

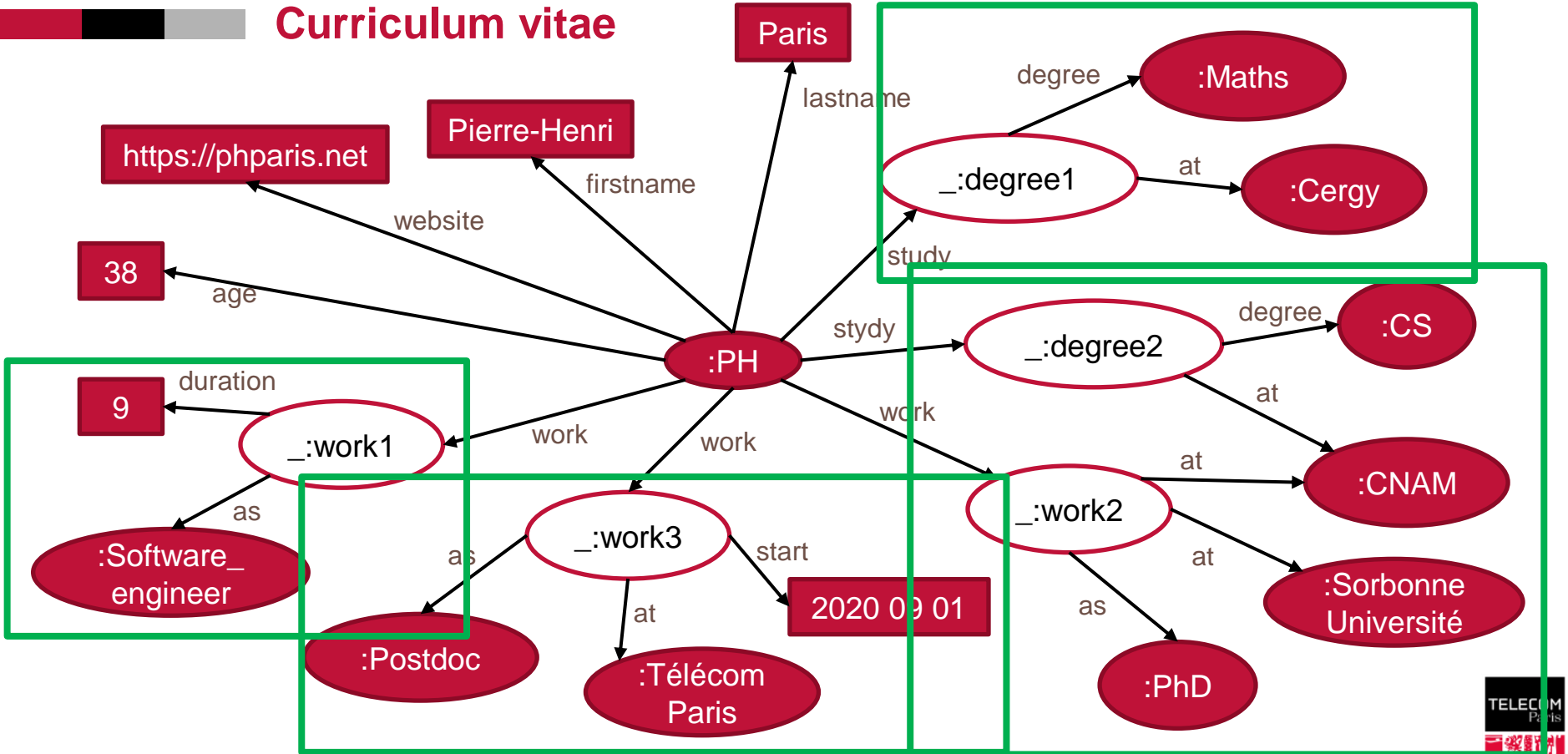


Vagueness, non-named entities, and knowledge representation

Pierre-Henri PARIS



Curriculum vitae





Outline

- 1. The Vagueness of Vagueness in Noun Phrases**
- 2. Non-Named Entities – The Silent Majority**
- 3. Knowledge Representation and Reasoning**

The Vagueness Problems

“In the **early 20th century**, **German researchers** made **advances** in linking smoking to **health harms**.”

- This is a factual sentence, and thus interesting for information extraction.
However, when is the early 20th century? Who are the German researchers? How many of them? What advances? What are the health harms?
- How to detect, extract, model and reason with vagueness?
- Vagueness has not been sufficiently studied

Types of Vagueness

■ Scalar vagueness

- expressions that can be interpreted as a scalar
- *“a rich person”, “long-running debate”, “early history”*

■ Quantitative vagueness

- Expressions that refer to an unspecified portion of an entity or to a set of entities whose number is not identified
- *“a part of the film”, “many scientists”, “scientists”*

■ Subjective vagueness

- Expressions that can apply to a certain degree, and where there is no consensus on how to measure this degree
- *“a beautiful man”, “high ethical standards”, “tensions”*



Vagueness often depends on its context, i.e. the sentence in which it is used:
“Computer scientists believe that AES encryption is safe” VS “Computer scientists did not exist in the Middle Ages”

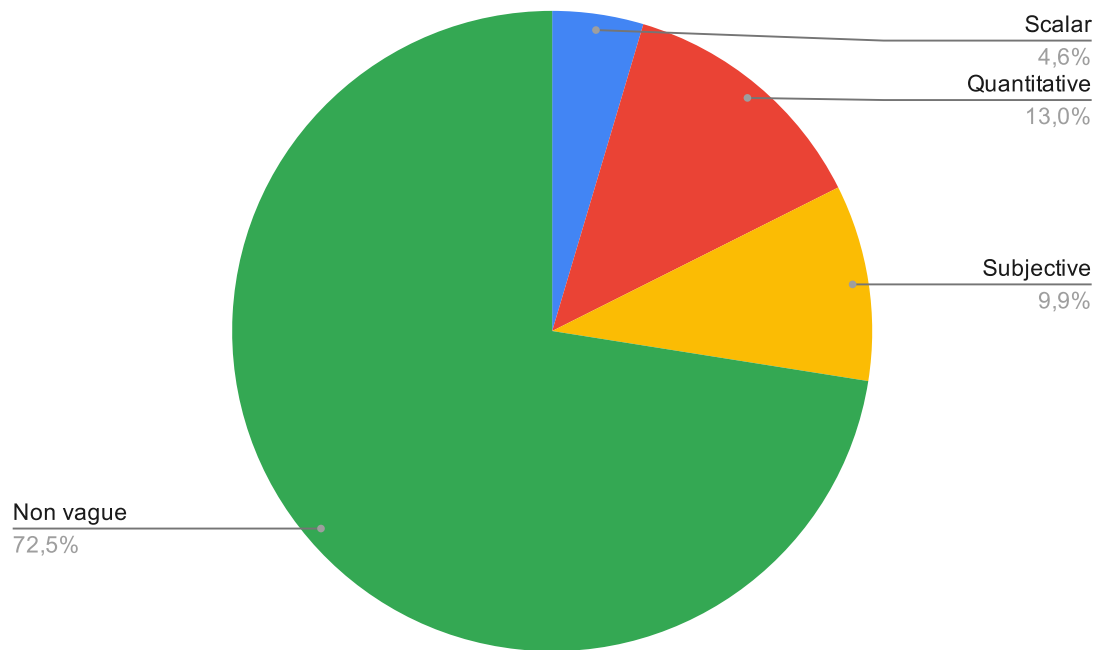
Case Study: Wikipedia

- Wikipedia featured articles (high quality articles)
- Wikipedia is a widely used standard reference, in both research and industry applications
- 1 article abstract for each of the 30 featured article topics

⇒ our study constitutes a probable lower bound on the frequency of vagueness



Case Study: Wikipedia



⇒ 1/4 of noun phrases are vague!

Existing Works: Reasoning with Vagueness

■ Logical Approaches

- Generalized quantifiers for quantitative vagueness
- Fuzzy Logic for scalar vagueness
- Fuzzy Logic + *ALC* Description Logic for scalar vagueness

■ Probabilistic Approaches

- Bayesian approach for scalar vagueness in adjectives
- Word co-occurrence for scalar vagueness in modifiers

⇒ Quantitative and scalar vagueness could be addressed if different approaches are used together, but subjective vagueness remains out of reach

Existing Work: Detecting Vagueness

■ Noun phrase level:

- **Vague definitions in ontologies:**
certain words indicate vagueness
- **Requirement specifications in the railway domain:**
difficult to define phrases that identify vagueness with a high accuracy
- **Privacy policies:**
collection of vague words to train a classifier, there is no exhaustive list of vague words

■ Sentences and noun phrases levels:

- **Crowdsourcing annotations in privacy policies to train two classifiers:**
vagueness is context-dependent

⇒ **Classifiers can detect vagueness on specific domains, could they detect vagueness in general?**



Conclusion

- **Vagueness is a common phenomenon in noun phrases, but current approaches do not address it comprehensively.**
- **Syrine El Aoud, Pierre-Henri Paris, and Fabian M. Suchanek. “*The Vagueness of Vagueness in Noun Phrase*”. In: ACL 2021 (under review)**



Outline

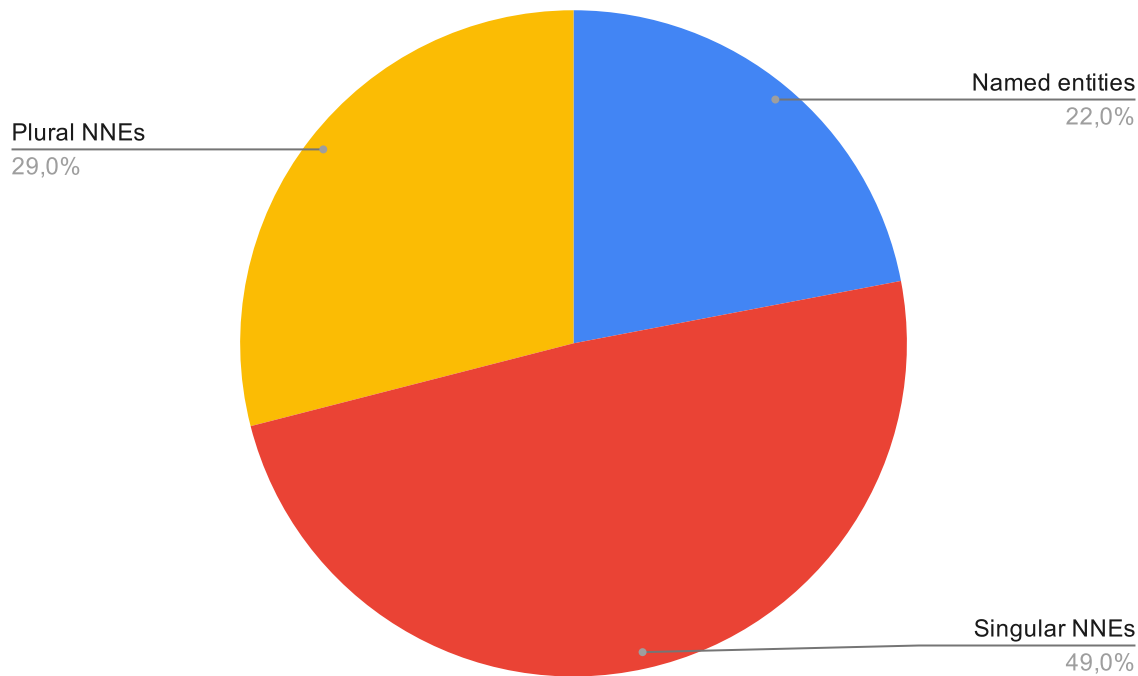
1. The Vagueness of Vagueness in Noun Phrases
2. Non-Named Entities – The Silent Majority
3. Knowledge Representation and Reasoning

Noun phrases

“The Arab Spring resulted in contentious battle between a consolidation of power by religious elites and the growing support for democracy.”

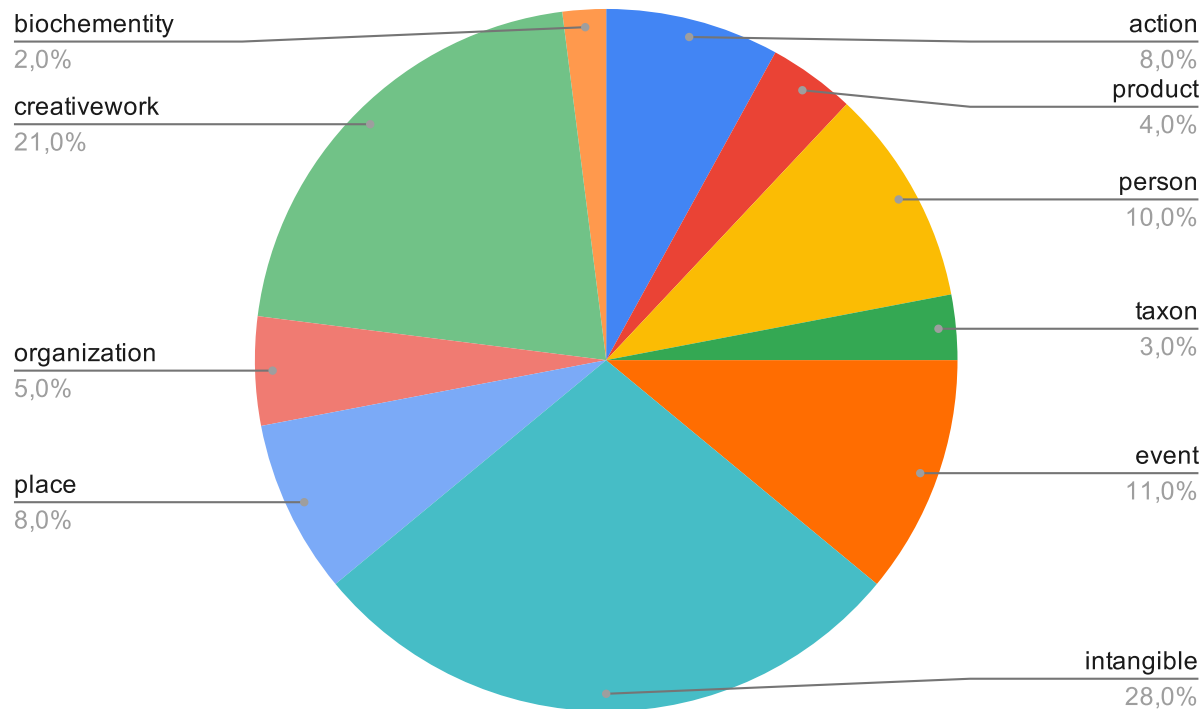
- This is a factual sentence, and thus interesting for information extraction.
- However, it contains only **a single named entity**. Many fact-extraction techniques will simply not consider **the other non-named entities**.

Manual Study of Non-Named Entities in Wikipedia

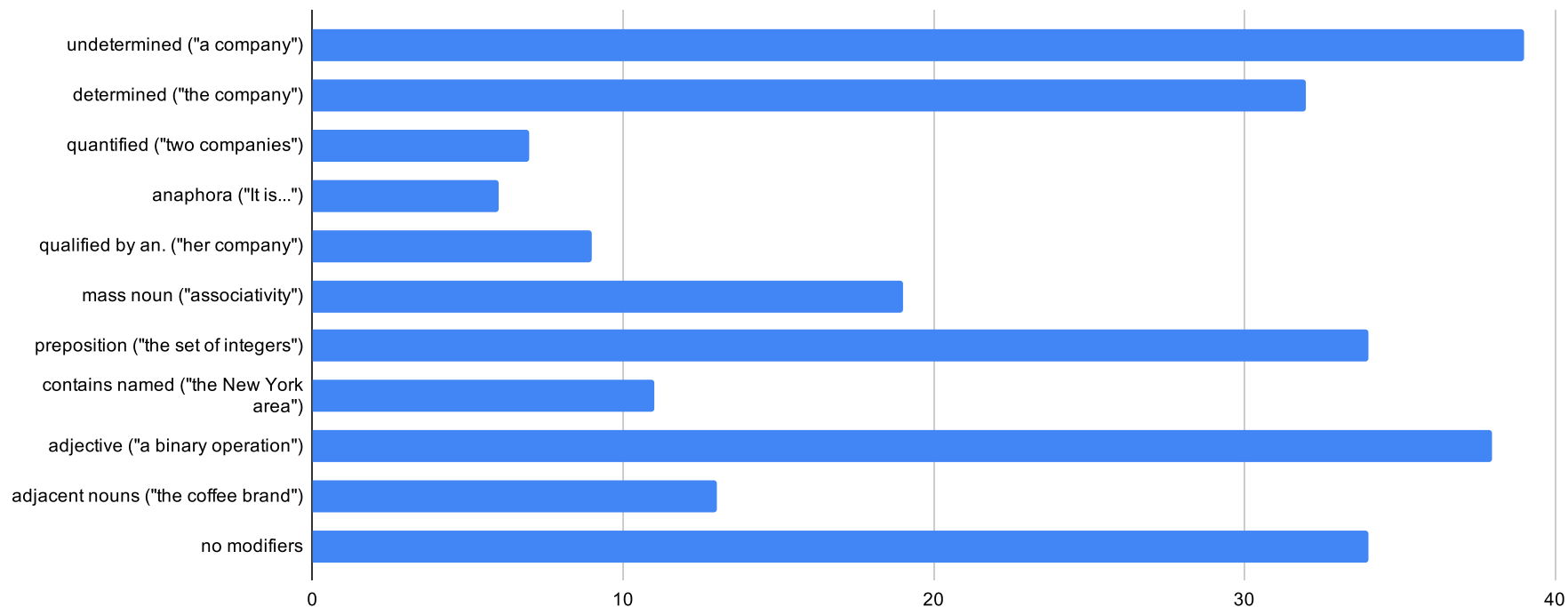


1 article abstract for each of the 30 featured Wikipedia article categories

Non-named entities by Yago class



Non-named entities by nature and modifiers



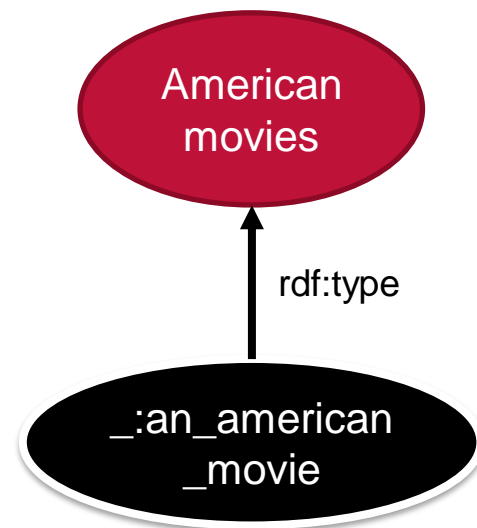
Adding Non-named Entities to RDFS Knowledge Bases

■ Extracting non-named entities

- Possible with Semantic parsers, AMR, Open Information Extraction systems...
- Named entities are mapped to the knowledge base

■ Modeling non-named entities

- Replace anaphoras and determined noun phrases with their referent
- Simple nested noun phrases are linked together
- Use of ad-hoc classes for plural noun phrases
- Make singular noun phrases anonymous instances of ad-hoc classes
- Replication of numbered noun phrases
- Mass nouns became classes



The Gap that remains

- **Knowledge representation:**
 - Some knowledge bases contain mainly classes, others mix classes and instances, and others duplicate them
 - Plural noun phrases like « *hundreds of soldiers* »
 - Vague noun phrases like « *large-scale settlement* »
 - Nested entities like « *the growing support for democracy in many Muslim-majority states* »
- **Canonicalization: Similar non-named entities in different contexts like two different rises of the same stock market**
- **Facts: Comparatives, superlatives, and temporal comparisons like « *A rose is more beautiful than a daisy* »**



Outline

1. The Vagueness of Vagueness in Noun Phrases
2. Non-Named Entities – The Silent Majority
3. Knowledge Representation and Reasoning

What we want to model: High arity relationship



Arity is the number of arguments of a relationship



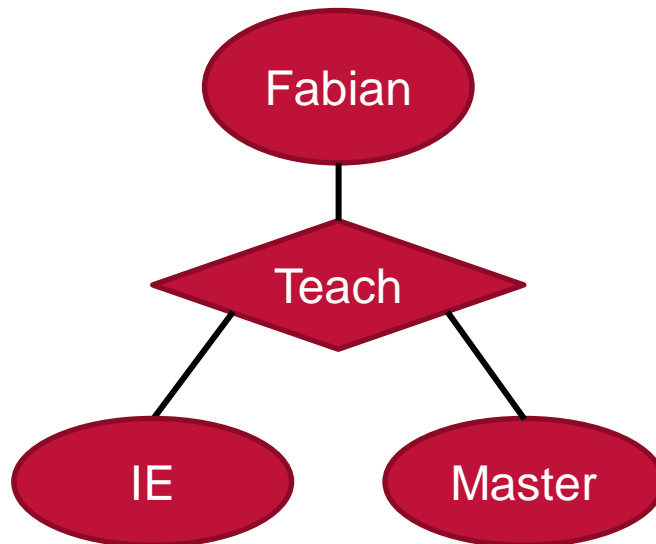
Because information is not always binary, we want to model relationship with arity > 2



"Fabian teaches Information Extraction in a specialized Master's program."



Reification and other constructs that tackle arbitrary relationships lead to undecidability of reasoning, if used in an unrestricted way.



What we want to model: Events



Events are objects in time



Many modeling/reasoning dimensions depend on events



What, when, topic, where and who?

“The Apollo 11 crew landed on the Moon on July 21, 1969.”



Evolution of long-lasting event? Use of a temporal logic or Event calculus?



What we want to model: Precedence



An event that happened before another one



Useful to model a narrative

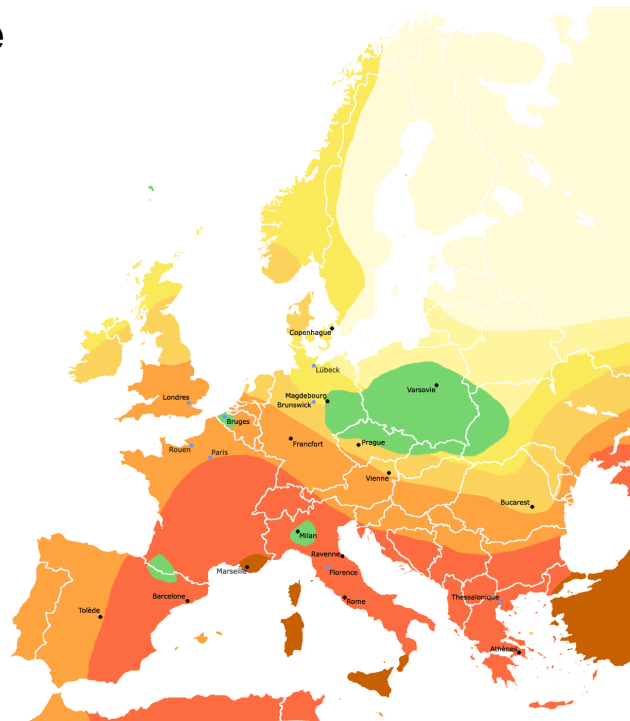


$\forall E_i: \text{before}(E_1, E_2), \text{before}(E_2, E_3)$
 $\Rightarrow \text{before}(E_1, E_3)$

*“WW1 happened before WW2 and
WW2 happened before the creation of
European Union” \Rightarrow “WW1 happened
before the creation of European Union”*



How to model temporal dimension?
Point-based or interval?



What we want to model: Causation



An event contributes to the production of another event



Causation is necessary to explain succession of events

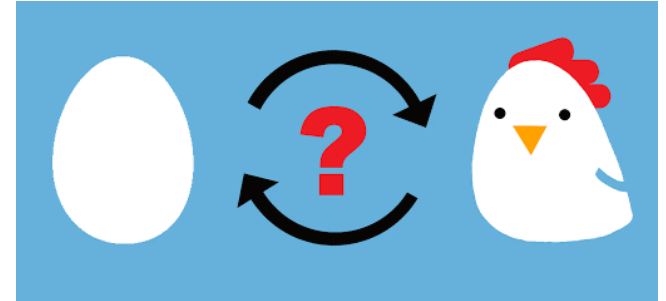


$\forall a, b, c: \text{causes}(a, b), \text{causes}(b, c) \Rightarrow \text{causes}(a, c)$
“The crisis of 2008 caused a recession. The number of homeless families is increasing due to the recession.”



Multiple implementations are possible

- **Necessary causes**
 - $\forall x, y: \text{necessary}(x, y) \text{ and } y \Rightarrow x$
 - “People who have a salary necessarily have a job.”
- **Sufficient causes**
 - $\forall x, y: \text{sufficient}(x, y) \text{ and } x \Rightarrow y$
 - “The recession caused the increase in the number of homeless families.”
- **Contributory causes**
 - $\forall x_i, y: \text{contributory}(x_1, \dots, x_n, y), x_1, \dots, x_n \Rightarrow y$
 - “Excessive risk-taking by banks contributed to the crisis of 2008.”



What we want to model: Negation



Modeling of false facts



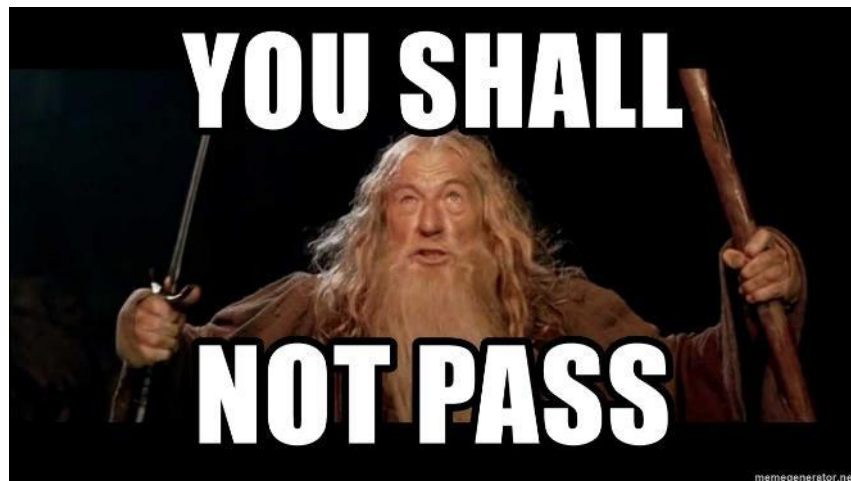
Most knowledge bases contain only positive facts, yet many interesting facts are negative



$\forall x: deathdate(x, d) \Rightarrow \neg alive(x)$
“Elvis died in 1977” \Rightarrow “Elvis is not alive”



The complexity increases as the arity of the negated relationship increases



What we want to model: Beliefs



A belief is an attitude that some proposition about the world is true



Allows to have contradicting statements in the same KB

“Flat-earthers believe the earth is flat, but the earth is almost a sphere.”



$\forall \phi: \forall x: \text{believes}(a, \phi),$
 $\text{type}(x, b) \Rightarrow \text{believes}(x, \phi)$

“All Republicans believe Trump and Trump believes he won the election.” \Rightarrow “Republicans believe they won the election.”



Introduction of the doxastic logic or the logic of context might lead to undecidability



What we want to model: Sentiments



Affective states toward something or someone



Sentiments are omnipresent in social interactions, especially between a company and its customers



$\forall x: \neg like(x, a) \Rightarrow \neg buy(x, i)$

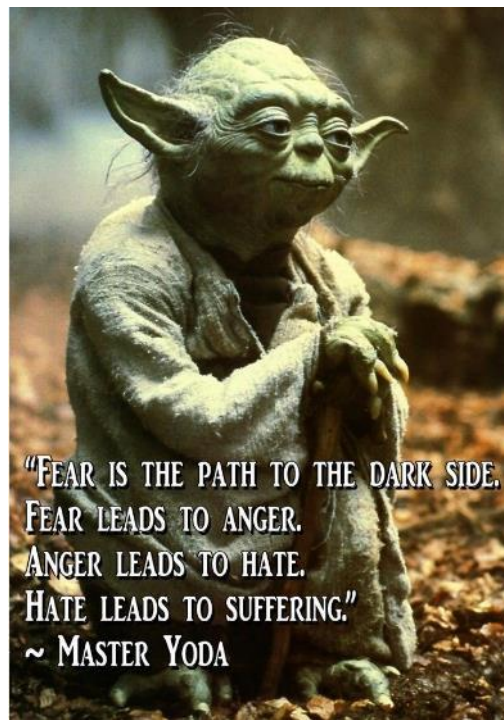
“John don’t like Apple” \Rightarrow “He won’t buy an iPhone”

$\forall x: like(x, h) \Rightarrow \psi(x)$

“John like to hunt” \Rightarrow “He won’t vote for a political ecologist party”



How to represent sentiments in knowledge bases?



What we want to model: Narratives



A series of logically and chronologically related events that describe something



Narratives are everywhere: Wikipedia articles, newspapers, books,...



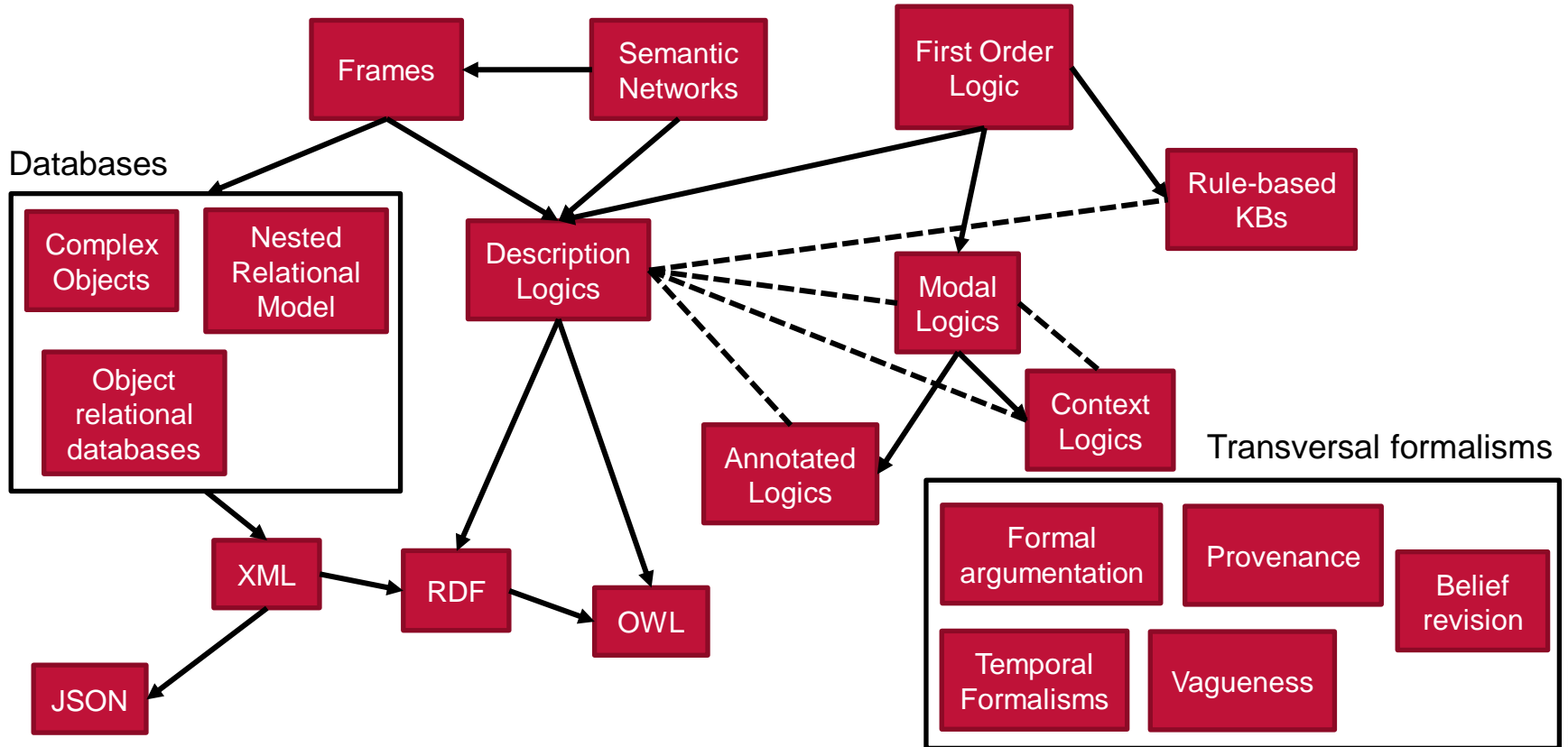
“Once upon a time...”



All of the preceding elements are necessary to properly represent and reason with narratives



Existing Formalisms: A bestiary





Existing Formalisms: Challenges

- **What do we want?**
To model and reason with high arity relationship, events, precedence, causation, negation, beliefs, sentiments and narratives
- **Some formalisms can represent what we need, but they are undecidable.**
- **We need a trade-off**

We need your use cases!

■ Some rule we have seen:

- $\forall E_i: \text{before}(E_1, E_2), \text{before}(E_2, E_3) \Rightarrow \text{before}(E_1, E_3)$
- $\forall a, b, c: \text{causes}(a, b), \text{causes}(b, c) \Rightarrow \text{causes}(a, c)$
- $\forall x, y: \text{necessary}(x, y) \text{ and } y \Rightarrow x$
- $\forall x, y: \text{sufficient}(x, y) \text{ and } x \Rightarrow y$
- $\forall x_i, y: \text{contributory}(x_1, \dots, x_n, y), x_1, \dots, x_n \Rightarrow y$
- $\forall x: \text{deathdate}(x, d) \Rightarrow \neg \text{alive}(x)$
- $\forall \phi: \forall x: \text{believes}(a, \phi), \text{type}(x, b) \Rightarrow \text{believes}(x, \phi)$
- $\forall x: \neg \text{like}(x, a) \Rightarrow \neg \text{buy}(x, i)$
- $\forall x: \text{like}(x, h) \Rightarrow \psi(x)$



■ What are your needs in terms of modeling and reasoning?

pierre-henri.paris@telecom-paris.fr