Introduction to NLP

11 Summary

Dr. Paul Buitelaar & Dr. John McCrae

Data Science Institute, NUI Galway

Lecture Schedule

- Introduction
- Foundations
  - Linguistic Concepts
  - Vector Space Model
  - Semantic Analysis
  - Language Modelling
  - Syntactic Analysis
- Applications
  - Information Extraction & Knowledge Graphs
  - Opinion Mining
- Summary and Q&A

Linguistic Concepts

…

cat

The **dog** lies on the mat

child

…

Paradigmatic Relations

Syntagmatic Relations

Linguistic Units - Tokenization

Types vs. Tokens

Multiword Expression

Morphology – Word Formation

Inflection
Derivation
Stemming vs. **Lemmatization**
Decomposition

Syntax - Grammar & Lexicon

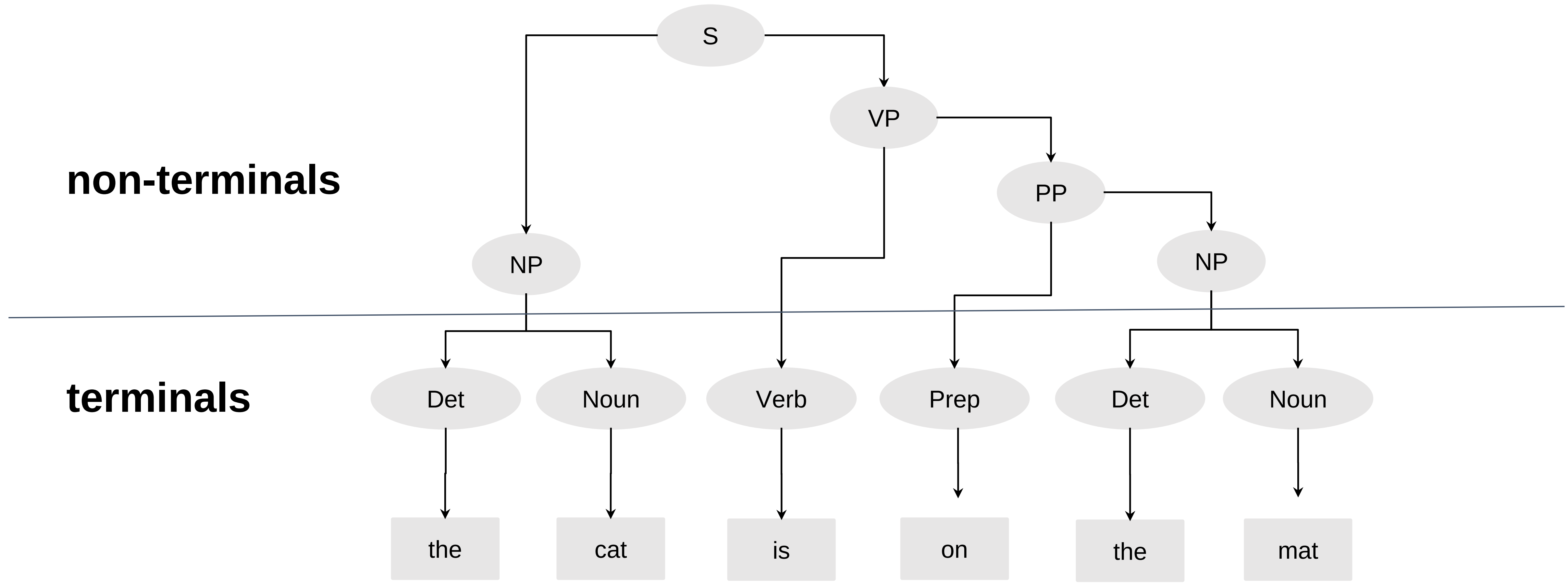
S	-> NP, VP
NP	-> Det, Noun
VP	-> Verb, PP
PP	-> Prep, NP

Grammar

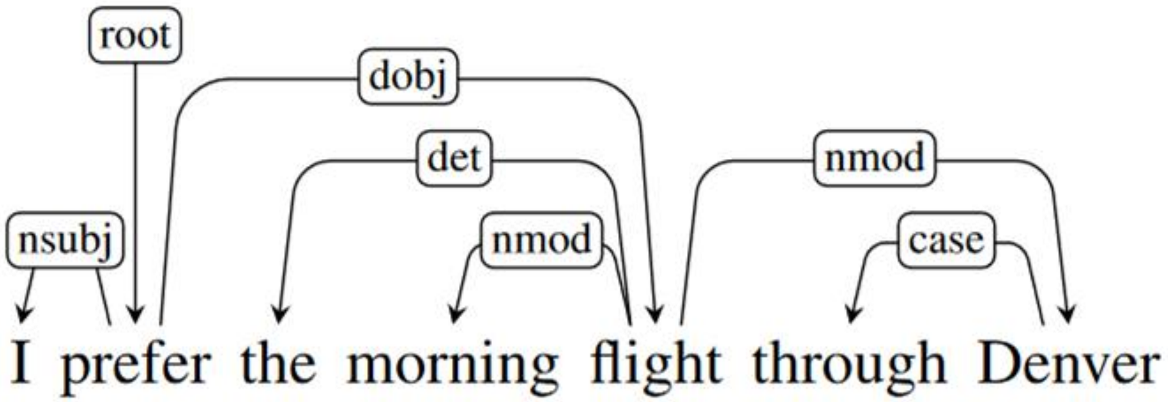
Noun	-> cat, mat
Verb	-> is [“to be “, 3rd, pres]
Preposition	-> on
Determiner	-> the

Lexicon

Syntax – Constituency/Phrase Structure



Syntax – Dependency Structure



(13.1)

Language Data

Corpus

- General vs. Domain-Specific
- Annotated vs. Unannotated (‘raw data’)
- Monolingual vs. Bilingual, Multilingual
- Parallel vs. Comparable
- Multimodal

Lexicon

Vector Space Model

Co-occurrence Matrix

	a	the	on	cat	dog	child	mat	floor	mouse	sits	lies	caught	chased
a	0	0	0	5	3	3	0	0	1	0	0	1	2
the	0	0	6	0	0	0	3	3	0	0	0	0	0
on	0	6	0	0	0	0	0	0	0	3	3	0	0
cat	5	0	0	0	0	0	0	0	0	1	1	1	0
dog	3	0	0	0	0	0	0	0	0	1	1	0	1
child	3	0	0	0	0	0	0	0	0	1	1	0	1
mat	0	3	0	0	0	0	0	0	0	0	0	0	0
floor	0	3	0	0	0	0	0	0	0	0	0	0	0
mouse	1	0	0	0	0	0	0	0	0	0	0	0	0
sits	0	0	3	1	1	1	0	0	0	0	0	0	0
lies	0	0	3	1	1	1	0	0	0	0	0	0	0
caught	1	0	0	1	0	0	0	0	0	0	0	0	0
chased	2	0	0	0	1	1	0	0	0	0	0	0	0

A cat sits on the mat.
A cat lies on the floor.
A cat caught a mouse.

A dog sits on the mat.
A dog lies on the floor.
A dog chased a cat.

A child sits on the mat.
A child lies on the floor.
A child chased a cat.

CORPUS

Word Vectors - Similarity

	a	sits	lies	caught	chased
cat	5	1	1	1	0
dog	3	1	1	0	1
child	3	1	1	0	1

$$\text{cosine}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

cos (cat, dog) *= 15+1+1+0+0 / (√25+1+1+1+0 * √9+1+1+0+1)*
*= 17 / (√28 * √12) = 17 / (5.29 * 3.46)* ***0.93***

cos (cat, child) *= 15+1+1+0+0 / (√25+1+1+1+0 * √9+1+1+0+1)*
*= 17 / (√28 * √12) = 17 / (5.29 * 3.46)* ***0.93***

cos (dog, child) *= 9+1+1+0+1 / (√9+1+1+0+1 * √9+1+1+0+1)*
*= 12 / (√12 * √12) = 12 / (3.46 * 3.46)* ***1.00***

Pointwise Mutual Information (PMI)

Pointwise mutual information:

Do events x and y co-occur more than if they were independent?

$$\mathrm{PMI}(X,Y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

PMI between two words: (Church & Hanks 1989)

Do words x and y co-occur more than if they were independent?

$$\mathrm{PMI}(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

Semantic Analysis

Lexical Semantic Ambiguity

Homonymy

Synonymy

Antonymy

Semantic Lexicons

WordNet, organized by 'synsets' – defines meaning by a set of synonyms

https://wordnet.princeton.edu/

FrameNet, organized by 'frames' – defines meaning by typical semantic roles

https://framenet.icsi.berkeley.edu/fndrupal/about

Word Sense Disambiguation

```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

 best-sense ← most frequent sense for word
 max-overlap ← 0
 context ← set of words in sentence
 for each sense in senses of word do
   signature ← set of words in the gloss and examples of sense
   overlap ← COMPUTEOVERLAP(signature, context)
   if overlap > max-overlap then
        max-overlap ← overlap
        best-sense ← sense
 end
 return(best-sense)
```

NUI Galway
OÉ Gaillimh

Semantic Role Labeling

AGENT	The volitional causer of an event	*The waiter* spilled the soup.
EXPERIENCER	The experiencer of an event	*John* has a headache.
FORCE	The non-volitional causer of the event	*The wind* blows debris from the mall into our yards.
THEME	The participant most directly affected by an event	Only after Benjamin Franklin broke *the ice*...
RESULT	The end product of an event	The city built a *regulation-size baseball diamond*...
CONTENT	The proposition or content of a propositional event	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	An instrument used in an event	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	The beneficiary of an event	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	The origin of the object of a transfer event	I flew in *from Boston*.
GOAL	The destination of an object of a transfer event	I drove *to Portland*.

Language Modelling

NUI Galway
OÉ Gaillimh

Language Modelling

Noisy Channel Model: $p(Y|X) \propto p(X|Y)\, p(Y)$
Applications to machine translation, spelling correction, etc.

Estimating probabilities by counting: $$p(w) = \frac{c(w)}{\sum_{v \in W} c(v)}$$

Add-one smoothing: $$p(w) = \frac{c(w) + 1}{\sum_{v \in V} c(v) + |V|}$$

n-gram Language Models: $$p(w_1 w_2 \dots w_n) = \prod_{k=1,\dots,n} p(w_k | w_{k-m+1} \dots w_{k-1})$$

NUI Galway
OÉ Gaillimh

Language Modelling

Why perform smoothing?

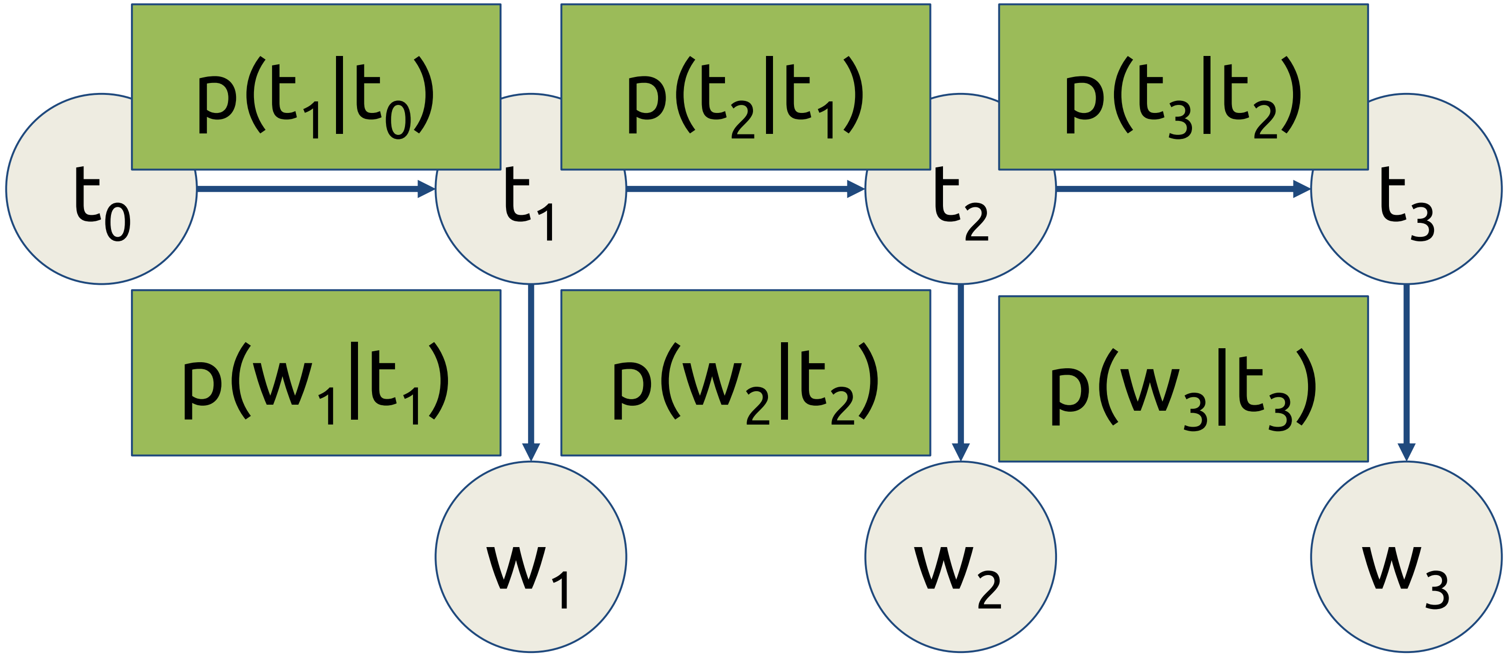
Back-off and Interpolation

Perplexity to evaluate language models

Syntactic Analysis

NUI Galway
OÉ Gaillimh

Tagging: Hidden Markov Model



$$p(w_1, \ldots, w_n, t_1, \ldots, t_n) \approx \prod_{i=1,\ldots,n} p(t_i | t_{i-1}) p(w_i | t_i)$$

NUI Galway
OÉ Gaillimh

Three fundamental problems for HMMs

1. What is the probability of a sequence given an observation and a model?
   a. What is P(DET N VBZ DET N|the cat chases the mouse, μ)?
2. What is the probability of an observation given the model?
   a. What is P(the cat chases the mouse|μ)?
3. What is the model that maximizes the likelihood of the observed data and known sequences?
   a. What μ maximizes P(the cat chases the mouse, DET N VBZ DET N|μ)?
   b. What μ maximizes P(the cat chases the mouse|μ)?

Viterbi algorithm

Set $\pi_{s,0}$=0 except for $\pi_{\text{Start},0}$=1

Set y_s = []

For i from 1 to T

  For s ∈ S

    Set $\pi_{s,i} = \max_{t \in S} \pi_{t,i-1} p(s|t) p(w_i|s)$

    Append t to y_s

Return y_s + s where s

  maximizes $\pi_{s,T}$

Forward Algorithm also!

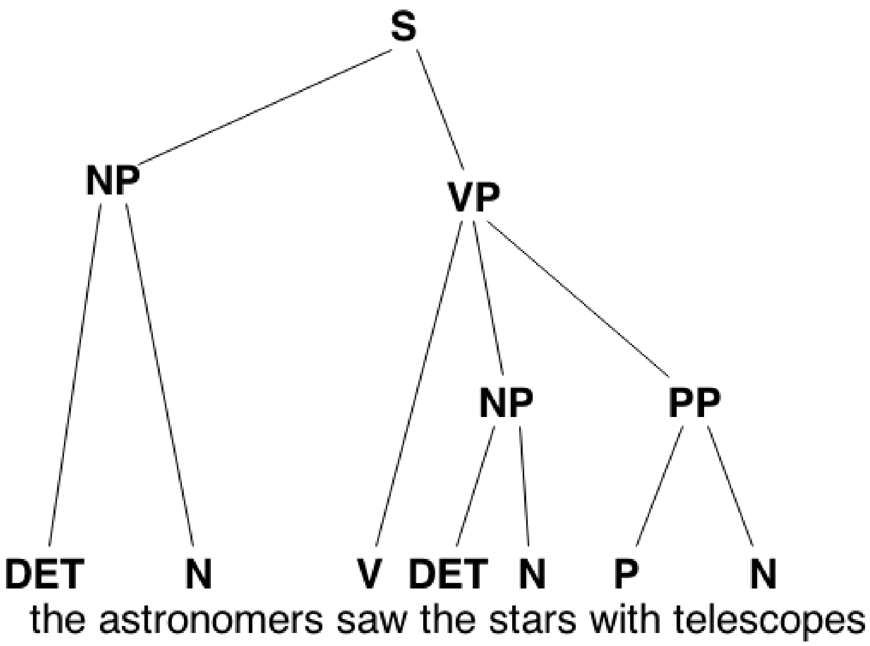
Supervised Learning of HMMs

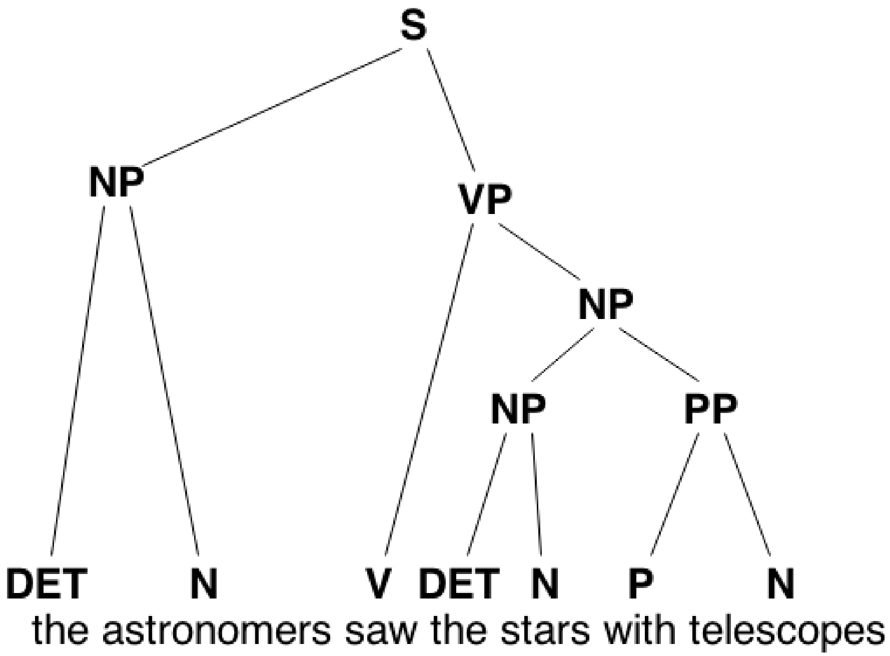
$$p(s_i|s_j) = \frac{c(t_{i-1} = s_j, t_i = s_i)}{\sum_{s'} c(t_{i-1} = s_j, t_i = s')}$$

$$p(w|s) = \frac{c(w_i = w, t_i = s)}{c(t_i = s)}$$

NUI Galway
OÉ Gaillimh

Parsing: Ambiguity





Context-free grammars

Recall, a context-free grammar G=(N,Σ,P,S) consists of:

- A set of non-terminal symbols N
  - e.g., 'N', 'VP', 'S'
- A set of terminal symbols Σ
  - e.g., 'cat', 'astronomer', 'the'
- A set of productions P
  - e.g., 'S → NP VP'
- A start symbol
  - Normally 'S'
- (PCFG) A probability function *D*

$$\sum_{\{\beta : A \to \beta \in P\}} D(A \to \beta) = 1 \quad \forall A \in N$$

NUI Galway
OÉ Gaillimh

CYK Algorithm

Set $t_{i,j,a} = -\infty$ for all values

For $i = 1, \ldots., n$

  For $A \rightarrow w_i \in P$

    $t_{i,i+1,A} = D(A \rightarrow w_i)$

For $k = 1, \ldots , n$; $i = 1, \ldots , n - k + 1$; $j = i + k$

  For $A \rightarrow \beta \in P$

    If β matches between i and j

      $S = D(A \rightarrow \beta) \times \prod_{i',j',A'} t_{i',j',A'}$ where $\{i',j',A\}$ are the matches

      If $s > t_{i,j,A}$

        $t_{i,j,A} = s$

NUI Galway
OÉ Gaillimh

Chomsky Normal Form

CYK is only polynomial if all rules are of the form

- $A \rightarrow BC$ or $A \rightarrow a$

Any PCFG can be easily transformed to Chomsky Normal Form

Problems of PCFGs

Lexical Dependencies
Lexical Attachment
Parse ambiguity not distinguished

Solutions

Lexicalized PCFGs
Dependency Grammars

Information Extraction & Knowledge Graphs

IE Approaches

Lexical lookup

Rules

Machine learning

NUI Galway
OÉ Gaillimh

Supervised Learning

IOB sequence annotation

Words	IOB Label	IO Label
American	B-ORG	I-ORG
Airlines	I-ORG	I-ORG
,	O	O
a	O	O
unit	O	O
of	O	O
AMR	B-ORG	I-ORG
Corp.	I-ORG	I-ORG
,	O	O
immediately	O	O
matched	O	O
the	O	O
move	O	O
,	O	O
spokesman	O	O
Tim	B-PER	I-PER
Wagner	I-PER	I-PER
said	O	O
.	O	O

Inter-Annotator Agreement

Cohen's kappa coefficient

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Semi-Supervised - Distant Learning

Wikipedia

- Info-box provides seeds
- Corresponding text can serve as 'annotation' for training purposes

Rex Wayne Tillerson (born March 23, 1952) is an American former government official and former energy executive who served as the 69th United States Secretary of State from February 1, 2017, to March 31, 2018, under President Donald Trump.[1][2][3] Originally a civil engineer, Tillerson joined Exxon in 1975. He rose to become chairman and chief executive officer of ExxonMobil, holding that position from 2006 until 2017, when he left to join the Trump administration.

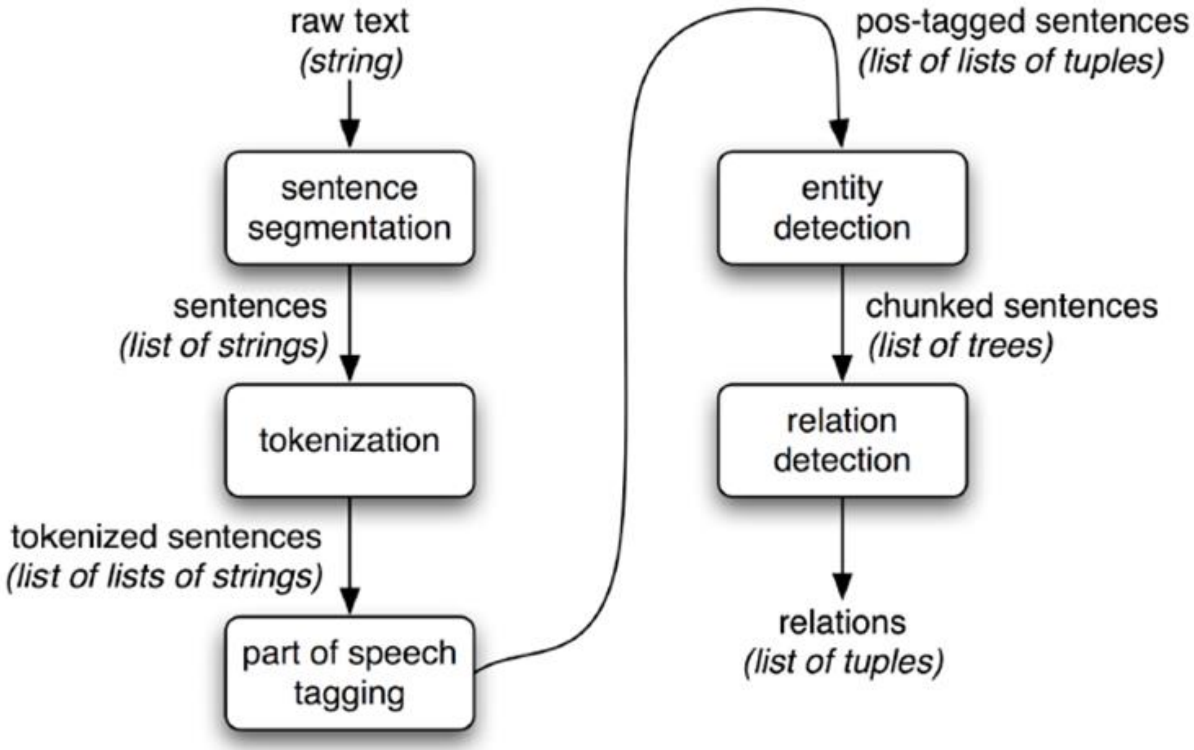
Rex Tillerson

69th United States Secretary of State

In office
February 1, 2017 – March 31, 2018[a]

President	Donald Trump
Deputy	John Sullivan
Preceded by	John Kerry
Succeeded by	Mike Pompeo
Personal details	
Born	Rex Wayne Tillerson March 23, 1952 (age 66) Wichita Falls, Texas, U.S.

Unsupervised - Open IE Architecture



IE Evaluation

Precision

$$P = \frac{\#\ correctly\ extracted\ items}{Total\ \#\ of\ extracted\ items}$$

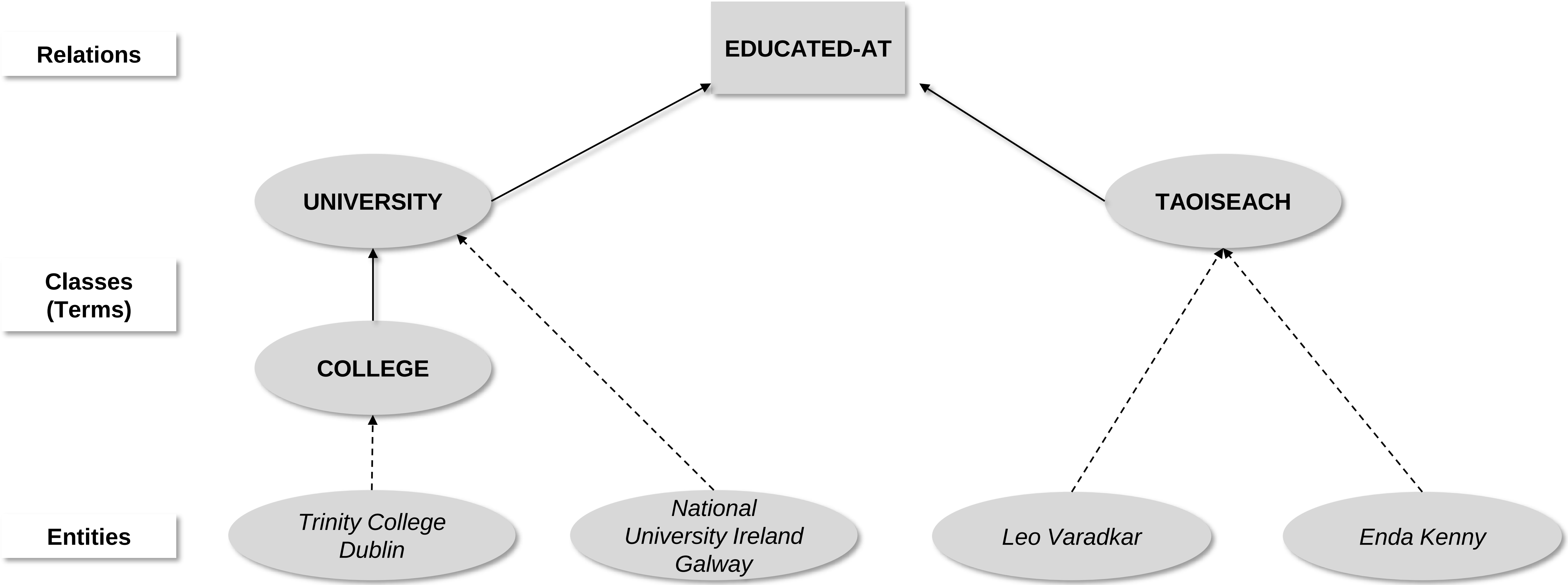
Recall

$$R = \frac{\#\ correctly\ extracted\ items}{Total\ \#\ of\ gold\ items}$$

F-Score (weighted harmonic mean)

$$F = \frac{2 \times P \times R}{P + R}$$

Knowledge Graph – Elements



NUI Galway
OÉ Gaillimh

Entity Linking – Disambiguation Features

Disambiguation according to

- Contextual information around the entity
- Popularity of the entity
- Coherence across different entities

Disambiguation features can be

- Text-based
- Graph-based

Term Extraction – Unsupervised Learning

Extract, rank and filter NPs and/or apply Hearst patterns

Taxonomy Extraction - Unsupervised

Substrings

Hearst patterns

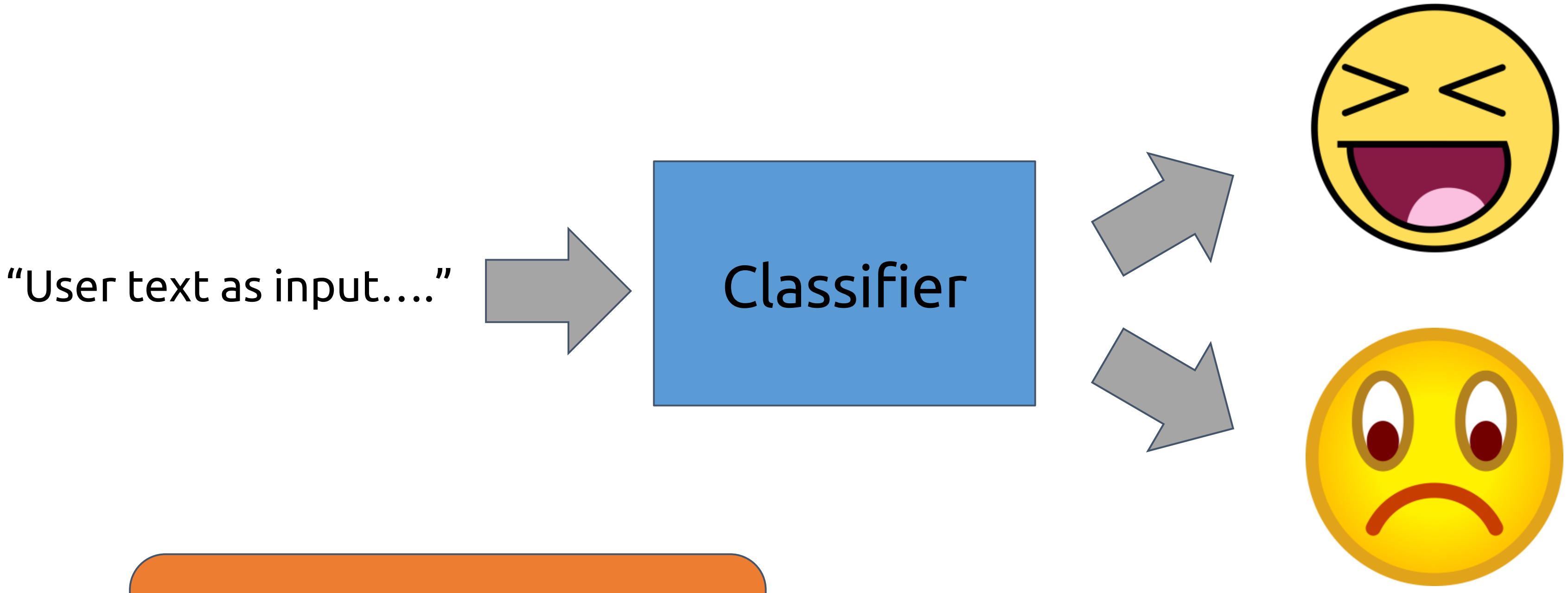
Distributional models

Taxonomy Extraction – Supervised

Training over pairs of sub/super classes

Opinion Mining

What is Sentiment Analysis?



f(text) = +1 or - 1

NUI Galway
OÉ Gaillimh

Features from a sentiment lexicon

The camera's focus was bad, but has a great size and is easy-to-use

$$\begin{bmatrix} 2/13 \\ 1/13 \end{bmatrix}$$

Count Vectors

Sentiment Lexicon

Positive:
good
great
happy
easy
….

Negative:
bad
sad
hard
poor
….

Negation Feature Examples

I do not NOT_like NOT_this NOT_new NOT_Nokia NOT_model

Bag-of-words vector

I	1
do	1
not	1
like	0
NOT_like	1
….	

Aspect-based Sentiment Analysis

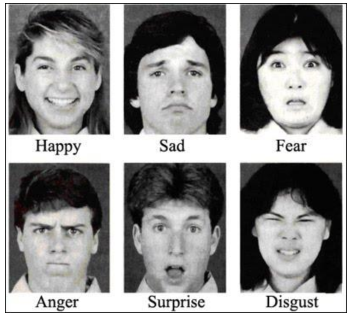
The staff was very friendly and informative. The bus stop, restaurants and railway station are at a walking distance. The breakfast did not have much variety but everything was fresh and tasted very good. The room was comfortable, but the bathroom was very small.

Identify aspect term/mention

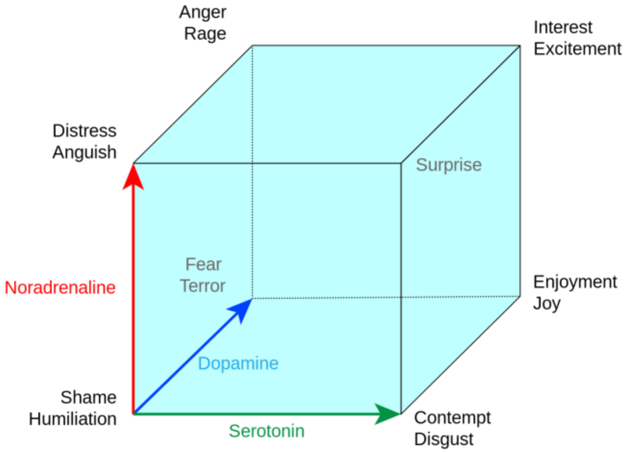
Aspect	Sentiment
Staff	positive
Location	positive
Breakfast	conflict
Room	positive
Bathroom	negative

Identify sentiments towards the aspect term.

Emotion Analysis

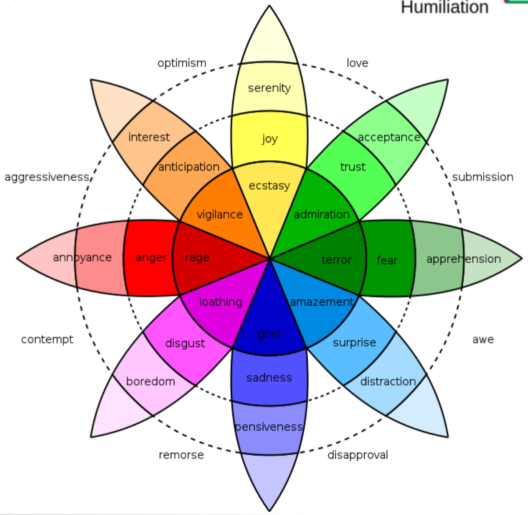


Ekman's 6 Emotions



Lövheim's 3D (VAD) Model

Plutchik's 8 (32) Emotions



Cambria's Hourglass