sequence Models

- Probabilistic sequence models allow integrating uncertainty over multiple, interdependent classifications and collectively determine the most likely global assignment.
- Two standard models
  - Hidden Markov Model (HMM)
  - Conditional Random Field (CRF)

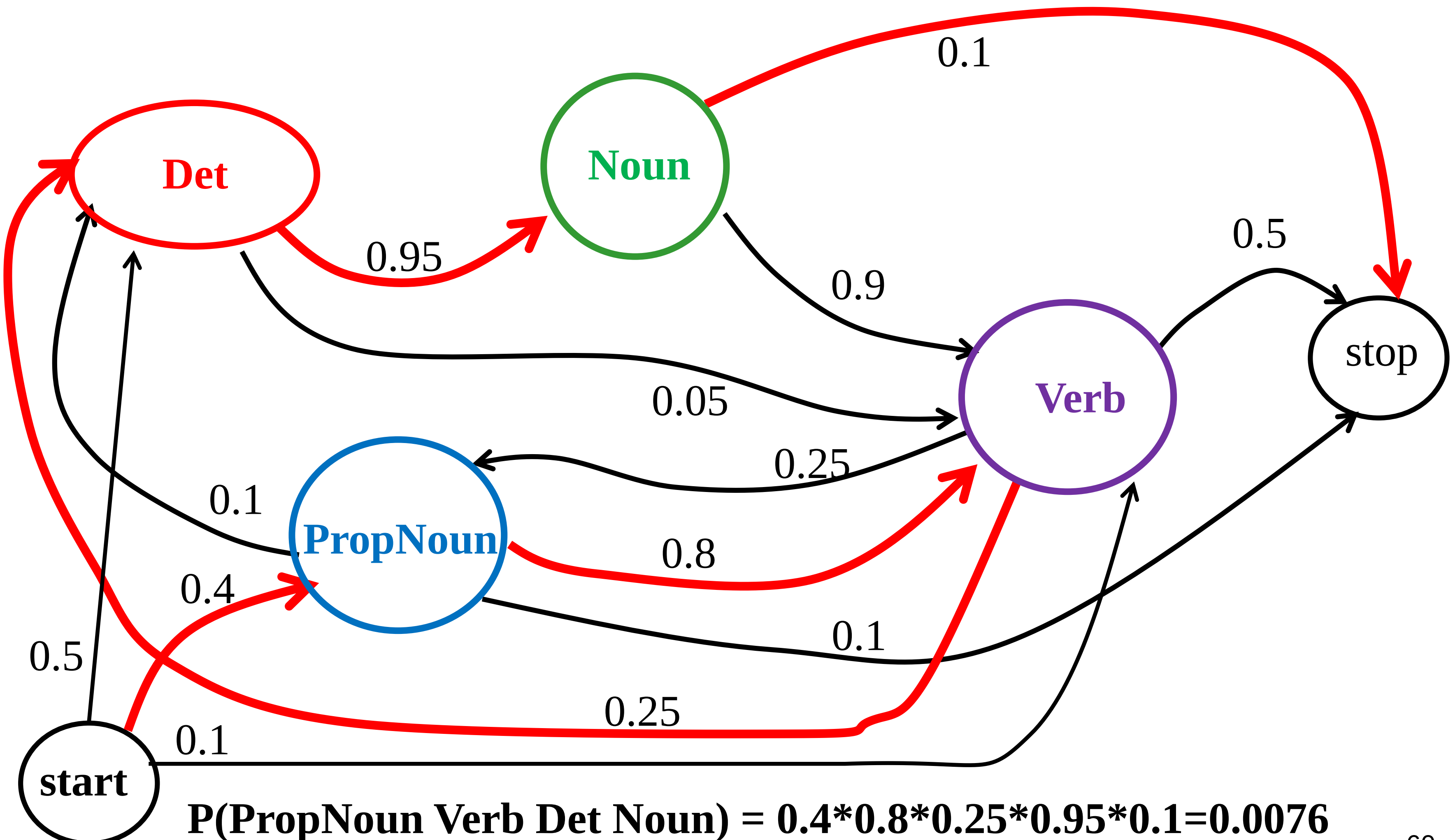
Markov Model / Markov Chain

- A finite state machine with probabilistic state transitions.
- Makes Markov assumption that next state only depends on the current state and independent of previous history.

Sample Markov Model for POS



Sample Markov Model for POS



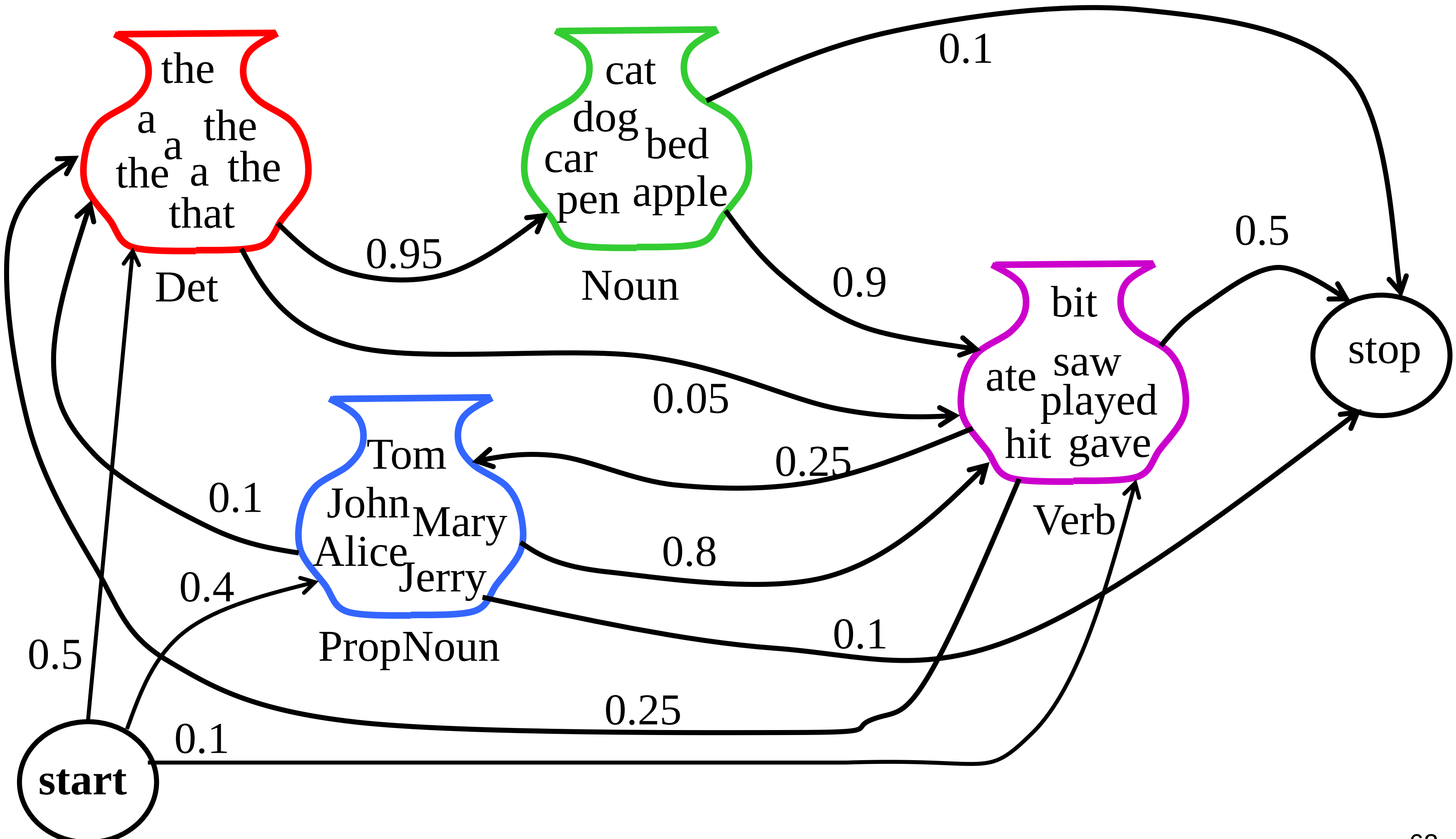
P(PropNoun Verb Det Noun) = 0.4*0.8*0.25*0.95*0.1=0.0076

Time/[V,N] flies/[V,N] like/[V,Prep] an/Det arrow/N

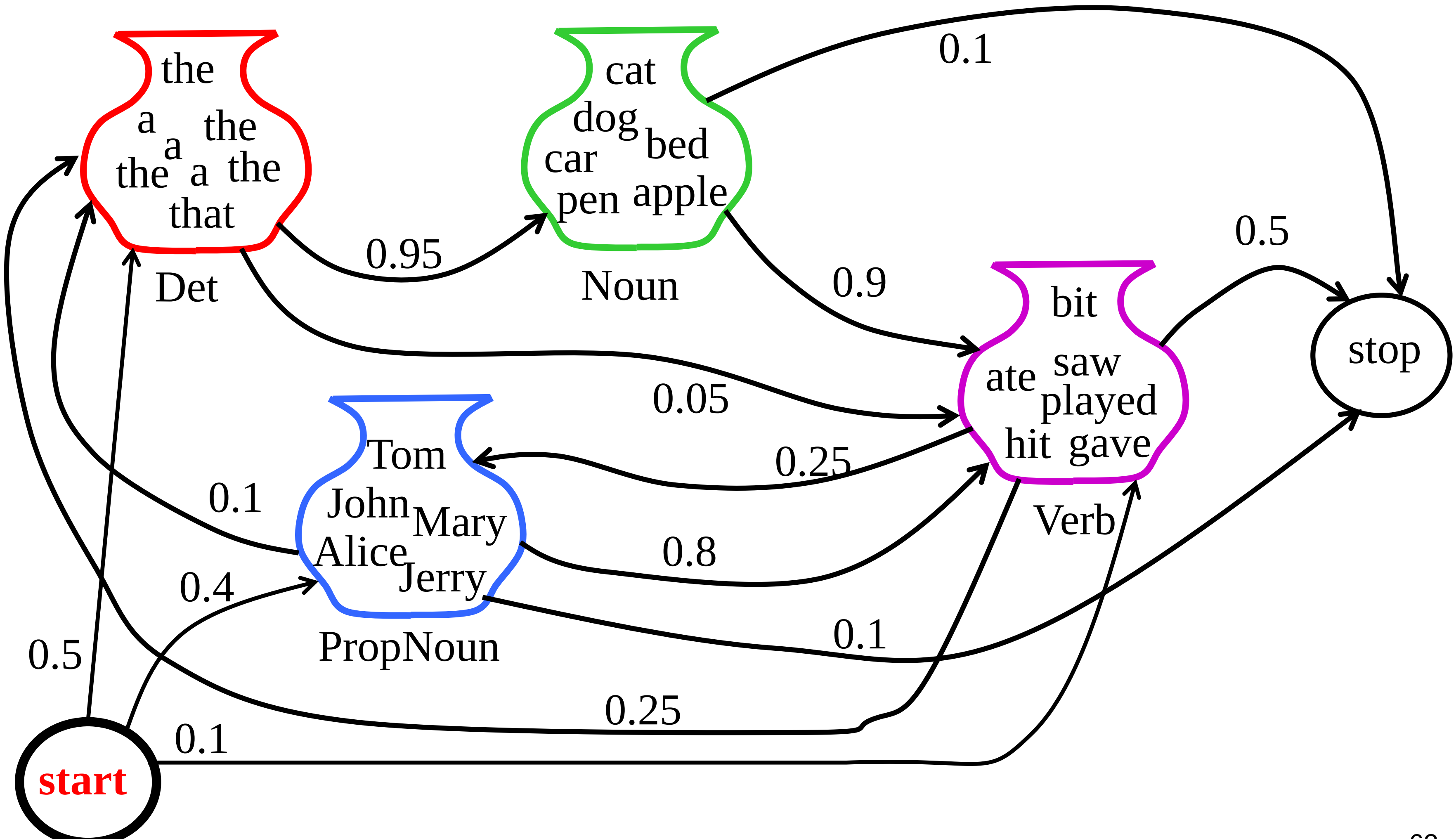
Hidden Markov Model

- Probabilistic generative model for sequences.
- Assume an underlying set of ***hidden*** (unobserved, latent) states in which the model can be (e.g. parts of speech).
- Assume probabilistic transitions between states over time (e.g. transition from POS to another POS as sequence is generated).
- Assume a ***probabilistic*** generation of tokens from states (e.g. words generated for each POS).
- May view as assigning POS tags that maximize the probabilities for the sentence(among all possible)!

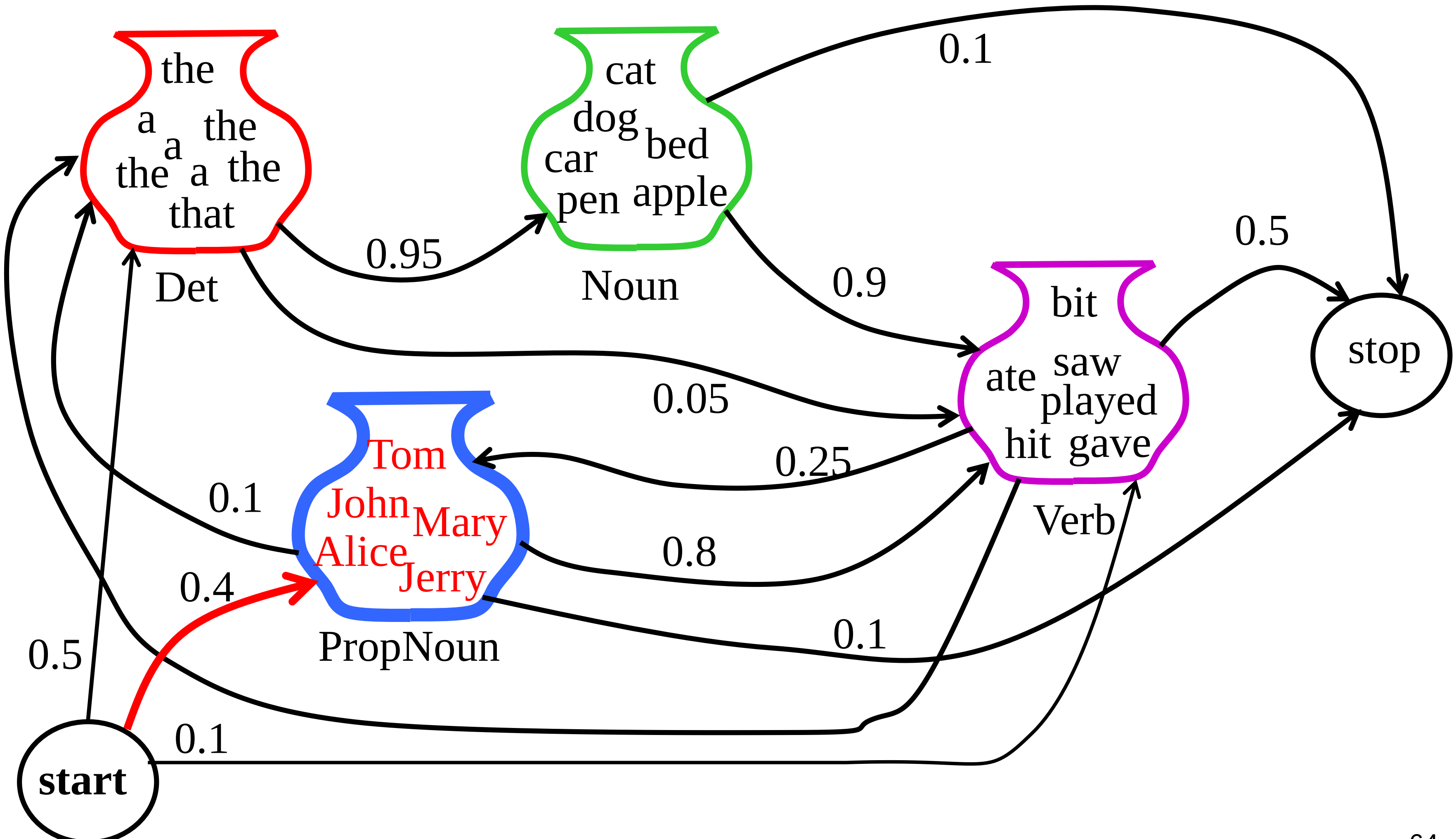
Sample HMM for POS



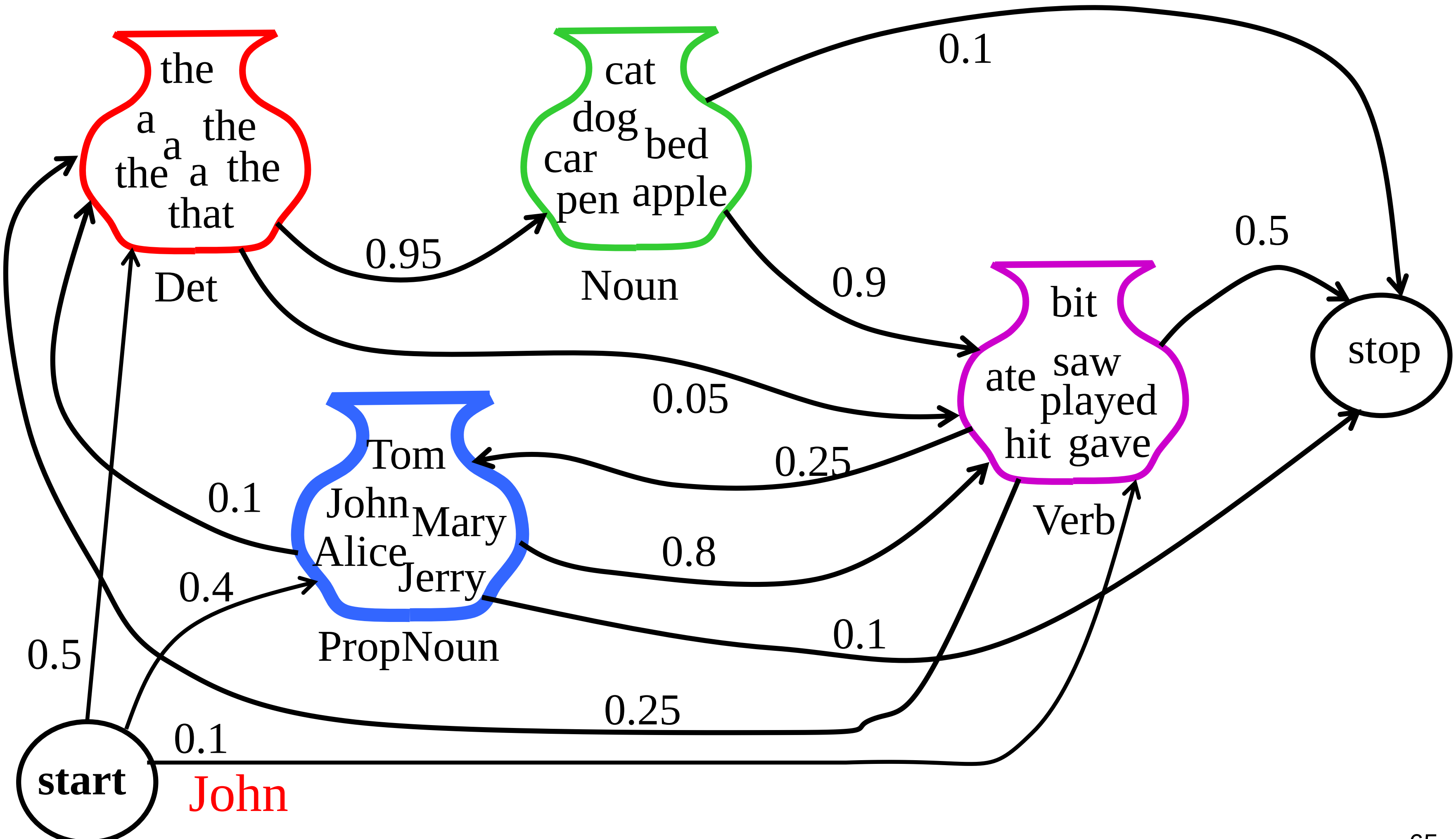
Sample HMM Generation



Sample HMM Generation

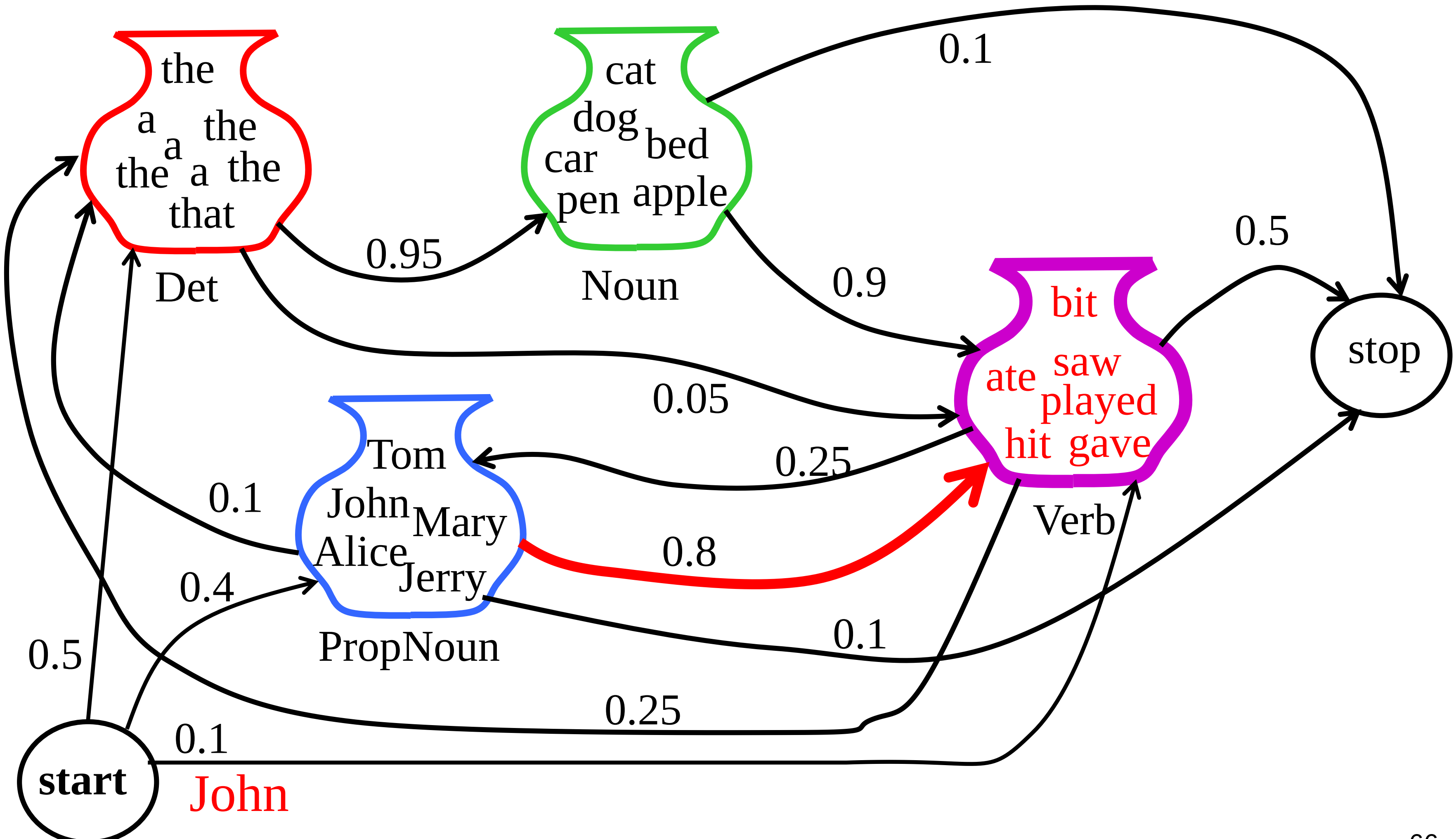


Sample HMM Generation

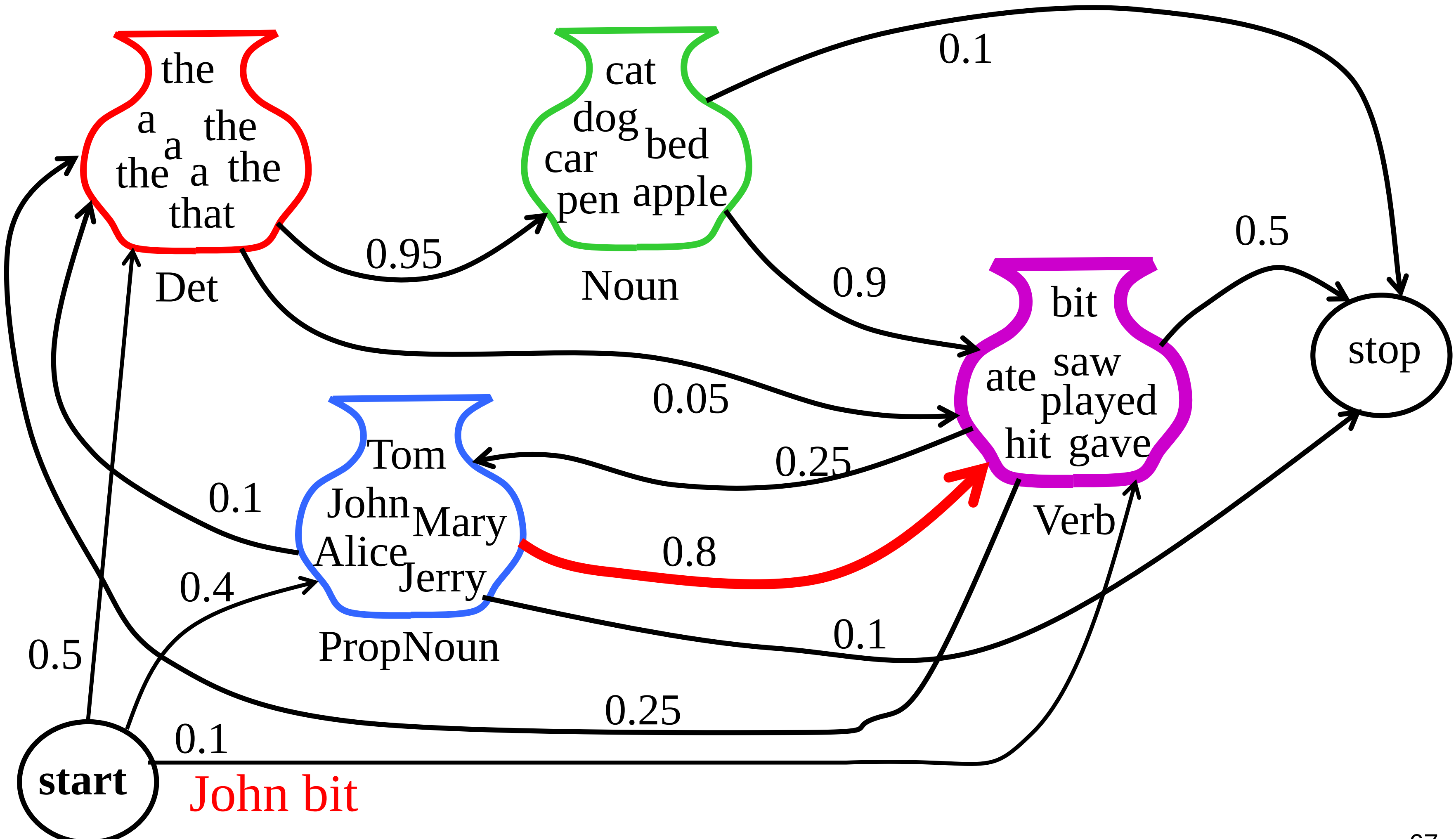


John

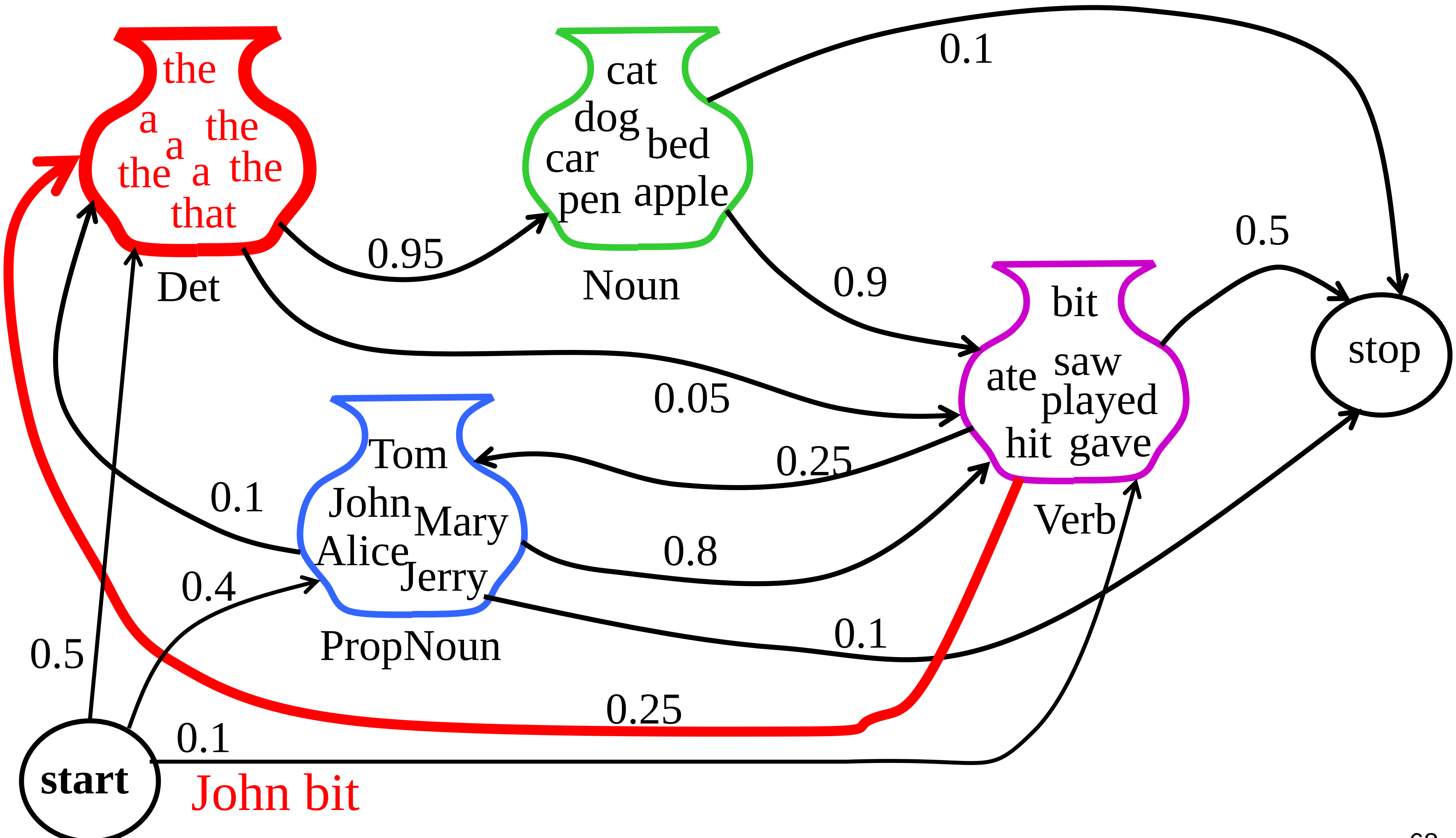
Sample HMM Generation



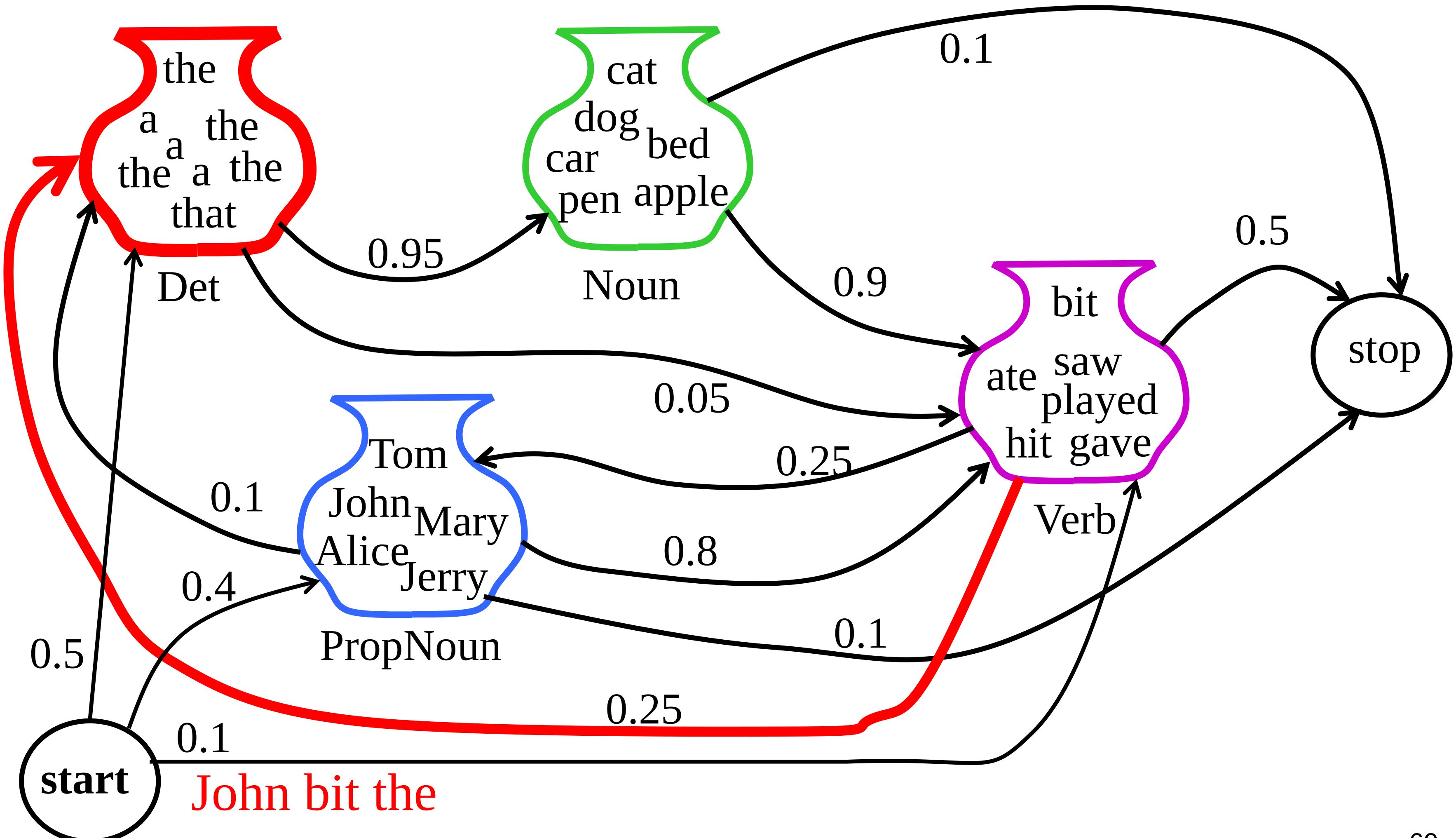
Sample HMM Generation



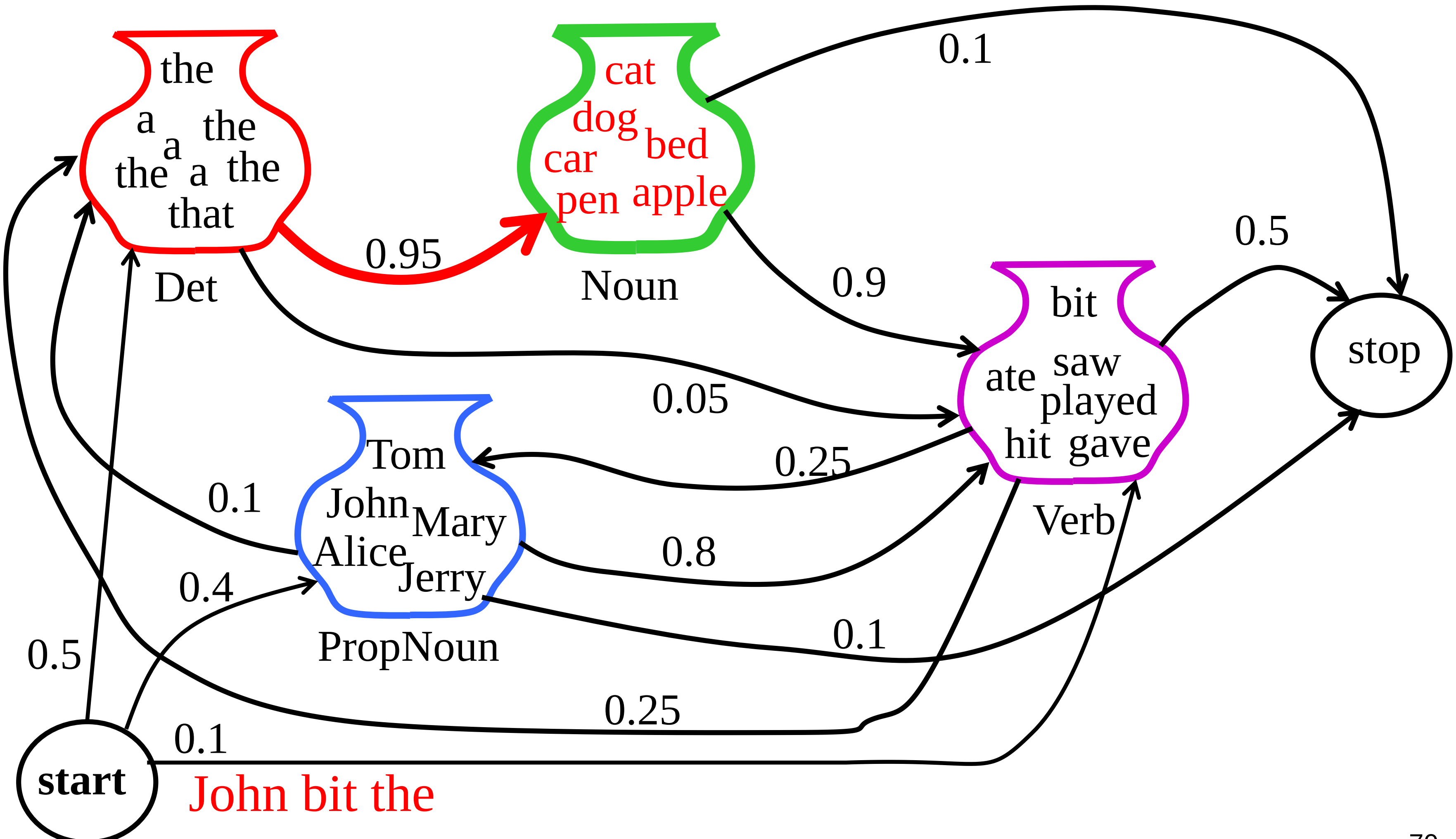
Sample HMM Generation



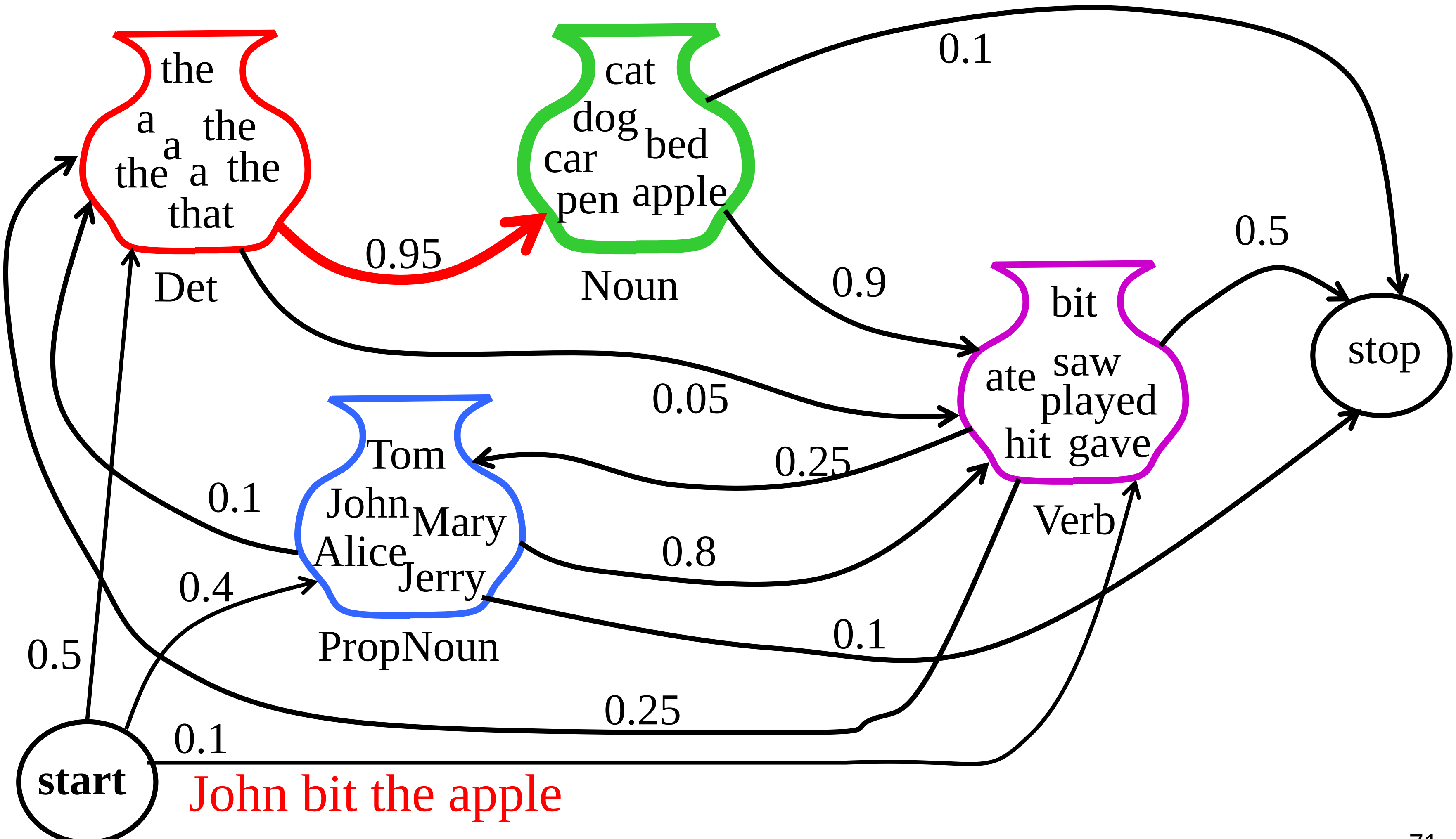
Sample HMM Generation



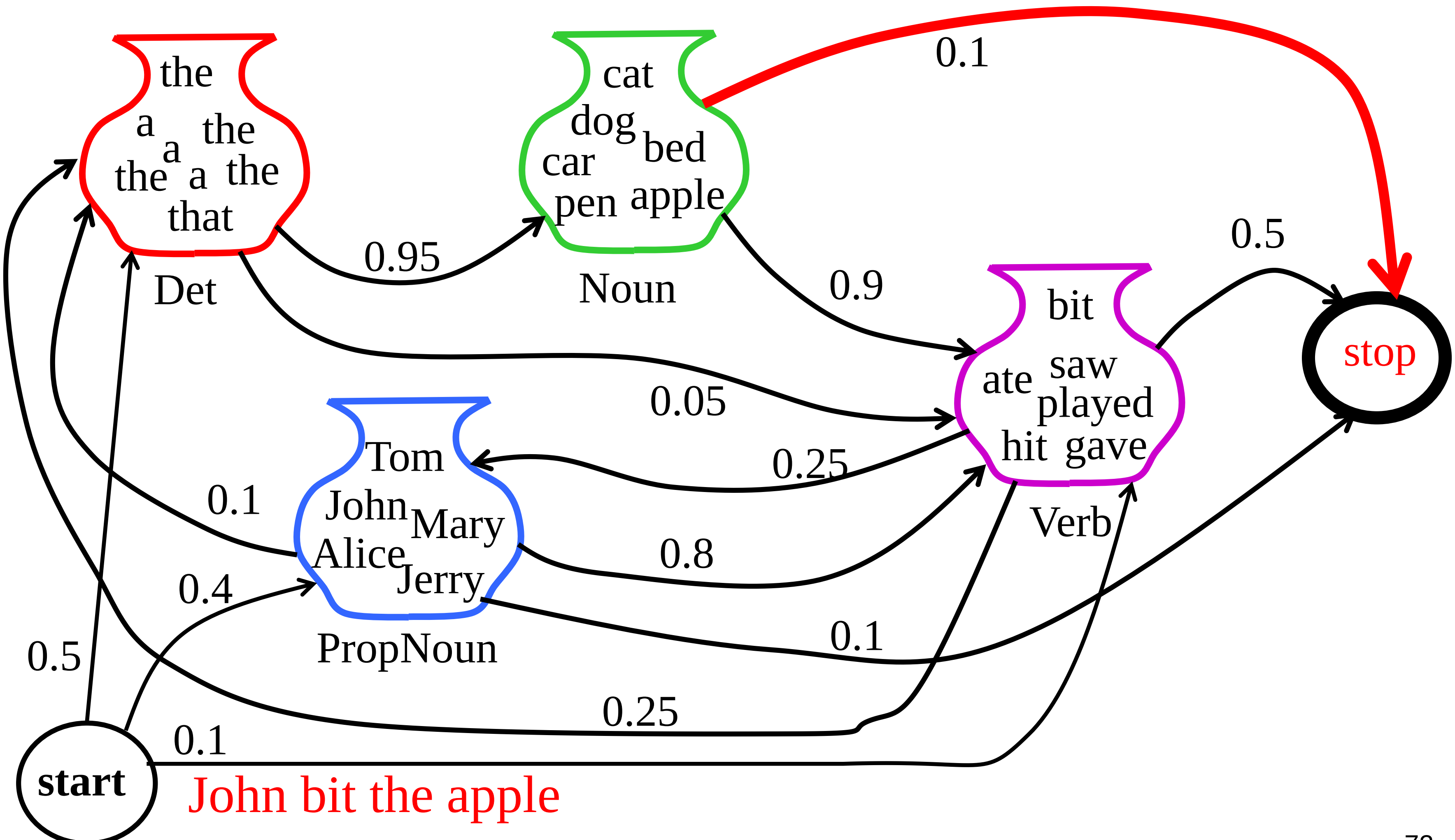
Sample HMM Generation



Sample HMM Generation



Sample HMM Generation



NER: Named Entity Recognition: who, where, when, how much

The task: identify atomic elements of information in text. Mostlt names, could be compound.

Imprtant in many tasks: IE, Translation, Summaries, Better IR

- person names
- company/organization names
- locations
- dates & times
- percentages
- monetary amounts
- **Can be viewed as an extension of POS**
- **Rule based of Machine learning, or combined**