ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΜΑΤΙΚΗΣ

ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Καραγιώργου Σοφία

Overview

- Computational Linguistics vs Natural Language Processing
- Some informative history
- Going Distributional (Part of Speech Tagging)
- Distributional Syntax (Language Modeling)
- Vector Space Modeling and Information Retrieval
- Word Sense Disambiguation

Text handling

- Text for human readers but also to be handled by computers
  - editing, storing and indexing, searching and retrieving, ranking, classifying, extracting information and knowledge, question answering,…

Handling text data/Text analytics

- Knowledge Representation & Reasoning/Tagging
  - Semantic Web
- Computational Linguistics
  - NLP
- Search & DB
- Data analysis
  - ML & Text mining

Early text learning

- Machine learning methods and tasks
  - Inspired by information retrieval, classifying document regarding relevance for a query
  - Personalized information delivery
  - Document categorization into class hierarchy

Machine Learning on text data

- Text representation
  - Words and word sequences as features in ML setting
  - Handling sparse feature vectors
  - Efficient feature selection

Text Analytics

- Information extraction from the Web
- Combining text, graphs, databases, images,…
- Capturing semantics
- Knowledge management

Example tasks

- Visualization of text data
- Semi-automatic ontology construction
- Text annotation
- Extracting triplets from text
- Document summarization
- Question answering
- Ontology construction and extension
- Social network analysis and text analytics

Technologies

- Statistical Machine Learning
- Data/Web/Text/Stream-Mining
- Graph/Social Network Analysis
- Complex Data Visualization
- Computational Linguistics
- Social Computing/Web2.0
- Light-Weight Semantic Technologies
- Deep Semantics & Reasoning

Linguistics

- The scientific study of language
- Goal is discovery of universal generalizations that apply to all languages, or sets of languages, or whatever the natural grouping is
- Discovery of what the natural kinds of language are (like whether or not there are natural groupings, etc.)

Computational Linguistics

- Applying computational tools to the study of language
- Corpus linguistics: using data sets of language to formulate and test theories of language
- Using the results from computational complexity to limit the types of theories of language we develop
- Stays true to the natural phenomenon of language

Natural Language Processing

- Text processing with a goal
- Dealing with spoken words is called “Speech Recognition”
  - Often the output of an ASR system is the input to an NLP system
- Sets of algorithms that solve problems that people want solved
  - Topic detection
  - Authorship detection
  - Even semantic understanding (not here yet, although on the way)
- Can use any technique that works
- Doesn’t pretend to stay true to cognitive

History

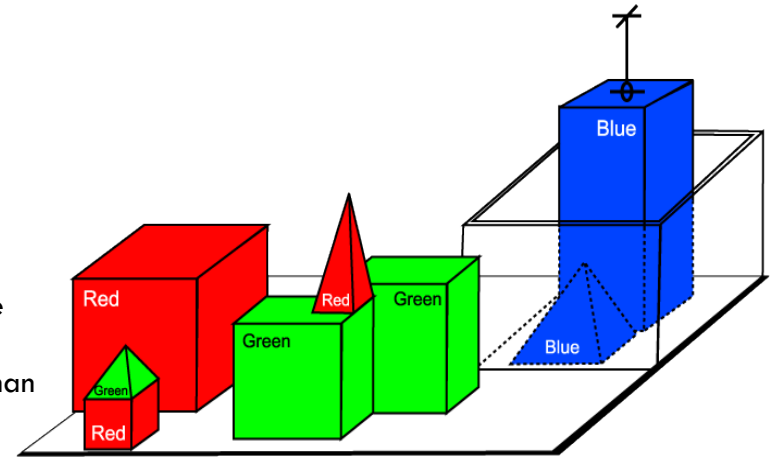
- Goals: systems that behaved comparably to humans on given linguistic tasks
- Models often build on models of human reasoning, e.g., reasoning by analogy, logic, etc.
- SHRDLU (Winograd, 1971): interaction with simulation of tabletop
- ELIZA (Weizenbaum, 1966): does it work?
- Weizenbaum thought no, secretary thought yes
- Worked because:
  - Restricted domain plus **LOTS of person-hours on given domain**
  - Interactive nature allowed for easy correction
- Very brittle

ELIZA

- An early natural language processing computer program - emulates a Rogerian psychotherapist
- Simulated conversation by using a 'pattern matching' and substitution methodology that gave users an illusion of understanding on the part of the program, but had no built in framework for contextualizing events
- Directives on how to interact were provided by 'scripts'

SHRDLU

- An early natural language understanding computer program
- A language parser that allowed user interaction using English terms
- Example:
  - Person: Pick up a big red block
  - Computer: OK
  - Person: Grasp the pyramid
  - Computer: I don't understand which pyramid you mean
  - Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box
  - Computer: By "it", I assume you mean the block which is taller than the one I am holding
  - Computer: OK



It is difficult to access unstructured information efficiently

- **Information extraction** tools can help you:
  - Save time and money on **management of text** and data from multiple sources
  - Find **hidden links** scattered across huge volumes of diverse information
  - Integrate **structured data** from variety of sources
  - **Interlink** text and data
  - **Collect information** and extract new facts

What is information extraction?

- Information Extraction (IE) is the automatic discovery of new, previously unknown information, by automatically extracting information from different textual resources
- A key element is the linking together of the extracted information together to form new facts or new hypotheses to be explored further
- In IE, the goal is to discover previously unknown information, i.e. something that no one yet knows

Statistical analysis

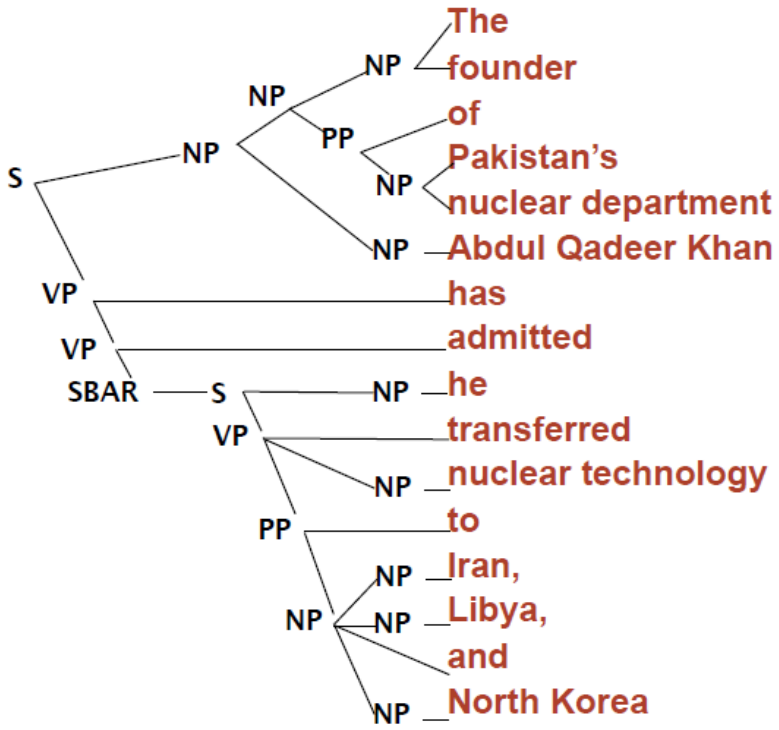
- Church (1985): statistical etymology classifier (important for stress in speech synthesis, calculus vs. spaghetti)
  - Classified a word based on probabilities over 3 letter segments (tri-grams)
- $P(xyz|l) = C_l(xyz)/C_l(*)$
- Choose the l that maximizes the probability
- Worked!

Modeling with words/word pairs

- The founder of Pakistan's nuclear program, Abdul Qadeer Khan, has admitted he transferred nuclear technology to Iran, Libya and North Korea, a Pakistani government official said Monday
- Khan made the confession in a written statement submitted "a couple of days ago" to investigators probing allegations of nuclear proliferation by Pakistan, the official told The Associated Press on condition on anonymity
- The transfers were made during the late 1980s and in the early and mid 1990s, and were motivated by "personal greed and ambition", the official said
- The official said the transfers were not authorized by the government

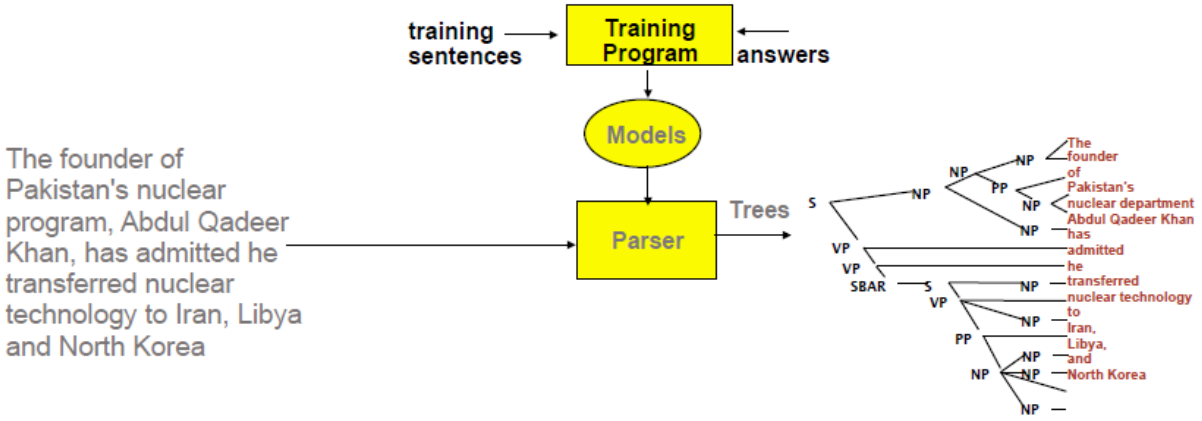
Word	# in Document
Khan	15
nuclear	14
Pakistan	10
transfers	9
official	8
scientists	5
journalists	5
government	5
Libya	5
officials	4
military	4

Phrase Structure requires annotation



Machine Learning with Corpora

- 1990: Best hand-built parsers ~40%-60% accuracy
- 1995+: Statistical parsers 90% accuracy



IE is not Data Mining

- Data mining is about using analytical techniques to find interesting patterns from large structured databases
- Examples:
  - using consumer purchasing patterns to predict which products to place close together on shelves in supermarkets
  - analyzing spending patterns on credit cards to detect fraudulent card use

IE is not Web Search

- IE is also different from traditional web search or IR
- In search, the user is typically looking for something that is already known and has been written by someone else
- The problem lies in sifting through all the material that currently isn't relevant to your needs, in order to find the information that is
- The solution often lies in better ways to ask the right question
- You can't ask Google to tell you about
  - all the speeches made by Tony Blair about foot and mouth disease while he was Prime Minister
  - all the documents in which a politician born in Sheffield is quoted as saying something about hospitals

Information Extraction Basics

- Entity recognition is required for…
- Relation extraction which is required for…
- Event recognition which is required for…
- Summarisation tasks

What is Entity Recognition?

- Entity Recognition is about recognizing and classifying key Named Entities and terms in the text
- A **Named Entity** is a Person, Location, Organization, Date etc.
- A **term** is a key concept or phrase that is representative of the text

Mitt Romney, the favorite to win the Republican nomination for president in 2012

Person Term Date

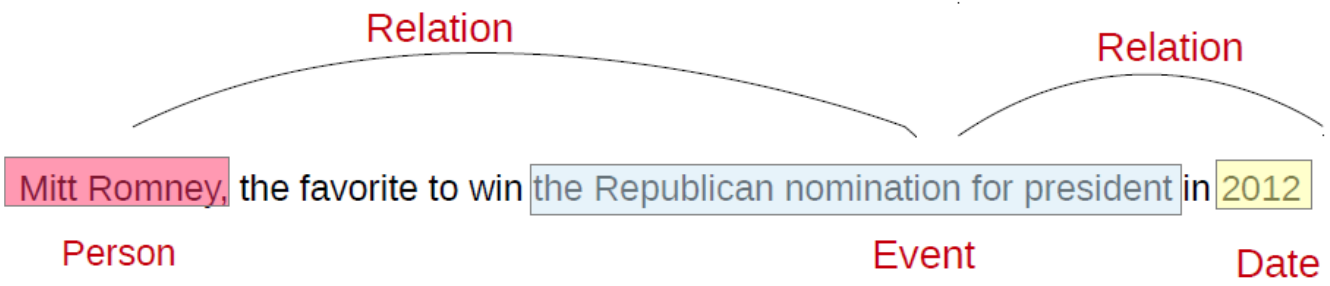
- Entities and terms may be described in different ways but refer to the same thing. We call this **co-reference**.

The GOP tweeted that they had knocked on 75,000 doors in Ohio the day

Organisation Location

What is Event Recognition?

- An event is an action or situation relevant to the domain expressed by some relation between entities or terms
- It is always grounded in time, e.g. the performance of a band, an election, the death of a person



Why are Entities and Events Useful?

- They can help answer the "Big 5" journalism questions (who, what, when, where, why)
- They can be used to categorize the texts in different ways
  - look at all texts about Obama
- They can be used as targets for opinion mining
  - find out what people think about President Obama
- When linked to an ontology and/or combined with other information, they can be used for reasoning about things not explicit in the text
  - seeing how opinions about different American presidents have changed over the years

Approaches to Information Extraction

- **Knowledge Engineering**
  - rule based
  - developed by experienced language engineers
  - make use of human intuition
  - easier to understand results
  - development could be very time consuming
  - some changes may be hard to accommodate
- **Learning Systems**
  - use statistics or other machine learning
  - developers do not need LE expertise
  - requires large amounts of annotated training data
  - some changes may require re-annotation of the entire training corpus

Typical components

- **Language Resources** (LRs), e.g. lexicons, corpora, ontologies
- **Processing Resources** (PRs), e.g. parsers, generators, taggers
- **Visual Resources** (VRs), i.e. visualization and editing components

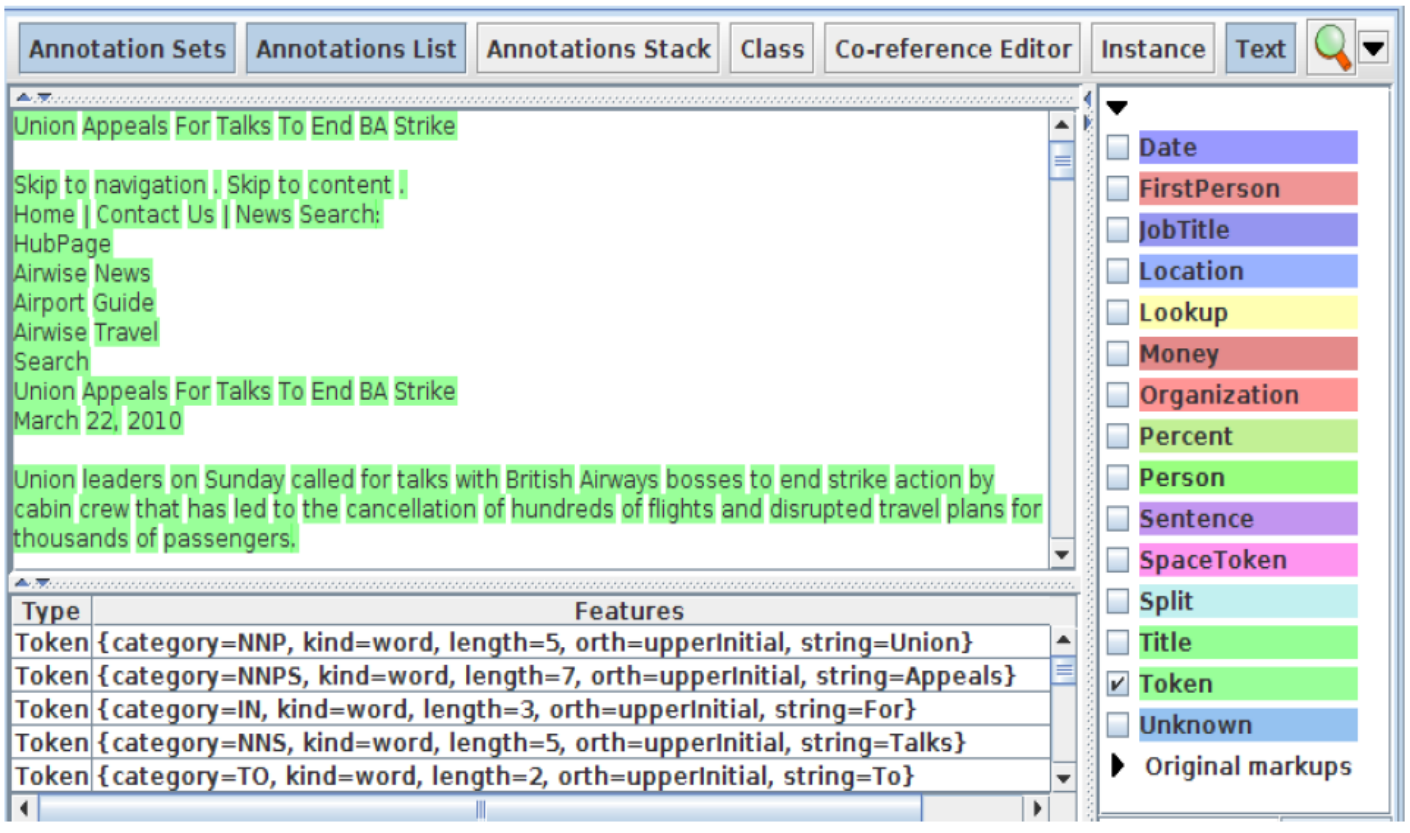
Information extraction step-by-step

- Rule-based IE system
  - Uses the language engineering approach
  - Uses a finite-state pattern-action rule language
  - Contains a reusable and easily extendable set of components:
    - generic preprocessing components for tokenization, sentence splitting etc.
    - components for performing NE on general open domain text

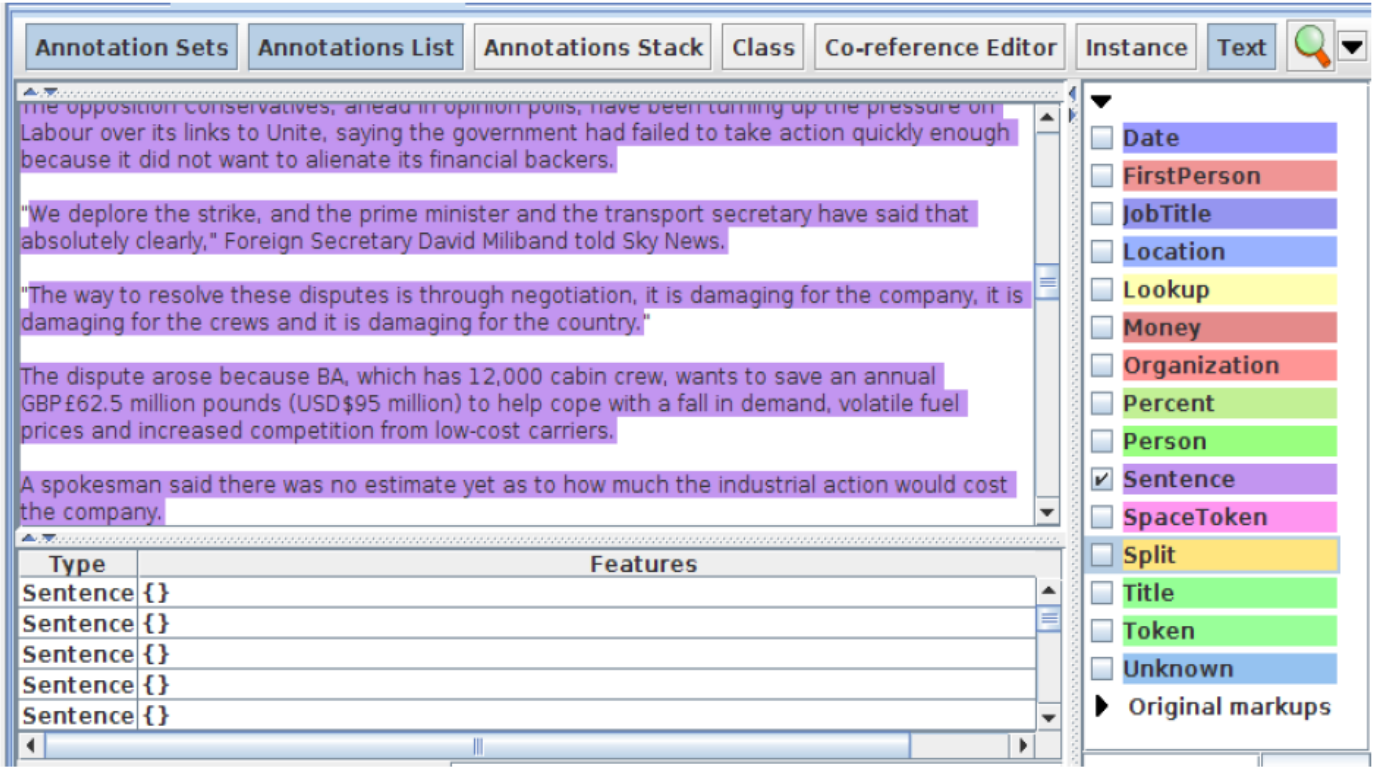
Processing resources

- Tokenizer
- Sentence Splitter
- Part Of Speech tagger
- Gazetteers
- Named entity tagger
- Orthomatcher (orthographic co-reference)

Example – GATE (tokenizer)



Example – GATE (sentence splitter)



Named Entity Recognition

- Gazetteers can be used to find terms that suggest entities
- However, the entries can often be ambiguous
  - “May Jones” vs “May 2010” vs “May I be excused?”
  - “Mr Parkinson” vs “Parkinson's Disease”
  - “General Motors” vs. “General Smith”
- Handcrafted grammars are used to define patterns over the Lookups and other annotations
- These patterns can help disambiguate, and they can combine different annotations, e.g. Dates can be comprised of day + number + month

Named Entity Grammars

- Hand-coded rules applied to annotations to identify NEs
- Phases run sequentially and constitute a cascade of Finite States over annotations
- Annotations from format analysis, tokenizer. splitter, POS tagger, morphological analysis, gazetteer etc.
- Because phases are sequential, annotations can be built up over a period of phases, as new information is gleaned
- Standard named entities: persons, locations, organisations, dates, addresses, money
- Basic NE grammars can be adapted for new applications, domains and languages

Example

```
Rule: PersonName
(
 {Lookup.majorType == firstname}
 {Token.category == NNP}
):tag
-->
:tag.Person = {kind = fullName}
```

Look for an entry in a gazetter list of first names

followed by a proper noun

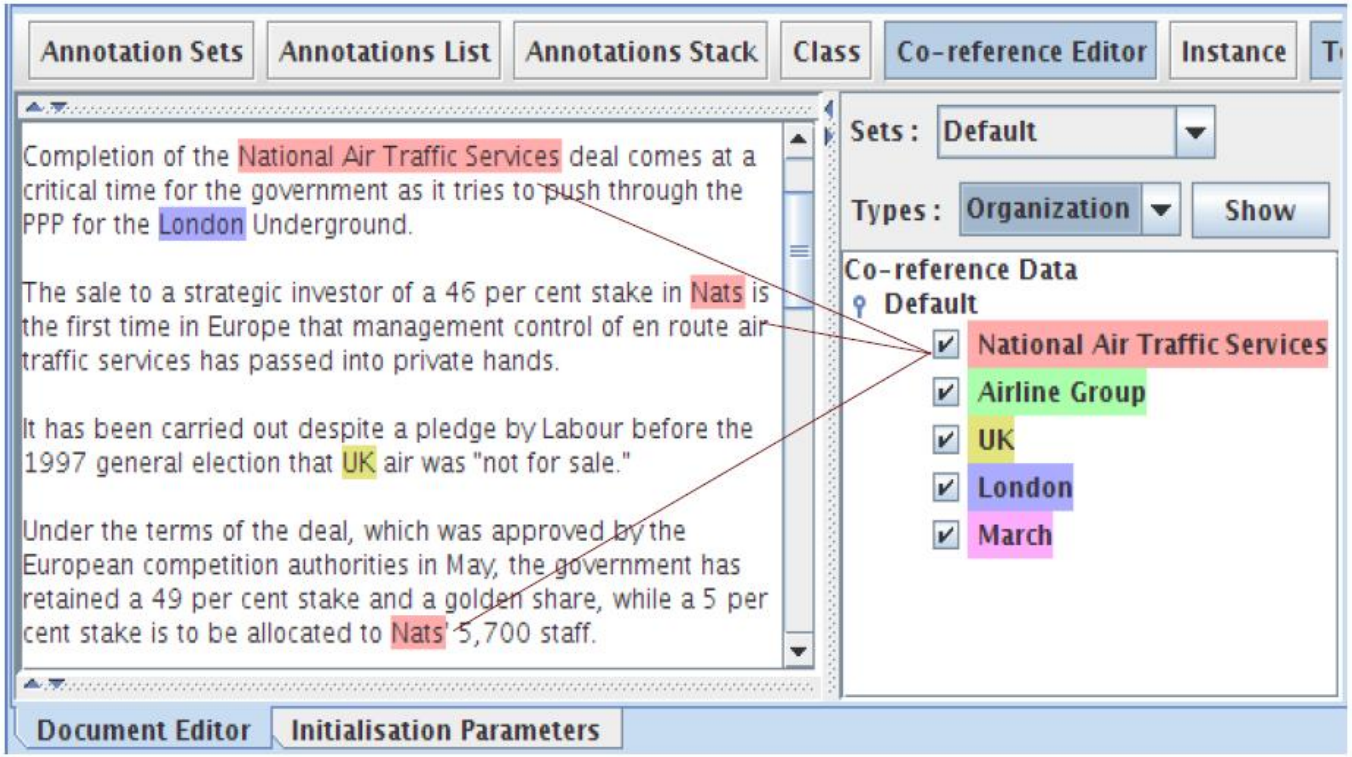
create an annotation of type “Person”

give the annotation the feature kind and value “fullName”

Using co-reference

- Different expressions may refer to the same entity
- Orthographic co-reference module (orthomatcher) matches proper names and their variants in a document
- [Mr Smith] and [John Smith] will be matched as the same person
- [International Business Machines Ltd.] will match [IBM]

Example


Annotation Sets
Annotations List
Annotations Stack
Class
Co-reference Editor
Instance
Completion of the National Air Traffic Services deal comes at a critical time for the government as it tries to push through the PPP for the London Underground.
The sale to a strategic investor of a 46 per cent stake in Nats is the first time in Europe that management control of en route air traffic services has passed into private hands.
It has been carried out despite a pledge by Labour before the 1997 general election that UK air was "not for sale."
Under the terms of the deal, which was approved by the European competition authorities in May, the government has retained a 49 per cent stake and a golden share, while a 5 per cent stake is to be allocated to Nats' 5,700 staff.
Sets : Default
Types : Organization
Show
Co-reference Data
Default
National Air Traffic Services
Airline Group
UK
London
March
Document Editor
Initialisation Parameters

Other NLP tools

- Stanford Tools
- UIMA
- OpenCalais
- Lingpipe
- OpenNLP

Analyzing Social Media

- Grundman:politics makes #climatechange scientific issue,people don't like knowitall rational voice tellin em wat 2do
- @adambation Try reading this article , it looks like it would be really helpful and not obvious at all. http://t.co/mo3vODoX
- Want to solve the problem of #ClimateChange? Just #vote for a #politician! Poof! Problem gone! #sarcasm #TVP #99%
- Human Caused #ClimateChange is a Monumental Scam! http://www.youtube.com/watch?v=LiX792kNQeE … Lying to us like MOFO's Tax The Air We Breath!

Challenges for NLP

- Noisy language: unusual punctuation, capitalization, spelling, use of slang, sarcasm etc.
- Terse nature of micro-posts such as tweets
- Use of hashtags, @mentions etc. causes problems for tokenisation #thisistricky
- Lack of context gives rise to ambiguities
- NER performs poorly on micro-posts, mainly because of linguistic pre-processing failure
  - Performance of standard IE tools decreases from ~90% to ~40% when run on tweets rather than news articles

Persons in news articles

Left context	Match	Right context
in dicated Atef, including	Douglas Feith	, the United States defence
, the group that killed	President Sadat	in 1981 as retribution for
. The current leader,	President Olusegun Obasanjo	, who recently came to
Kuwait, whose information minister	Sheikh Ahmed Fahed al-Sabah	met editors of local newspapers
The current defence minister,	Theophilus Danjuma	, has also been threatened
The three right-wing MPs,	Andrew Rosindell	(Romford), Andrew
Late on Wednesday night,	Justice Oputa	, who chairs the commission
the militarily-manoeuvred civilian elec...	President Obasanjo	in 1999 and is widely
after the mysterious death of	General Sani Abacha	in 1998.
have learnt that one of	Bin Laden	's closest and most senior
evidence confirms the involvement of	Osama bin Laden	in those attacks."
. He is one of	Bin Laden	's two most senior associates
for future civilian office.	General Buhari	took power in a 1983
$5m price on	Atef	's head and prosecutors have
Afghanistan. He was once	Bin Laden	's chief media adviser and
thinking in the Tory party	Iain Duncan Smith	has ordered three Tory MPs
club and the party,	David Maclean	, the Tory Chief Whip
Centre and the Pentagon.	Mohammed Atef	, who is thought to
are still very powerful.	General Babangida	supported the militarily-manoeuvred civ
sexual orientation or religion.	Mr Duncan Smith	's purge of the Monday
," he said.	Atef	, who is reported variously
of the late singer,	Fela Kuti	□ which took place while
field in Penn sylvania.	President Bush	included Atef in an order
. It is believed that	Mr Duncan Smith	intended to launch his crackdown

Persons in tweets

Left context	Match	Right context
i was your age ,	spencer	from iCarly was Crazy Steve
iCarly was Crazy Steve ,	Carly	was Megan and Josh was
bath , shut up ,	sam	's coming tomorrow and steve
. All are welcome ,	joe	included
. All are welcome ,	joe	included
teachers , chinese takeaways ,	gatt holly	, phil collins , the
takeaways , gatt holly ,	phil collins	, the skin of a
@GdnPolitics : RT AlJahom :	Blair	: " I'm gonna
Empls of the Month :	Deborah L	#Speech #Pathologist-Childrens
be the next Pope "	Brown	: " I won't
(via POPSUGAR)	Sarah Jessica Parker	and Gwen Stefani Wrap Up
and is smexy !!;)And	Chelsea Handler	is hilarious ! Finally got
him befnrjustthen about	kenny	signing his book but it
three kinds of reactions after	Ayodhya	verdict .
, Carly was Megan and	Josh	was fat . #damnteenquotes
sam 's coming tomorrow and	steve	and tanya will be round
coming tomorrow and steve and	tanya	will be round at 10am
photo caption contest- Nadal and	Novak	in the tub http://ow.ly/2G3Jh
) Sarah Jessica Parker and	Gwen Stefani	Wrap Up Another Successful New
#Pathologist-Childrens Rehab and	Patricia M	#Referral/#Auth #
Just casually stalking Cheryl AND	Dermot	tomorrow .... NO BIGGIE
did tweet him befnr	justthen	about kenny signing his book
Test : We just congratulated	Lindsay	an hour ago on h
the funny photo caption contest-	Nadal	and Novak in the tub

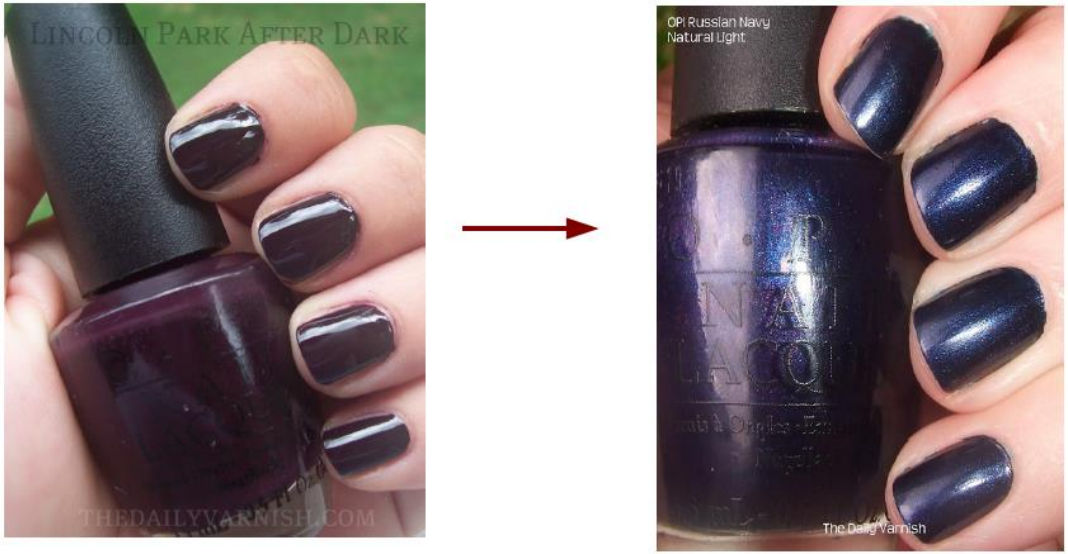
Lack of context causes ambiguity

- Branching out from Lincoln park after dark ... Hello Russian Navy, it's like the same thing but with glitter!

???

Getting the NEs right is crucial

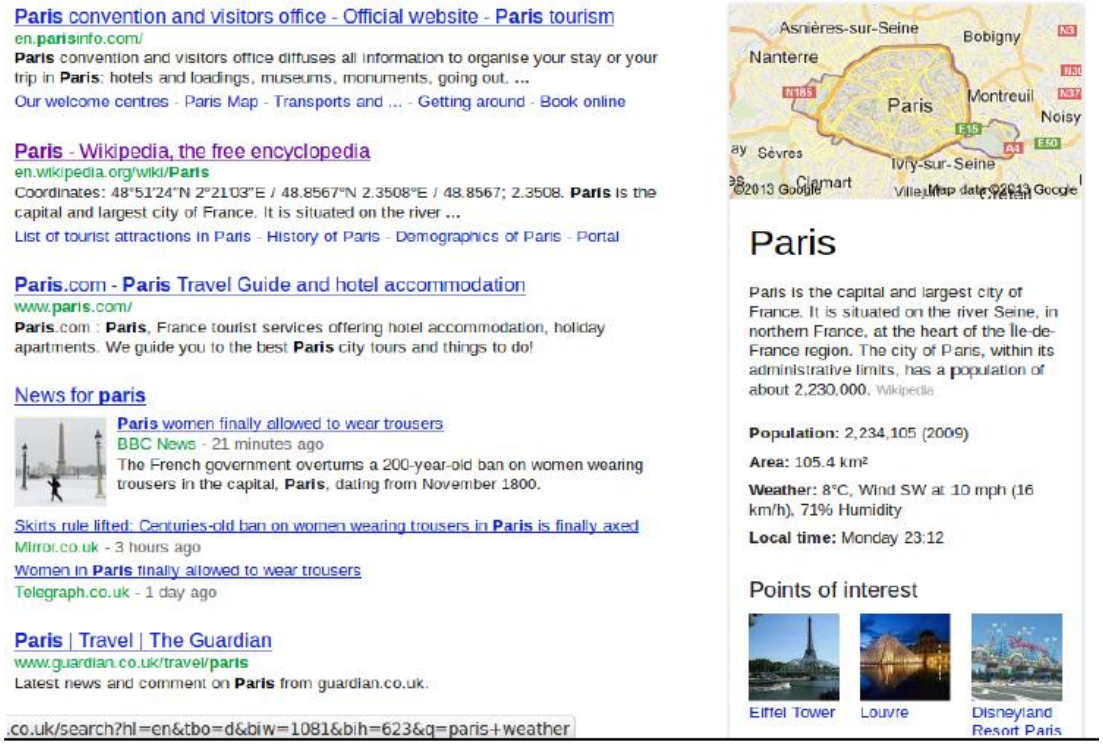
- Branching out from Lincoln park after dark ... Hello Russian Navy, it's like the same thing but with glitter!



How do we deal with this kind of text?

- Typical NLP pipeline means that degraded performance has a knock-on effect along the chain
- Short sentences confuse language identification tools
- Linguistic processing tools have to be adapted to the domain
- Retraining individual components on large volumes of data
- Adaptation of techniques from e.g. SMS analysis
- Development of new Twitter-specific tools (e.g. GATE's TwitIE)
- But....lack of standards, easily accessible data, common evaluation etc. are holding back development

Text Analytics for Semantic Search



Searching for Things, not Strings

- 500 million entities that Google "knows" about
- Used to provide more accurate search results
- Summaries of information about the entity being searched

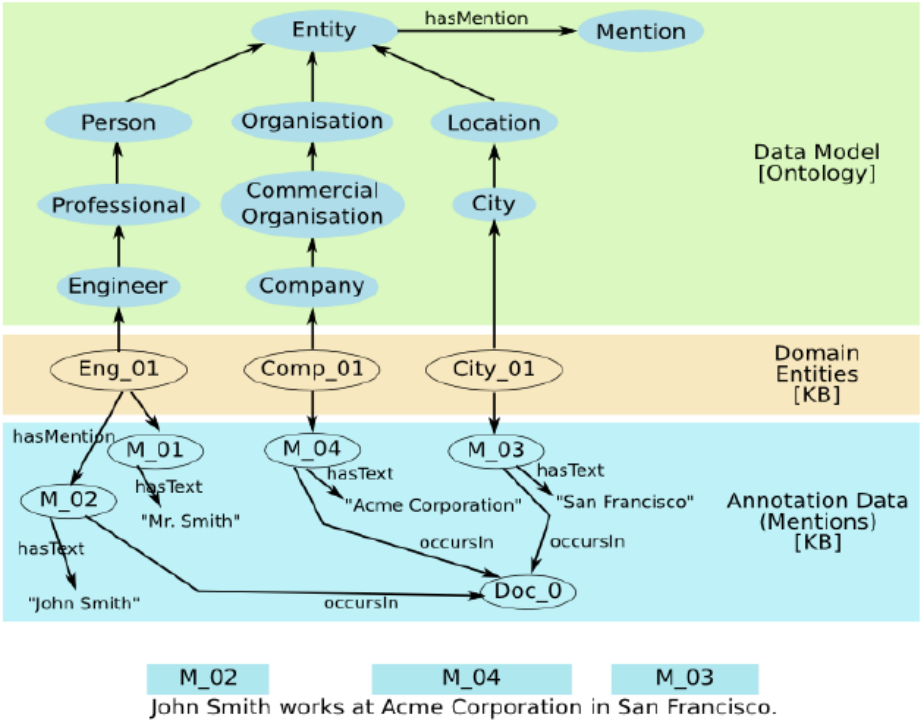
Semantic Enrichment

- Textual mentions aren't actually that useful in isolation
  - knowing that something is a “Person" isn't very helpful
  - knowing which Person the mention refers to can be very useful
- Disambiguating mentions against an ontology provides extra context
- This is where **semantic enrichment** comes in
- The end product is a set of textual mentions linked to an ontology, otherwise known as **semantic annotations**
- Annotations on their own can be useful but they can also
  - be used to generate corpus level statistics
  - be used for further ontology population
  - form the basis of summaries
  - be indexed to provide semantic search
- What about **Automatic Semantic Enrichment**? - Automatically extend article metadata to improve search quality

Semantic Annotation

- **Semantic annotation**: rather than just annotating the word "Cambridge" as a location, link it to an ontology instance
  - Differentiate between *Cambridge, UK* and *Cambridge, Mass*
- **Semantic search via reasoning**
  - So we can infer that this document mentions a city in Europe
  - Ontologies tell us that this particular Cambridge is part of the country called the UK, which is part of the continent Europe
- **Knowledge source**
  - If I want to annotate *strikes* in baseball reports, the ontology will tell me that a *strike* involves a *batter* who is a *person*
  - In the text "BA went on strike", using the knowledge that BA is a company and not a person, the IE system can conclude that this is not the kind of strike it is interested in

Semantic Annotation



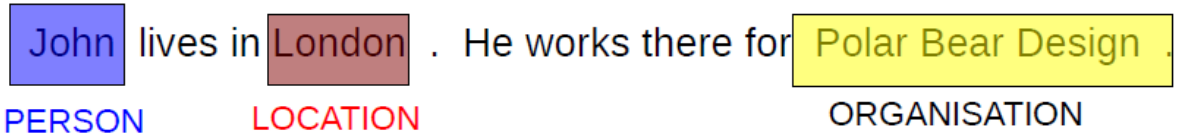
What is an ontology?

- Set of concepts (instances and classes)
- Relationships between them (is-a, part-of, located-in)
- Multiple inheritance
- Classes can have more than one parent
- Instances can have more than one class
- Ontologies are graphs, not trees

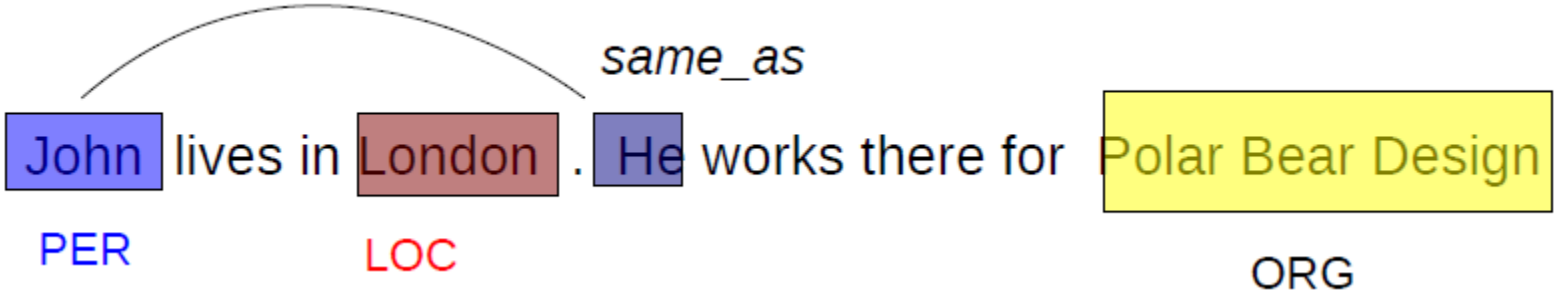
Information Extraction for the Semantic Web

- Traditional IE is based on a flat structure, e.g. recognizing Person, Location, Organisation, Date, Time etc.
- For the Semantic Web, we need information in a hierarchical structure
- Idea is that we attach semantic metadata to the documents, pointing to concepts in an ontology
- Information can be exported as an ontology annotated with instances, or as text annotated with links to the ontology

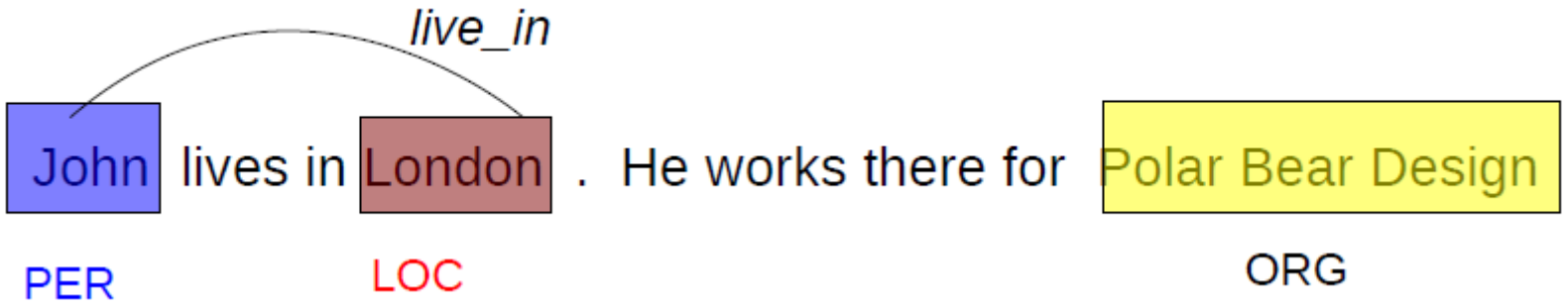
Traditional NE recognition



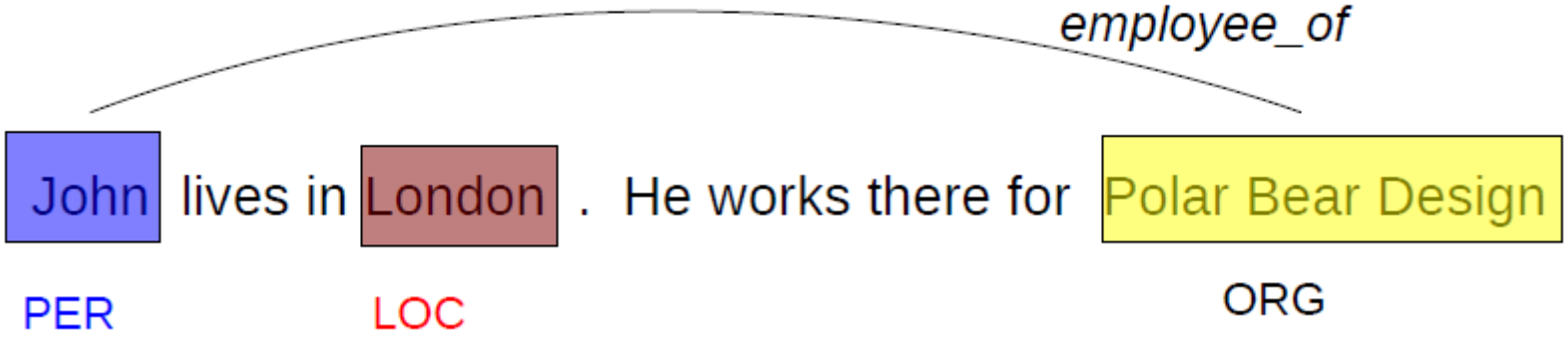
Co-reference



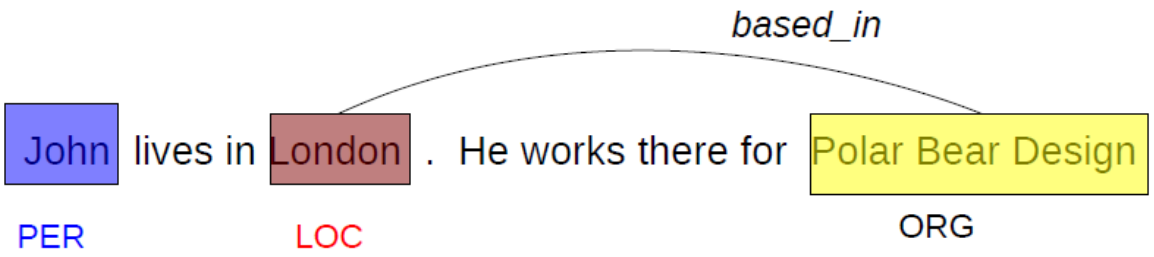
Relations


live_in
John
lives in
London
. He works there for
Polar Bear Design
PER
LOC
ORG

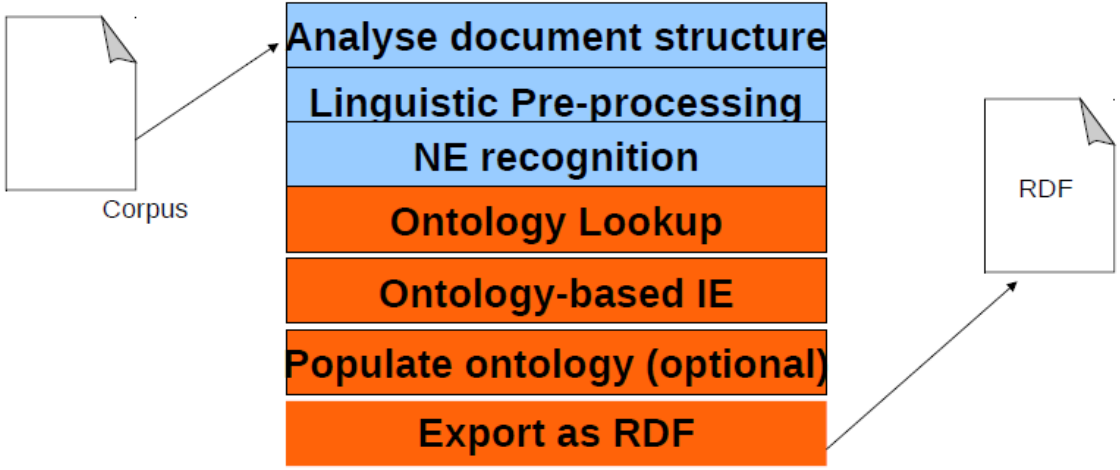
Relations


employee_of
John
lives in
London
. He works there for
Polar Bear Design
PER
LOC
ORG

Relations


based_in
John
lives in
London
. He works there for
Polar Bear Design
PER
LOC
ORG

Typical semantic annotation pipeline



Automatic Semantic Annotation

- Locations (linked to DBpedia and GeoNames)
  - Annotate the place name itself (e.g. Norwich) with the corresponding DBpedia and GeoNames URIs
  - Also use knowledge of the implied reference to the levels 1, 2, and 3 sub-divisions from the Nomenclature of Territorial Units for Statistics (NUTS)
  - For Norwich, these are East of England (UKH – level 1), East Anglia (UKH1 – level 2), and Norfolk (UKH13 – level 3)
  - Similarly use knowledge to retrieve nearby places

Automatic Semantic Annotation

- Organizations (linked to DBpedia)
  - Names of companies, government organizations, committees, agencies, universities, and other organizations
- Dates
  - Absolute (e.g. 31/03/2012) and relative (yesterday)
- Measurements and Percentages
  - e.g. 8,596 km2 , 1 km, one fifth, 10%

Why is Machine Translation Hard?

- Human languages are:
  - Elegant
  - Efficient
  - Flexible
  - Complex
- One word/sentence may mean many things
- Many ways of saying the same thing
- Meaning depends on context
- Literal and figurative language (metaphor)
- Language and culture (different ways of conceptualizing the same thing)
- Word order
- Morphology

Language and Translation is Complex

- Language/translation is complex
- We cannot compute it exactly
- We tried: rule-based MT
- What do we do?
- Machine Learning
  - Learns from data ⇒ data is all important
  - Approximate solution ⇒ not perfect, needs help
    - human professional translators
    - Post-editing
    - Automated Translation ≠ Automatic

How does modern MT work?

- Statistical MT learns from data
- Two kinds of data:
  - Human translations
  - Text in the target language
- The more data the better!
- Also: the right kind of data!

GERMAN

Einleitung

I. Von dem Unterschiede der reinen und empirischen Erkenntnis

Daß alle unsere Erkenntnis mit der Erfahrung anfange, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verstandestätigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.

ENGLISH

Introduction

I. Of the difference between Pure and Empirical Knowledge

That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.

FRENCH

Introduction

I. De la différence de la connaissance pure et de la connaissance empirique.

Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaître pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.

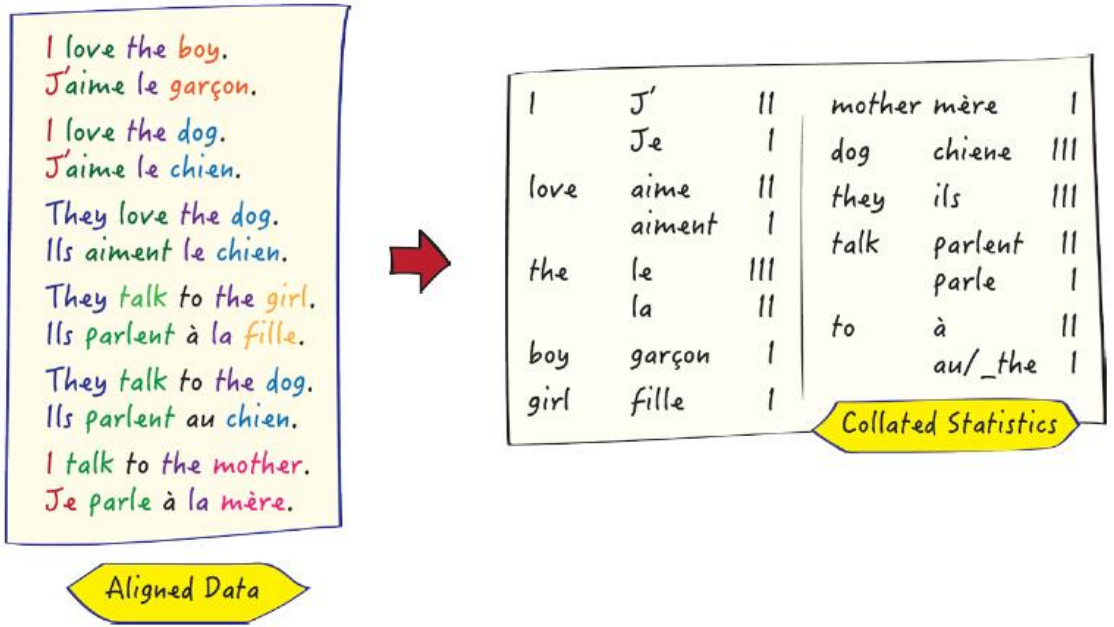
What can/do we learn from data?

- Which sentences translate as which: sentence alignment
- Which words translate as which: word alignment + translation probabilities
- What is good target language like: language model

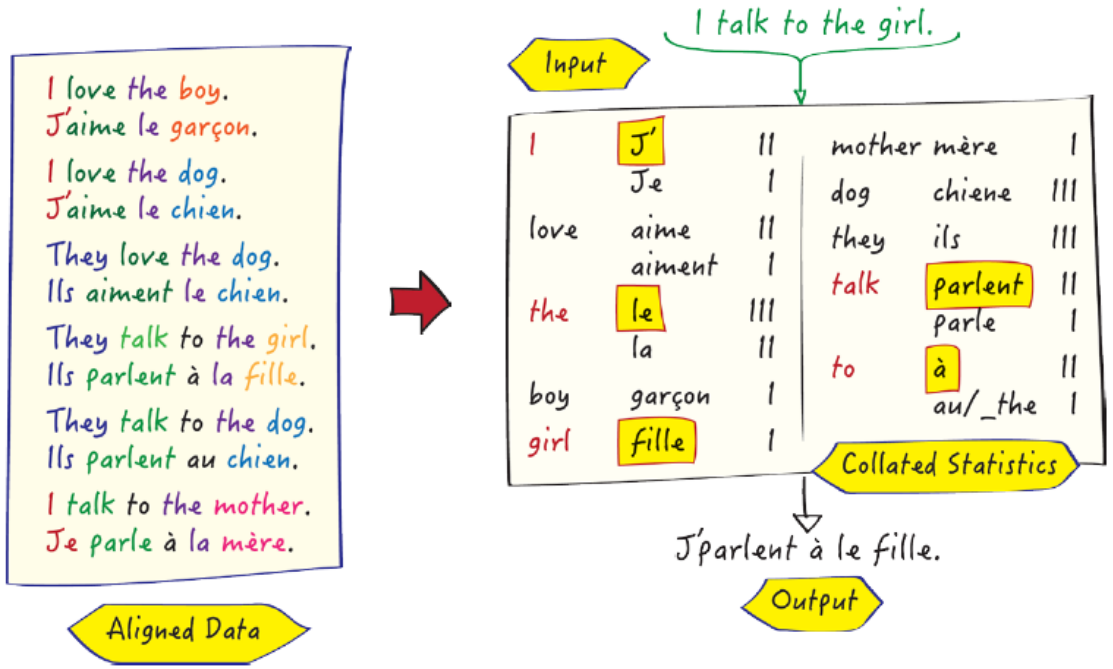
What can/do we learn from data?

- Sentence alignment
- Word alignment
- Given word aligned translation data, can we learn a translation dictionary? - Yes, really easy …

Statistical ML



Statistical ML



Statistical ML

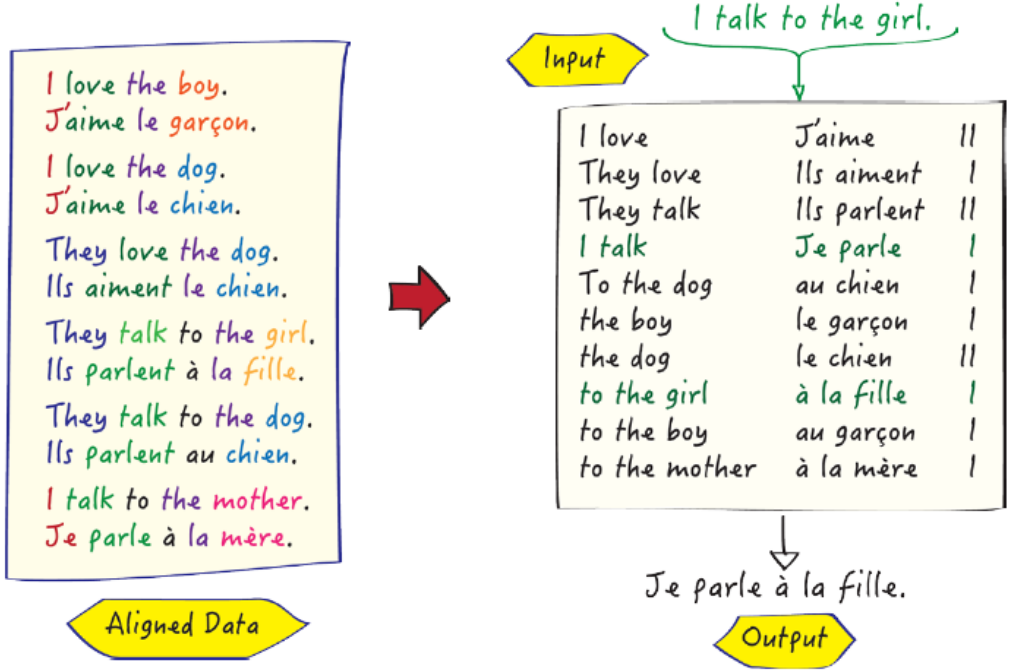
I	talk	to	the	girl	
J'	parlent		au	le	fille
2/3	2/3		2/3	3/5	1/1
Je	parle	à	la	fille	
1/3	1/3	1/3	2/5	1/1	

How to choose?

Phrase-Based SMT

- So far: translating single words
- Loses context: such as agreement (le fille…) etc.
- To some extent “repaired” by language model
- A better model:
  - Not just translations of single words
  - But also phrase translations:
    - the girl : la fille
    - to the girl : a la fille
    - I talk : Je parle

Phrase Based -Statistical Machine Translation



Phrase Based -Statistical Machine Translation

- Much better than word-based SMT!
- Standard technology: Google, Microsoft, Baidu, Global Localization & Translation Industry
- Moses Open Source PB-SMT
- Most widely used SMT system

Machine Translation and Data

- Statistical Machine Translation is all about data
- SMT learns how to translate from data
- Data
  - translations (bilingual data)
  - monolingual data (target language text)
  - dictionaries, terminology, ontologies, named entities
- Like people SMT is good at what it has learned

Conclusions

- Text mining is a very useful pre-requisite for doing all kinds of more interesting things, like semantic search
- Semantic web search allows you to do much more interesting kinds of search than a standard text-based search
- Text mining is hard, so it won't always be correct, however
- This is especially true on lower quality text such as social media
- On the other hand, social media has some of the most interesting material to look at
- Still plenty of work to be done, but there are lots of tools that you can use now and get useful results

Conclusions

- Hand-crafted rules are brittle
- Distributional information can often be a proxy for rich structure
- Large-scale annotations of rich structure can produce better models (with machine learning) than hand-built methods
- Large-scale annotation is VERY expensive and time consuming
- On going research:
  - Big data including stream of text / Deep Learning
  - Cross-lingual, cross-modal, cross-domain