From linguistic predicate-arguments to Linked Data and ontologies

Extracting n-ary relations

Stamatia Dasiopoulou
Gerard Casamayor
Simon Mille

Universitat Pompeu Fabra, Barcelona, Spain

Heraklion Greece, May 29th 2016

Tutorial Objectives

- Introduction to Natural Language Processing tools and resources for predicate-argument identification in text

- Overview of models and methods for mapping linguistic structures to SW representations

Tutorial Structure

- **9:00-9:20** Part I - Introduction
- **9:20-10:30** Part II - From text to semantic structures
- **10:30-11:00** Coffee break
- **11:00-12:30** Part III - From linguistic representations to RDF/OWL
- **12:30-14:00** Lunch break
- **14:00-14:45** Part IV – Example applications & Evaluation methods
- **14:45-15:30** Hands-on session (I)
- **15:30-16:00** Coffee break
- **16:00-17:30** Hands-on session (II)

Part I: *From linguistic predicate-arguments to LD and ontologies – Extracting n-ary relations*

Introduction

Knowledge Extraction from NL texts for the Semantic Web

- **Goal:** capture information in NL texts and model it in a SW compliant form to allow further processing
- What further processing?
  - Ontology learning and population
  - Natural Language Interfaces
    - Question Answering over KBs
    - NL dialogue systems
  - Semantic search & retrieval
  - Natural Language Generation from ontologies & KBs
    - summarisation, ontology verbalizers, etc.

Typical KE tasks

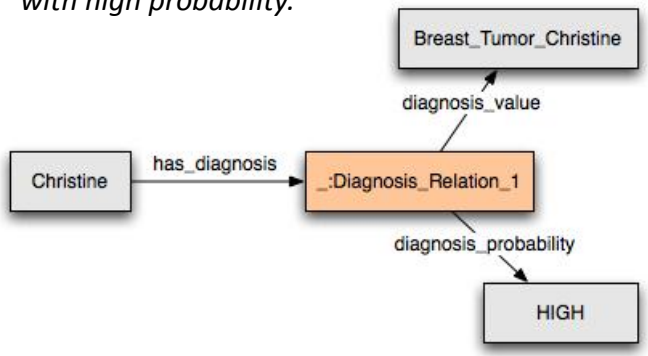
- **Named Entity Recognition (NER)**
  - Detects NEs, often referred to with proper names: people, places, organizations, etc.
  - Tools: Stanford NER, Illinois NET, OpenCalais NER, etc.
- **Entity Linking (EL)**
  - Identity resolution by linking to external knoweldge resources
    - Resources: WordNet, Wikipedia, DBpedia, BabelNet, etc.
  - Tools: Babelfy, DBpedia Spotlight, AGDISTIS, NERD, etc.
- **Binary relation extraction**
  - Corresponds to the typical <subject, predicate, object> triple
  - Tools: REVERB, DBpedia population tools, etc.
- **N-ary relation extraction**
  - Complex relational contexts (events, situations) involving multiple entities
  - Tools: FRED, PIKES, talnSRL, etc.

NLP tasks relevant to KE

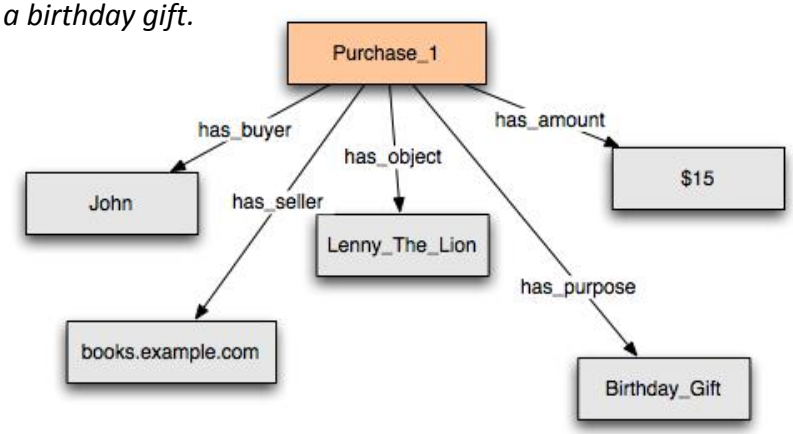
- **Coreference resolution**
  - Detects text fragments referring to the same entity
  - Tools: Stanford CoreNLP, Berkeley Coreference Resolution System, etc
- **Word Sense Disambiguation (WSD)**
  - Assigns senses to words according to some dictionary of senses.
  - Resources: WordNet, BabelNet, etc.
  - Tools: UKB, Babelfy, etc.
- **Syntactic parsing**
  - Captures the functional structure of sentences
  - Resources: Penn Treebank, CCGBank, etc.
  - Tools: Mate tools, Stanford CoreNLP, Berkeley tools, etc.
- **Semantic Role Labelling (SRL)**
  - Assigns semantic types to predicate-argument structures
  - Resources: PropBank, NomBank, VerbNet, FrameNet, etc.
  - Tools: Semafor, Mate tools, etc.
- **Semantic Parsing**
  - Produces (linguistic) semantic structures of sentences of whole texts
  - Resources: DRS, ARM,
  - Tools: Boxer, ARM parsers, talnSLR

N-ary relations

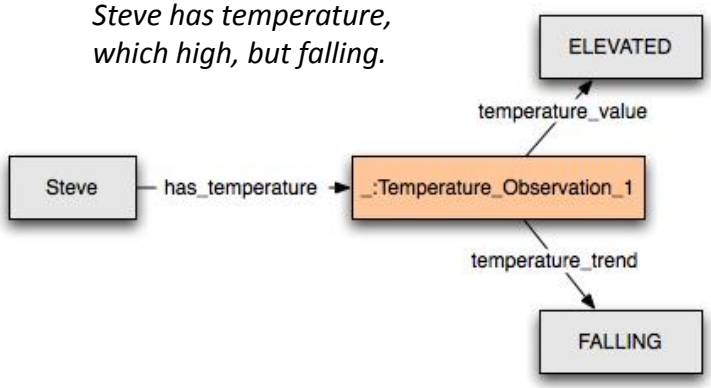
Christine has breast tumour with high probability.



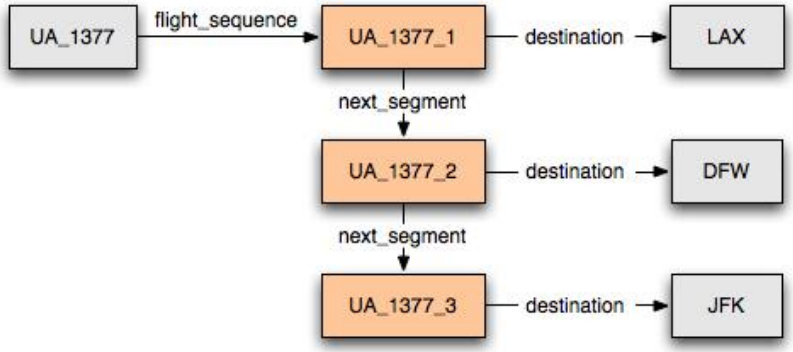
John buys a "Lenny the Lion" book from books.example.com for $15 as a birthday gift.



Steve has temperature, which high, but falling.



United Airlines flight 3177 visits the following airports: LAX, DFW, and JFK.



https://www.w3.org/TR/swbp-n-aryRelations/

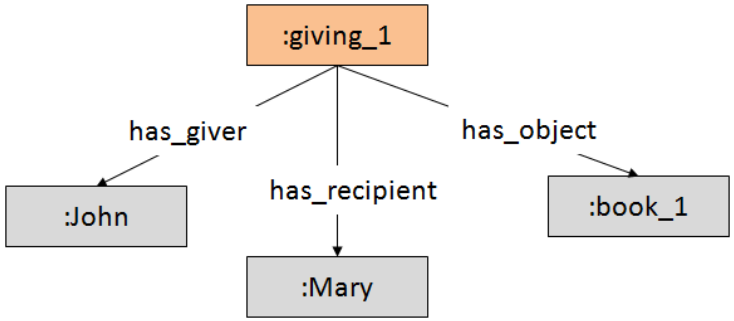
N-ary relations

- **Goal:** Capture the n-ary dependencies!
  - e.g. "*John gave a book to Mary on Wednesday.*"
    "John gave a book to Alice on Thursday."
    "John gave a pen to Paul on Wednesday."

- Query data precisely!
  - What did John give to Mary?
  - To whom did John give a book?
  - What did John on Wednesday?
  - What events/situations did John instigate?
  - What events/situations involve John and Mary?

N-ary relation extraction

- How to go from text to an ontology
  - Many sentences with the same meaning!
  - Conceptual gap!



- *John gave Mary a book.*
- *John gave a book to Mary.*
- *Mary was given a book by John.*
- *What John gave to Mary was a book.*
- *John a donné un livre à Mary.*

N-ary relation extraction

- Easier to get there in several steps
  - A linguistic structure that is close to the text – structure over all the words and only these words
    - *Syntactic structure*
  - A linguistic structure that abstracts from language-specific features and is closer to the meaning
    - *Semantic structure*

Predicates & Arguments

- **Predicate-argument** structures are the basis of many semantic structures
  - e.g. Discourse Representation Theory (DRT), Abstract Meaning Representation (ARM), SemS in Meaning-Text Theory
- A **predicate** is a linguistic element whose meaning isn't complete without its arguments
  - e.g. *yellow* submarine
  - e.g. John *bought* a book from Mary for 5€

Predicates & Arguments

- Predicate-argument structures can be derived from the output of **syntactic parsers**
- However:
  - Parsers do not assign semantic types
  - May not identify semantic relations not expressed syntactically
  - Functional words need to be removed
  - Do not determine the specific entities or concepts participating in the relation
- Other tools are needed to produce semantic structures:
  - SRL, semantic parsers, NER, EL, WSD, etc.

Semantic analysis

- **Semantic Role Labelling (SRL) tools** type predicate-argument relations
  - Semantic role resources: NomBank, PropBank, VerbNet, FrameNet, etc.
  - E.g. Semafor, Mate Tools SRL, Illinois SRL, etc.
- **Semantic parsers** produce semantic **structures** for sentences or whole texts
  - Tools: Boxer (DRS), JAMR (AMR), etc.
- Semantic role **resources**: NomBank, PropBank, VerbNet, FrameNet, etc.
- E.g. "*John <u>bought</u> a book from Mary for 5€*"

SRL resource	John	buy	book	Mary	5€
PropBank	*A0*	*buy.01*	*A1*	*A2*	*A3*
VerbNet	*Agent*	*get-13.5.1*	*Theme*	*Source*	*Asset*
FrameNet	*Buyer*	*Commerce_buy*	*Goods*	*Seller*	*Money*

Bridging NLP and SW

- Part II – Abstracting text
  - Coverage of linguistic phenomena
  - Richness and expressiveness of semantic structures and types
- Part III – From linguistic representations to SW
  - Still a large gap between extracted NLP structures and formal semantics in the SW
    - Relations modelled using RDF versions of lexical resources
    - Non-trivial design & re-engineering is required

Part II – From text to linguistic semantic structures

Outline

- Syntactic parsing
  - Constituency parsing
  - Dependency parsing
- Semantic resources and tools
  - SRL
    - ProbBank/NomBank
    - VerbNet
    - FrameNet
  - Semantic parsing:
    - DRS
    - AMR
    - Semantic Dependencies
- From syntax to semantics

The dog chases the mouse

Individual X

↓

Individual Y

- Understanding words is obviously not enough in order to understand a text
- Need for structuring natural language in order to access the meaning of a sentence
  - More or less abstract ways
  - Syntactic parsing VS Semantic Role Labeling

Syntactic parsing

- Objective: capture the **sentence** structure!
  - how **words** relate together
  - specific to each particular language
  - 2 main paradigms:
    a. constituency-based models: grouping
    b. dependency-based models: linking

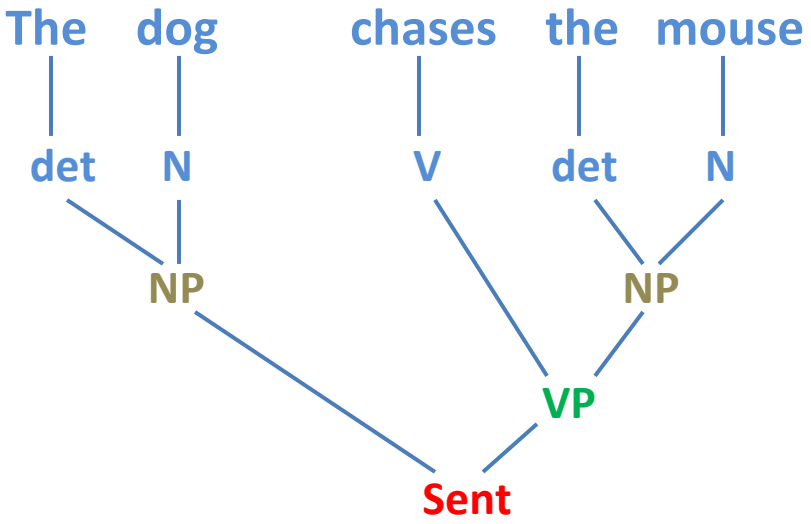
(a) [[The dog] [chases [the mouse]]]

(b) The dog chases the mouse

Constituency Parsing (I)

- Describes phrase groupings (constituents)
  - Based on word order
- Non-terminal elements are **constituent** names
  - Noun Phrase, Verb Phrase, etc.
  - Eg: a well-formed NP is:
    - a sequence of determiner and a noun
    - a sequence of a determiner, an adjective and a noun, etc.

[[The dog]$_{NP}$ [chases [the mouse]$_{NP}$] $_{VP}$] $_{Sent}$



Constituency Parsing (II)

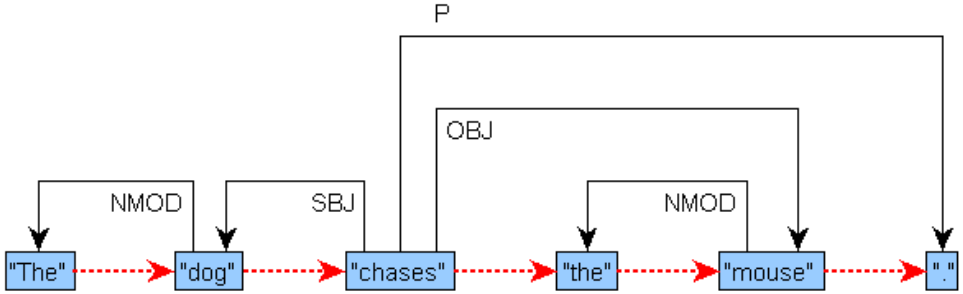
- A constituency parser:
  - automatically produces word groupings
  - predicts if a sentence is gramatically correct or not

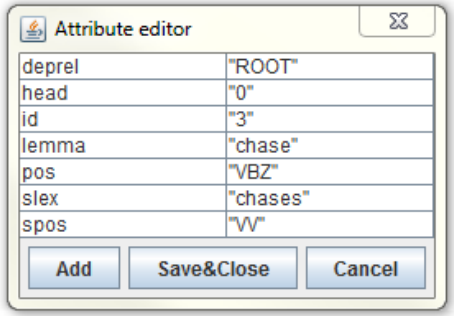
The	dog	chases	the	mouse
NP[nb]/N	N	(S[dcl]\NP)/NP	NP[nb]/N	N
NP[nb]			NP[nb]	
		S[dcl]\NP		
		S[dcl]		

  - has problems with free-order languages
  - Predicate-argument relations can be mapped indirectly
  - Automatic derivation of approximate dependencies

Dependency Parsing (I)

- Describes syntactic relations between words (dependencies)
  - Independent of word order
- Edge labels indicate grammatical function of a **dependent** to a **governor**:
  - Encode agreements, ordering, functional word requirements
  - E.g.: a *subject* is typically an element that triggers number and person agreement on the verb and that is realized before it in a neutral sentence.



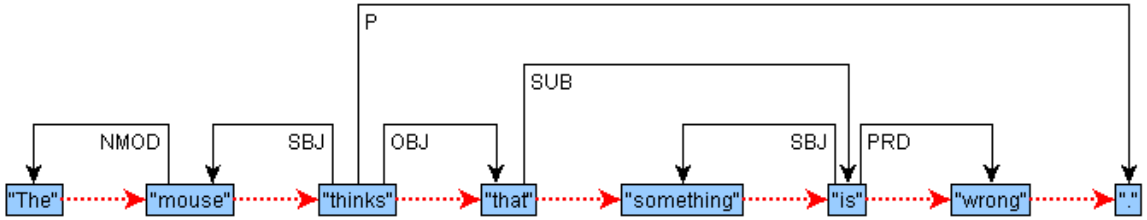


Dependency Parsing (II)

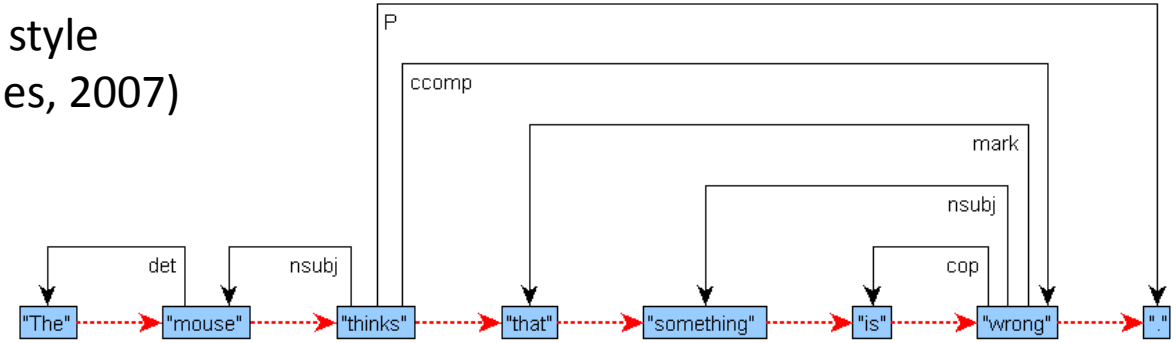
- A dependency parser:
  - automatically produces dependencies between words
  - does not predict if a sentence is gramatically correct or not
    - ... but can model any sentence
  - has no problem with free-order languages
  - Straightforward mapping to predicate-argument relations
  - Automatic derivation of approximate constituency structure

Dependency Parsing (III)

- Many different annotation philosophies
- Different perspectives regarding:
  - notion of governor/dependent
  - edge label nomenclature



Penn Treebank style
(Johansson & Nugues, 2007)



USD style
(De Marneffe & Manning, 2008)

Parsing Pipelines

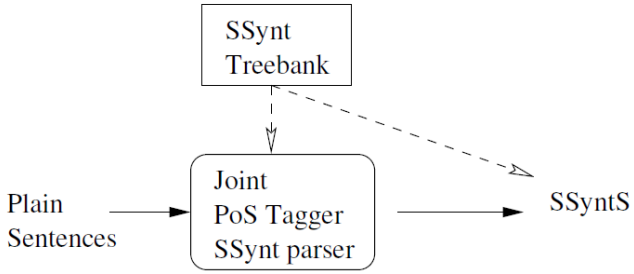
- Typical parsing pipeline
  - Input — won't
  - Tokenization — wo + n't
  - Lemmatization — will + not
  - Part-of-Speech tagging — $will_V$ + not_{RB}
  - Dependency labelling — $will_V$ –*ADV*-> not_{RB}

  or

  Constituency labelling — $[will\ not]_{VP}$

Tools for parsing

- Rule-based parsers
  - not always reliable (coverage issues)
  - output depends on the rules
- Statistical parsers
  - parser + model trained on annotated data
  - output depends on the training data
    - one parser can work with different models
  - always return a complete structure



A few popular parsers

- Dependency (Statistical)
  - Stack LSTM (new): neural networks, state-of-the-art
    - (Dyer et al., 2015)
  - MATE-Tools (PTB)
    - (Björkelund et al., 2010) http://en.sempar.ims.uni-stuttgart.de/
  - TurboParser (USD)
    - (Martins et al., 2010)http://www.cs.cmu.edu/~ark/TurboParser/
- Constituency (Statistical and Rule-based)
  - CCG parsers
    - (Clark & Curran, 2007) http://4.easy-ccg.appspot.com/
  - Stanford parser
    - (Klein & Manning, 2003) http://nlp.stanford.edu:8080/parser/index.jsp

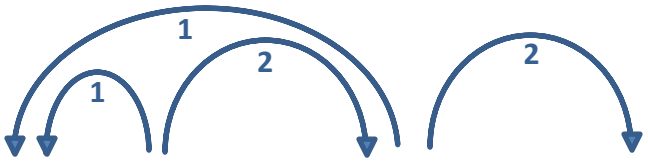
Towards semantic parsing (I)

- Constituency parsing tells us that:
  - “the mouse” and “the dog” are noun groups
  - “chases the mouse” is a verb group
  - “the dog chases the mouse” is a complete sentence
- Dependency parsing tells us that:
  - “dog” is subject of “chases”
  - “mouse” is object of “chases”
- How to go towards linguistic meaning?

Towards semantic parsing (II)

- Objective: get closer to the **meaning**
  - abstract away from language-specific features
    - syntactic relations (constituencies, dependencies)
    - functional words
      - determiners, void conjunctions/prepositions, auxiliaries, etc.
  - capture relations not expressed in syntax
  - common use of predicate-argument structures
    - meaningful words + formalized semantic relations

Towards semantic parsing (III)



The **dog wants** to **chase** the **mouse**

The **mouse** is **chased** by the **dog**

- Lexical resources for typing of semantic-predicate structures:
  - PropBank (Kingsbury & Palmer, 2002) /Nombank (Meyers et al., 2004)
  - VerbNet (Schuler, 2005)
  - FrameNet (Baker et al., 1988)
- Formalisms for semantic structures:
  - Discourse Representation Structures (DRS, Kamp et al., 2011)
  - Abstract Meaning Representation (AMR, Banarescu et al., 2012)
  - Others: Meaning-Text Theory SemS, Prague T-Layer, Minimal Recursion Semantics, Enju, Collapsed Stanford, etc.

Outline

- Syntactic parsing
  - Constituency parsing
  - Dependency parsing
- Semantic resources and tools
  - SRL
    - ProbBank/NomBank
    - VerbNet
    - FrameNet
  - Semantic parsing:
    - DRS
    - AMR
    - Semantic Dependencies
- From syntax to semantics

PropBank/NomBank (I)

- 2 English lexicons http://verbs.colorado.edu/verb-index/
- Contents
  - about 11,000 **predicative words** with one or more **senses**:
    - e.g., senses for verb chase: chase.01, chase.02
  - for each sense, arguments are assigned **semantic roles**:
    - Universal set of predicate-argument roles: A0, A1, A2, A3, A4, A5
    - +AM-.., R-A.., C-A..
- Corpus
  - Penn Treebank syntactic dependencies annotated with PropBank/NomBank

PropBank/NomBank (III)


```
<?xml version="1.0"?>
<!DOCTYPE frameset SYSTEM "frameset.dtd">
- <frameset>
  - <predicate lemma="chase">
      <note> Frames file for 'chase' based on sentences in financial subcorpus. Verbnet entry chase-
        51.6, other framed members include 'follow' and 'pursue'. Comparison with follow. </note>
    - <roleset vncls="51.6" name="follow, pursue" id="chase.01">
      - <roles>
        - <role n="0" descr="follower">
            <vnrole vncls="51.6" vntheta="Agent"/>
          </role>
        - <role n="1" descr="thing followed">
            <vnrole vncls="51.6" vntheta="Theme"/>
          </role>
        </roles>
      - <example name="with causal agent">
          <text> John chased his dinner with a shot of antifreeze. </text>
          <arg n="A">John</arg>
          <rel>chased</rel>
          <arg n="1">his dinner</arg>
          <arg n="M" f="MNR">a shot of antifreeze</arg>
        </example>
      - <example name="with directional">
          <text> Higher margins would chase away dozens of smaller traders who help larger
            traders buy and sell, they say. </text>
          <arg n="0">Higher margins</arg>
          <arg n="M" f="MOD">would</arg>
          <rel>chase</rel>
          <arg n="M" f="DIR">away</arg>
          <arg n="1">dozens of smaller traders who help larger traders buy and sell</arg>
        </example>
        <note> The directional changes the meaning so dramatically that it might be wise to make it
          a different roleset. </note>
      </roleset>
    </predicate>
  </frameset>
```

PropBank/NomBank (IV)

- Sample tool: Mate tools SRL
  - http://en.sempar.ims.uni-stuttgart.de/parse

	The	dog	wants	to	chase	the	mouse
want.01	A0			A1			
chase.01	A0					A1	

- !! **PropBank:** Senses are specific to each linguistic predicate
  - no common semantic types for synonymous or quasi-synonymous verb senses
    - e.g. buy.01 and purchase.01
- !! **Mate tools:** Predicate-argument structures can contain syntactic information
  - relative clauses, circumstantials
  - no distinction between semantic and functional prepositions
- !! **Mate tools:** Structures can be incomplete
  - disconnections (partial coverage of all PoS)
  - annotates arguments as spans

VerbNet (I)

- Theory: Levin verb classes
- English lexicon http://verbs.colorado.edu/verb-index/
- Contents
  - 270 **semantic classes** of verbs inspired from Levin
    - e.g.: Putting, Removing, etc.
    - each class has a list of members (PropBank entries)
  - each class is described with a set of arguments and a set of syntactic frames:
    - about 40 **universal thematic roles**: Agent, Patient, Beneficiary, etc.
    - selectional restrictions: concrete, abstract, location, organization, etc.
    - correspondence with generic predicate-argument structure
- Corpus
  - collection of examples that illustrate VerbNet entries

VerbNet (III)


```
- <FRAMES>
  - <FRAME>
      <DESCRIPTION xtag="0.2" secondary="Basic Transitive" primary="NP V NP"
        descriptionNumber="0.2"/>
    - <EXAMPLES>
        <EXAMPLE>Jackie chased the thief.</EXAMPLE>
      </EXAMPLES>
    - <SYNTAX>
      - <NP value="Agent">
          <SYNRESTRS/>
        </NP>
        <VERB/>
      - <NP value="Theme">
          <SYNRESTRS/>
        </NP>
      </SYNTAX>
    - <SEMANTICS>
      - <PRED value="motion">
        - <ARGS>
            <ARG type="Event" value="during(E)"/>
            <ARG type="ThemRole" value="Agent"/>
          </ARGS>
        </PRED>
      - <PRED value="motion">
        - <ARGS>
            <ARG type="Event" value="during(E)"/>
            <ARG type="ThemRole" value="Theme"/>
          </ARGS>
        </PRED>
      </SEMANTICS>
    </FRAME>
  - <FRAME>
      <DESCRIPTION xtag="" secondary="NP-PPpath-PP" primary="NP V NP
        PP.location" descriptionNumber="0.2"/>
    - <EXAMPLES>
        <EXAMPLE>Jackie chased the thief down the street.</EXAMPLE>
      </EXAMPLES>
```

VerbNet (IV)

- Demo: see DRS (next)
- Very precise description of participants
  - semantics + syntax
- Easily reachable from predicate-argument structures produced by syntactic structures
- Semantic classes abstract over language-specific predicates
- !! **VerbNet:** Classes are coarse-grained
- !! **VerbNet:** Unstructured set of classes, without semantic relations between them.

FrameNet (I)

- Linguistic theory: Frame Semantics (Fillmore)
- English lexicon https://framenet.icsi.berkeley.edu/fndrupal/frameIndex
- Contents
  - 1,000 + **semantic frames**
    - frames describe types of situations, objects, or events
      - e.g. Grooming, Absorb_heat, Usefulness, Food, etc.
    - and are associated with Lexical Units, i.e. words, that evoke them
      - e.g. Grooming LUs include : bathe, shower, comb, floss, shave, etc.
  - for each frame, a set of frame-specific roles is defined
    - about 330 **frame elements**: Buyer.commerce_buy, Ingestor.ingestion, etc.
- Corpus
  - collection of sentences that illustrate FrameNet entries

FrameNet (II)

Cotheme

Lexical Unit Index

Definition:

This frame contains words that necessarily indicate the motion of two distinct objects. The Theme is typically animate and is expressed the same way a Self-mover is expressed in the Self-motion frame--i.e. as the subject of a target verb. The Cotheme may or may not be animate and is typically expressed as a direct object or an oblique. Source, Path, Goal, and the other frame elements common to motion words also regularly occur with the words in this frame.

She ACCOMPANIED him in the ambulance to Hollywood Memorial Hospital.

Police PERSUE teenage joyrider across three counties.

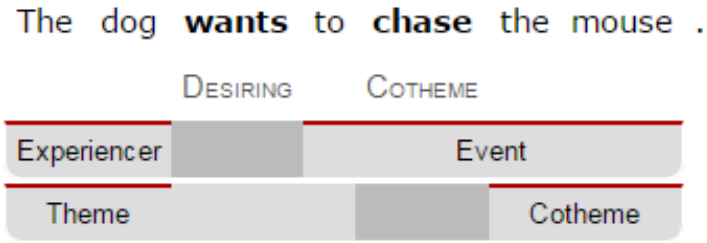
My PURSUIT of academic superiority shows that I was still striving to be the person my parents wanted me to be.

FrameNet (III)

Area [Area] **Semantic Type:** Location	Area marks expressions which describe a general area in which the motion of Theme and Cotheme takes place when the motion is understood to be irregular and not to consist of a single linear path. Locative setting adjuncts of motion expressions may also be assigned this frame element. The police FOLLOWED the suspects all around town.
Cotheme [Thm_c] **Semantic Type:** Physical_object	This frame element is the second moving object, expressed as a direct object or an oblique: Pat ACCOMPANIED me down the street. The squirrel CHASED after the nut.
Direction [dir] **Excludes:** Area	The direction in which the Theme and Cotheme move. The tracking dogs CHASED their target towards the river, where they lost it.
Goal [Goal] **Semantic Type:** Goal **Excludes:** Area	Any expression which tells where the Cotheme ends up as a result of the motion expresses the frame element Goal. Note that if the Cotheme is animate, the Theme need not also end up in the same place. Some particles imply the existence of a Goal which is understood in the context of utterance. The children CHASED the ball into the park. (The children end up in the park.) The children CHASED the dog into the park. (The dog ends up in the park; the children may not have entered the park.)
Path [Path] **Semantic Type:** Path **Excludes:** Area	Path marks phrases that describe the Theme and Cotheme's trajectory of motion and which are neither expressions of the Source nor the Goal of motion. The notion Path also includes directional expressions. The bikers FOLLOWED the truck west. The bikers FOLLOWED the truck through the desert.
Road [Road] **Excludes:** Cotheme	Phrases that denote a physical path on which the motion of Theme and Cotheme takes place are marked Road. The police FOLLOWED the suspect on the interstate for several miles.
Source [Src] **Semantic Type:** Source **Excludes:** Area	Source marks any expression which implies a definite starting-point of the motion of the Cotheme. In prepositional phrases, the object expresses the starting point of motion. With particles, the starting point of motion is understood from context. The cat CHASED the mouse out of the house. The cat CHASED the mouse out.
Theme [Thm] **Semantic Type:** Physical_object	This is the entity, frequently a living being, which moves in relation to the Cotheme. Normally the Theme frame element is expressed as an external argument. Pat ACCOMPANIED me for five miles in a blue Toyota. The squirrel CHASED the nut across the road.

FrameNet (IV)

- Sample tool: Semafor
  - http://demo.ark.cs.cmu.edu/parse



- Frames abstract over language-specific predicates
- Frame to Frame relations
  - inheritance*, subframe, using (perspectivises), causativeOf & inchoactiveOf, precedence, and seeAlso
- Corpus-based descriptions of syntactic frames
- !! **FrameNet:** Mapping from predicate-arguments to roles is complex
- !! **Semafor:** structures can be incomplete
  - disconnections (partial coverage of all PoS)
  - annotates arguments as spans

Linking between PropBank/NomBank, VerbNet and FrameNet

- SemLink
  - (Palmer, 2009)
  - PropBank-VerbNet
  - VerbNet-FrameNet

```
<predicate lemma="chase">
        <argmap pb-roleset="chase.01" vn-class="51.6">
                <role pb-arg="0" vn-theta="Agent" />
                <role pb-arg="1" vn-theta="Theme" />
        </argmap>
```

```
<vncls class='51.6' fnframe='Cotheme'>
  <roles>
    <role fnrole='Theme' vnrole='Agent'/>
    <role fnrole='Cotheme' vnrole='Theme'/>
  </roles>
</vncls>
```

- PredicateMatrix
  - (De Lacalle et al., 2014)
  - PropBank, VerbNet, FrameNet, WordNet, Multilingual Central Repository

vn:chase-51.6 vn:51.6 vn:NULL vn:NULL vn:chase_after vn:Agent
wn:chase_after%2:38:00 mcr:ili-30-02001858-v fn:Cotheme fn:chase.v fn:Theme pb:chase.01
pb:0 mcr:0 mcr:factotum mcr:Pursuing mcr:Dynamic;Location; mcr:motion
mcr:go%2:38:00 wn:10 wn:017 SEMLINK;SYNONYMS

Outline

- Syntactic parsing
  - Constituency parsing
  - Dependency parsing
- Semantic resources and tools
  - SRL
    - ProbBank/NomBank
    - VerbNet
    - FrameNet
  - Semantic parsing:
    - DRS
    - AMR
    - Semantic Dependencies
- From syntax to semantics

Discourse Representation Structures (I)

- Based on Discourse Representation Theory (DRT)
  - First Order Logic formal theory of meaning
- Phenomena covered:
  - Predicate-argument relations
  - Negation, equivalence, quantifiers, sets, modality, etc.
- DRSs building blocks
  - Discourse referents, i.e. variables corresponding to discourse entities (event, objects, etc.)
  - Conditions, i.e. constraints (properties, relations) on discourse entities: e.g. chase(x, y), dog(x), mouse(y)
- Corpus
  - Groningen Meaning Bank: http://gmb.let.rug.nl/
  - Whole texts are annotated rather than isolated sentences

Discourse Representation Structure (II)

- DRS semantic parser: **Boxer**
  - http://gmb.let.rug.nl/webdemo/demo.php#drs
  - Event-based Neo-Davidsonian representation (reified n-ary relations) using VerbNet thematic roles
  - e.g. *"The dog wants to chase the mouse."*

```
x1 p1 e1

dog(x1)
p1:  x2 e2
     mouse(x2)
     chase(e2)
     Actor(e2, x1)
     Theme(e2, x2)
want(e1)
Actor(e1, x1)
Topic(e1, p1)
```

- FOL representation

$$\exists A : \exists B : ((n1mouse(A) \wedge n1dog(B)) \wedge \exists C : \exists D : (r1Topic(C, D) \wedge (r1Actor(C, B) \wedge (v1want(C) \wedge \exists E : (r1Theme(E, A) \wedge (r1Actor(E, B) \wedge v1chase(E)))))))$$

Abstract Meaning Representation (I)

- Abstract Meaning Representation (AMR)
  http://amr.isi.edu/language.html
- Phenomena covered:
  - Predicate-argument relations
  - Within-sentence coreference
  - NEs and types
  - Modality, negation, questions, quantities, etc.
- AMRs are rooted, labelled graphs
- Makes extensive use of PropBank
- Corpus: AMR Bank containing English sentences.

Abstract Meaning Representation (II)

- AMR **concepts** English words (boy), PropBank framesets (want-01), special keywords.
- **Keywords** entity types (date-entity, world-region), quantities (monetary-quantity, distance-quantity), and logical conjunctions (and).
- Probank **arguments** arg0, arg1, etc.
- Universal set of **semantic relations**: accompanier, age, beneficiary, cause, etc.

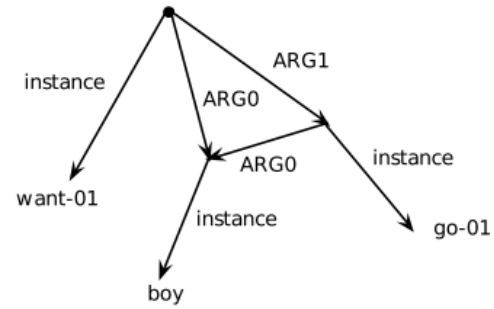
LOGIC format:

$\exists$ w, b, g:
instance(w, want-01) $\wedge$ instance(g, go-01) $\wedge$
instance(b, boy) $\wedge$ arg0(w, b) $\wedge$
arg1(w, g) $\wedge$ arg0(g, b)

AMR format (based on PENMAN):

```
(w / want-01
  :arg0 (b / boy)
  :arg1 (g / go-01
          :arg0 b))
```

GRAPH format:

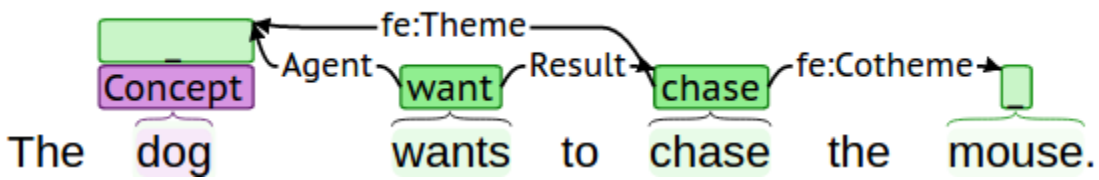


Abstract Meaning Representation (III)

- AMR semantic parser: **JAMR** https://github.com/jflanigan/jamr
- !! **AMR:** Annotation at sentence-level only (no discourse).
- !! **AMR:** Uses propbank
- !! **AMR:** Heaviliy geared towards English
- !! **JAMR:** English only

Semantic Dependencies (I)

- Semantic dependencies
- Similar formalism to syntactic dependencies, but with word-to-word semantic relations between content words.
- Phenomena covered:
  - Predicate-arguments relations.
  - Modality, appositions, argument sharing in coordinations and complex verb constructions, etc.
- Can integrate with PropBank, VerbNet or FrameNet.



Semantic Dependencies (II)

- Semantic dependencies parser: talnSRL http://dparse.multisensor.taln.upf.edu/demoMS/relExtr/
- Multilingual analysis (in progress).
- NER and EL
- !! **Dependencies:** Annotation at sentence-level only (no discourse).
- !! **Dependencies:** No semantic interpretation of negations, sets and quantifiers (yet).

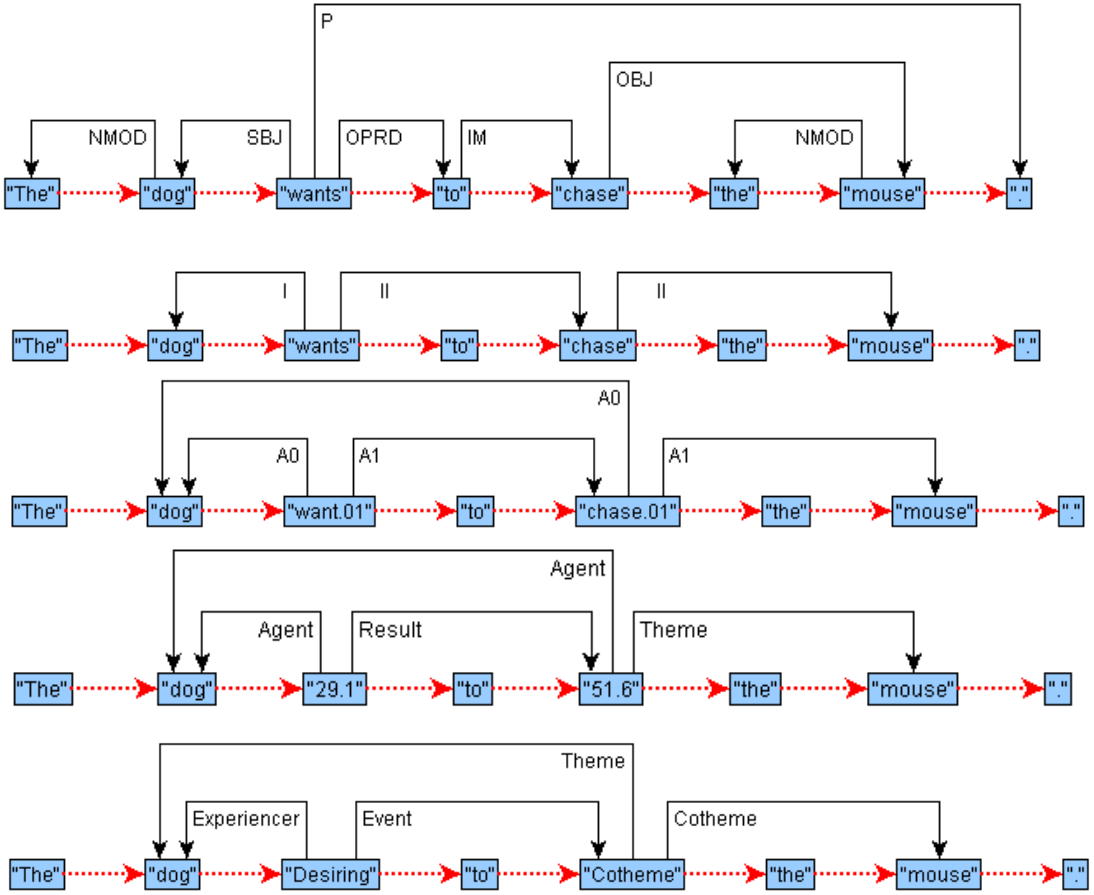
Outline

- Syntactic parsing
  - Constituency parsing
  - Dependency parsing
- Semantic resources and tools
  - SRL
    - ProbBank/NomBank
    - VerbNet
    - FrameNet
  - Semantic parsing:
    - DRS
    - AMR
    - Semantic Dependencies
- From syntax to semantics

From dependencies to semantics (I)

- Parallelism between consecutive outputs (talnSRL)

1. surface-syntax
2. deep-syntax
3. PropBank/NomBank
4. VerbNet
5. FrameNet

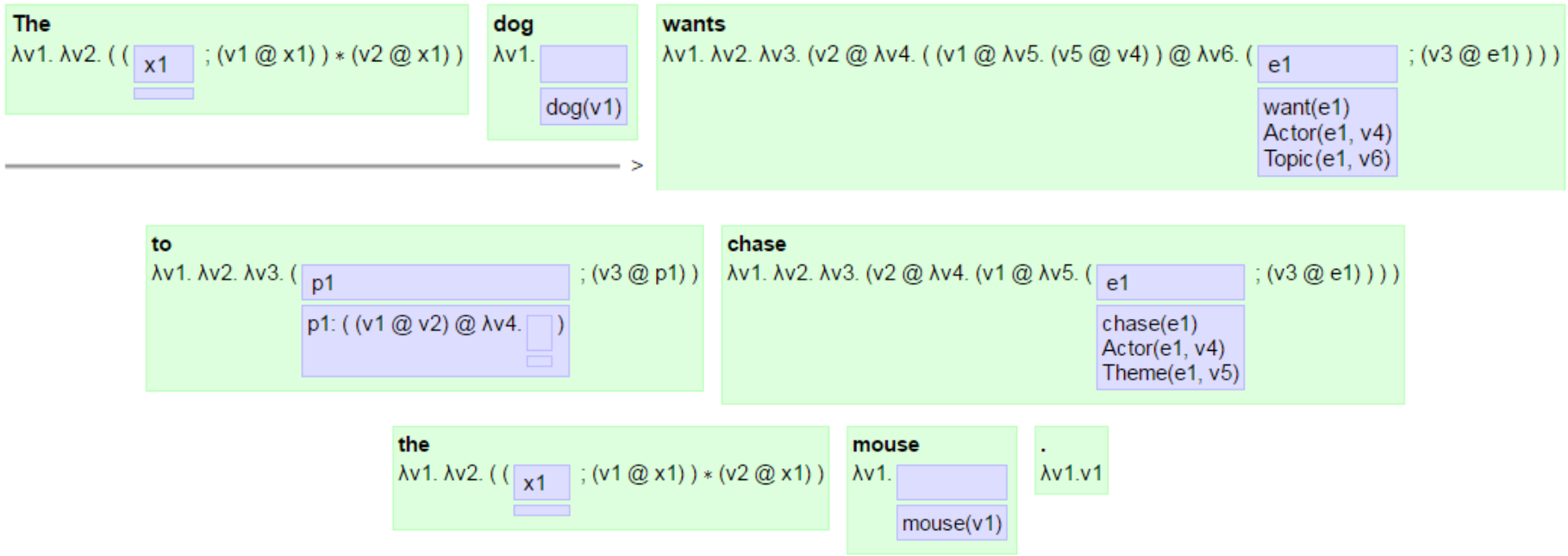


From dependencies to semantics (II)

- What has to be done starting from syntactic dependencies:
  - Identify functional nodes to remove
  - Find new governor/dependent pairs
  - Label the edges
- ML (using annotated corpora) or rules (using SemLink)
- Mapping not always one-to-one
  - e.g. SBJ-> I/II if active/passive
  - Even more complex examples!

From constituents to semantics (I)

- Example: CCG to DRS
  - The mapping is not straightforward
  - Semantic types (lambda expressions) can be manually assigned to words and to constituents



From constituents to semantics (II)

the mouse
λv1. (x1 * (v1 @ x1))
mouse(x1)

chase the mouse
λv1. λv2. (v1 @ λv3. (x1 e1 ; (v2 @ e1)))
mouse(x1)
chase(e1)
Actor(e1, v3)
Theme(e1, x1)

to chase the mouse
λv1. λv2. (p1 ; (v2 @ p1))
p1: (v1 @ λv3. x1 e1)
mouse(x1)
chase(e1)
Actor(e1, v3)
Theme(e1, x1)

wants to chase the mouse
λv1. λv2. (v1 @ λv3. (p1 e1 ; (v2 @ e1)))
p1: x1 e2
mouse(x1)
chase(e2)
Actor(e2, v3)
Theme(e2, x1)
want(e1)
Actor(e1, v3)
Topic(e1, p1)

The dog wants to chase the mouse
λv1. (x1 p1 e1 ; (v1 @ e1))
dog(x1)
p1: x2 e2
mouse(x2)
chase(e2)
Actor(e2, x1)
Theme(e2, x2)
want(e1)
Actor(e1, x1)
Topic(e1, p1)

From constituents to semantics (III)

- What has to be done starting from constituency structures:
  - ensure that there is a correct type corresponding to each constituent
  - ensure that they can combine properly
- Lambda-expressions must pre-exist
  - !! potential coverage issues

From syntax to semantics

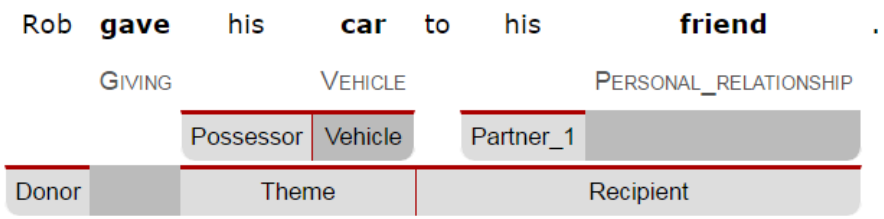
- Disambiguation needed when the predicate has multiple senses/VerbNet classes/frames
  - e.g. play: play.01, play.02?
- Semantic types of arguments can help to determine roles and resolve selectional restrictions
  - Play football with Peter
  - Play Peter with football

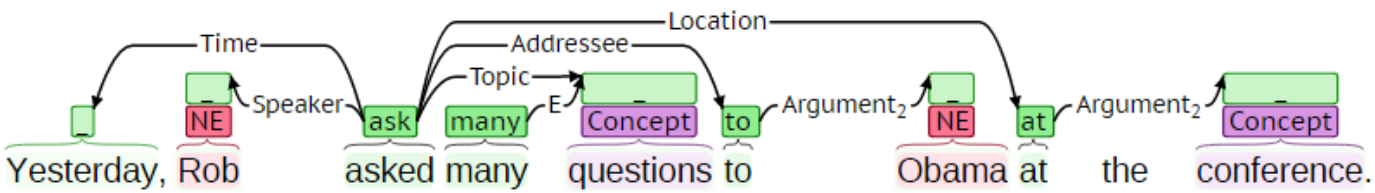
Summary (I)

- Variety of linguistic types:
  - Predicates: disambiguated words, VerbNet classes, FrameNet frames
  - Argument roles: A0, Agent, Buyer.commerce_buy...
  - Structures: DRS, AMR, semantic dependencies, etc.
- Both constituency and dependency parsing can be used to extract these representations
- Linguistic resources are crucial!
  - Linguistic theories -> formal definitions of structures
  - Annotated corpora -> to train models for tools
  - Semantic lexicons -> to assign semantic types
  - Rule-sets -> to implement parsers or mappings between structures

Summary (II)

- Semantic parsers and SRL tools work naturally with with n-ary relations!





x1 x2 x3 e1 x4 x5

named(x1, rob, per)
male(x2)
of(x3, x2)
car(x3)
give(e1)
Actor(e1, x1)
Theme(e1, x3)
male(x4)
of(x5, x4)
named(x5, friend, tim)
to(e1, x5)

	The	shares	raised	from	1,25	$	to	1,78	$	this	morning	.
raise.01	A1			A3						AM-TMP		
1,78.01												

Summary (III)

- What semantic representation is best?
  - Depends on the application:
    - For knowledge extraction, disconnected representations may be ok, e.g. predicate-argument structures identified by SLR tools.
    - For summarization, comprehensive structures covering whole sentences are needed!
  - In general, the more expressive and accurate the abstraction the better starting point it serves
- How to formalize linguistic structures to Semantic Web compliant ones?
  - Part III

Part III – From linguistic representations to RDF/OWL

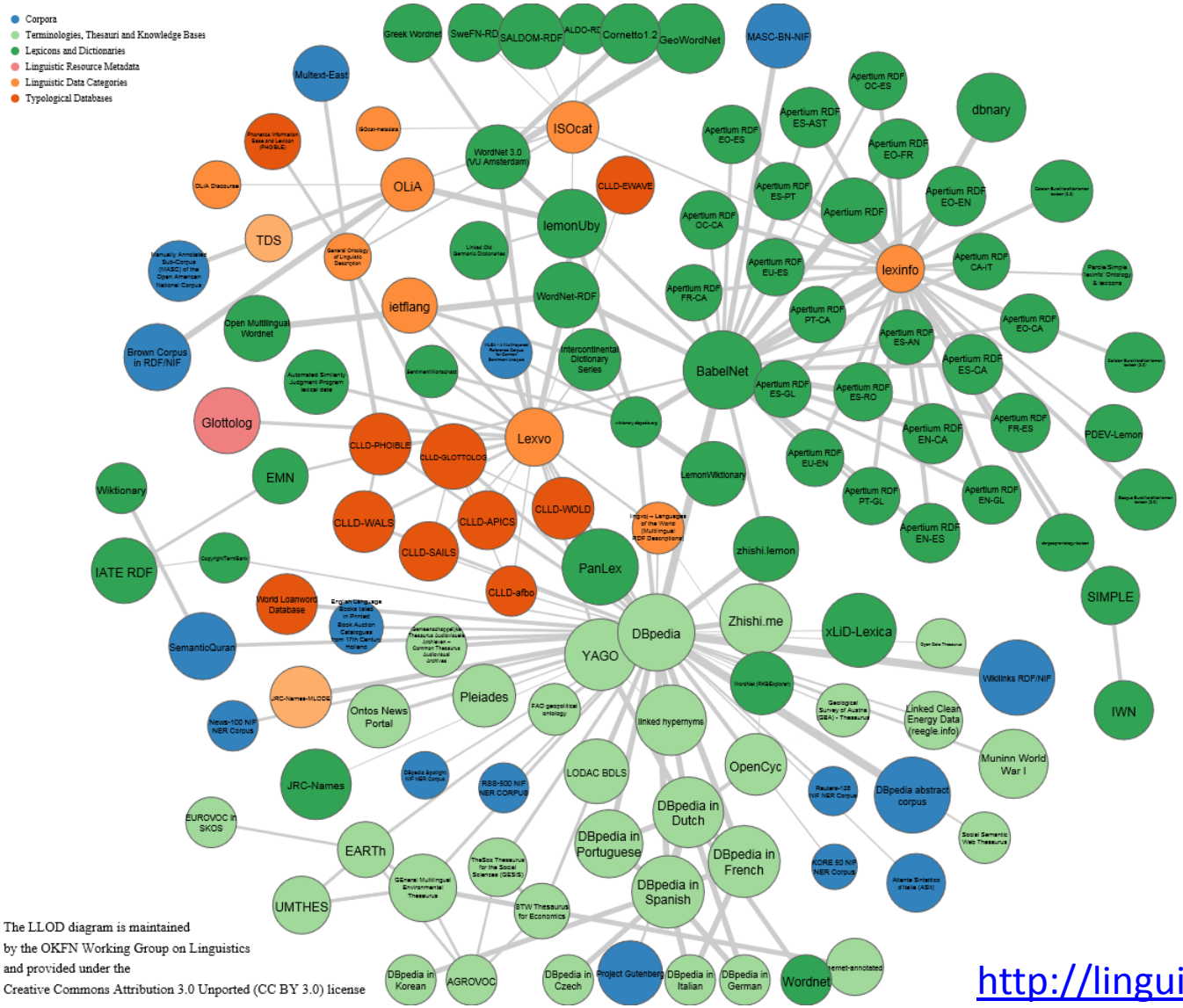
Outline

- Porting linguistic data to RDF
  - Models
    - NIF, EARMARK
    - PROV-O, ITS and NERD, OLiA
    - lemon
    - PreMOn
- Mapping linguistic structures to SW representations
  - Some relevant background
    - DUL DnS, c.DnS, LMM
  - Approaches
    - LODifier
    - talnSRL
    - PIKES
    - FRED

Bridging NL and the SW

1. Port linguistic resources and output of NLP tools to RDF
   - Motivated by reuse & interoperability of linguistic resources and tools
   - Publish Linguistic Open Linked Data
2. Translate linguistic structures into SW knowledge representations
   - Distil conveyed knowledge
   - Allow for automated reasoning

The Linguistic LOD cloud (LLOD)



The LLOD diagram is maintained
by the OKFN Working Group on Linguistics
and provided under the
Creative Commons Attribution 3.0 Unported (CC BY 3.0) license

http://linguistic-lod.org/

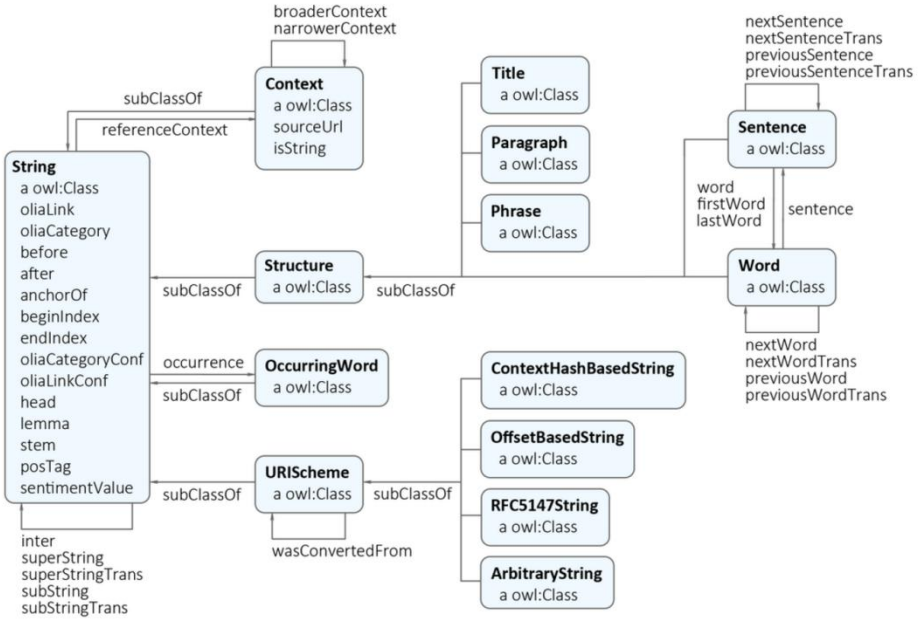
Porting NLP to the SW

- Relevant models for publishing LLOD
  1. Ontologies used to model the textual annotations produced by NLP and KE tools
     - e.g. NIF, EARMARK, PROV-O, OLiA, ITS, NERD
  2. Ontologies used to associate linguistic information to existing ontologies and LD
     - e.g. lemon, OntoLex
  3. Existing Linked Data versions of linguistic resources, e.g. lexica and corpora
     - e.g. PreMOn, WordNet RDF, FrameNet RDF, etc.

NIF

- Set of specifications, ontologies and software to achieve interoperability between NLP tools, resources and annotations.
- Core ontology for modelling of stand-off annotations of text.
- Used by TALN-UPF SRL, FRED and PIKES.

http://persistence.uni-leipzig.org/nlp2rdf/

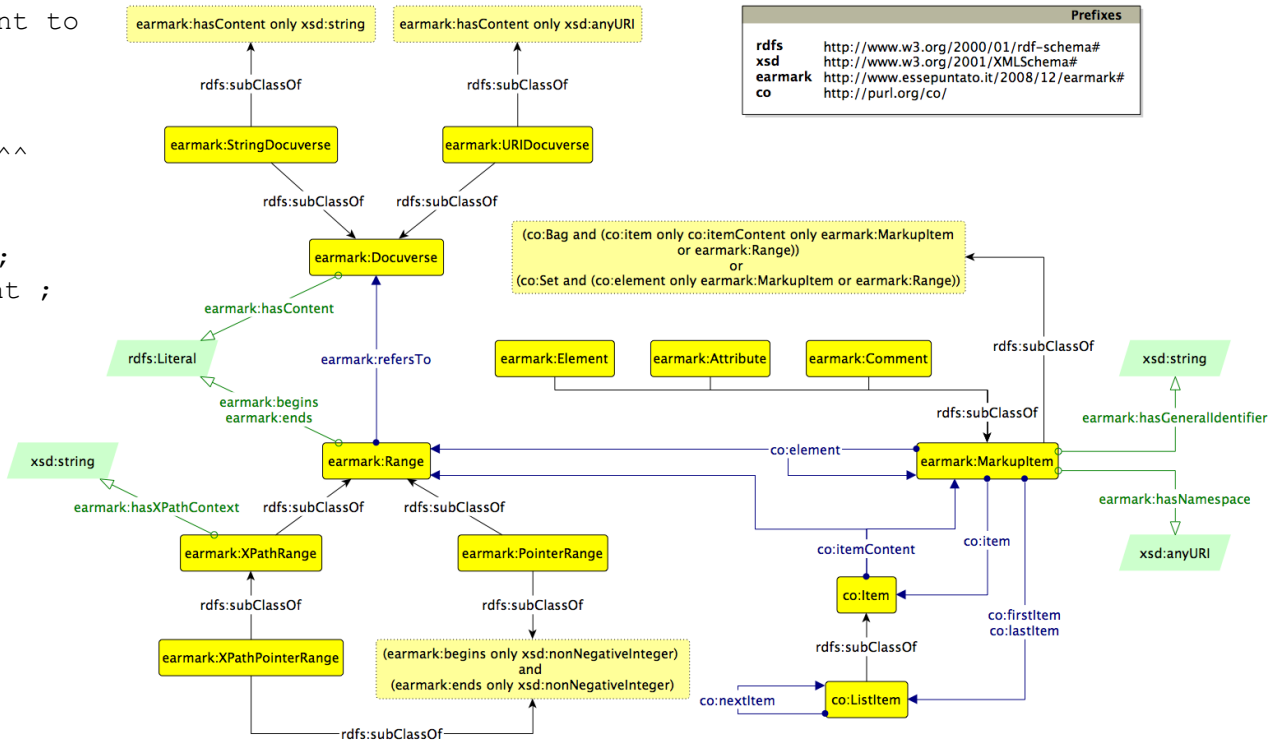



```
<http://brown.nlp2rdf.org/corpus/a01.xml#offset_4_10>
        a nif:String , nif:Word , nif:OffsetBasedString ;
        nif:anchorOf "Fulton"^^xsd:string ;
        nif:referenceContext <http://brown.nlp2rdf.org/corpus/a01.xml#offset_0_161> ;
        nif:beginIndex "4"^^xsd:int ;
        nif:endIndex "10"^^xsd:int ;
        a nif:EntityMention ;
        nif:confidence "0.98"^^xsd:decimal .
```

EARMARK

- Ontology for modelling non-embedded markup on documents, such as XML and XHTML.
- It has also been used to model stand-off semantic annotations of text in FRED.

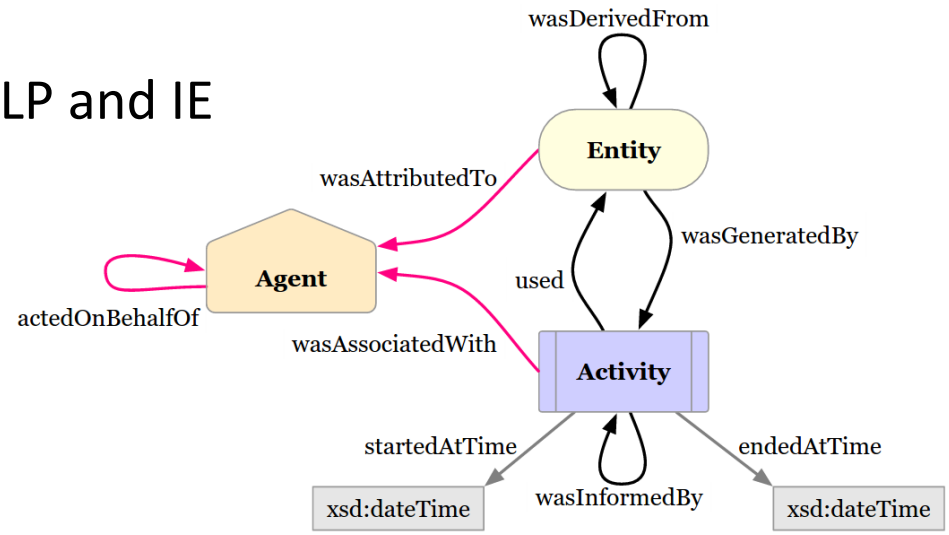
```
# The textual content of the document to
annotate
:content    a earmark:URIDocuverse ;
            earmark : hasContent
"http://ota.ox.ac.uk/text/5725.xml"^^
xsd:anyURI .
#The string "Ari."
:ari-string a earmark:PointerRange ;
            earmark:refersTo :content ;
            earmark:begins
"33974"^^xsd:nonNegativeInteger ;
            earmark:ends
"33978"^^xsd:nonNegativeInteger .
```



PROV-O

- Ontology of the PROV set of documents to model provenance information.
- Extended to model provenance of NLP and IE annotations in NIF 2.0 and 2.1 draft.

  https://www.w3.org/TR/prov-o/




```
<http://brown.nlp2rdf.org/corpus/a01.xml#offset_4_10>
    a nif:Word ;
    nif:anchorOf "Fulton"^^xsd:string ;
    nif:annotationUnit [
        a nif:EntityMention ;
        nif:confidence "0.98"^^xsd:decimal
        prov:wasAttributedTo <http://aksw.org/MarkusAckermann.rdf> ] ;
    nif:annotationUnit [
        nif:taNerdCoreClassRef nerd:Location ;
        nif:confidence "0.85"^^xsd:decimal ;
        prov:wasAttributedTo <http://aksw.org/MartinBruemmer.rdf> ] ;
    nif:annotationUnit [
        itsrdf:taIdentRef <http://dbpedia.org/page/Fulton_County,_Georgia> ;
        nif:confidence "0.72"^^xsd:decimal ;
    prov:wasAttributedTo http://aksw.org/MartinBruemmer.rdf ] .
```

ITS and NERD

- Internationalizing Tag Set (ITS) 2.0
- Provides metadata for internationalization, translation and localization of XML, HTML and XML Schema.

  https://www.w3.org/TR/its20/, https://github.com/w3c/itsrdf

- NERD is an ontology that models a set of taxonomies of NE types and provides mappings between them.

  http://nerd.eurecom.fr/ontology

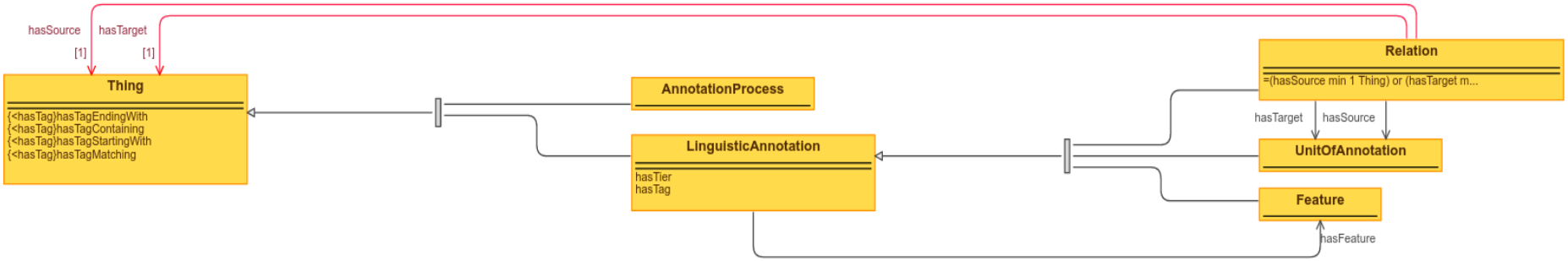
- Coupled with NIF, ITS and NERD have been used to model in RDF stand-off annotations resulting from NER, EL and concept extraction tools.

```
<#char=2797,2811>
          a nif:Phrase ;
          nif:anchorOf "European Union" ;
          nif:beginIndex "2797"^^xsd:nonNegativeInteger ;
          nif:endIndex "2811"^^xsd:nonNegativeInteger ;
          its:taClassRef nerd:Organization ;
          its:taIdentRef bn:00021127n ;
          its:taIdentRef dbpedia:European_Union;
```

OLiA

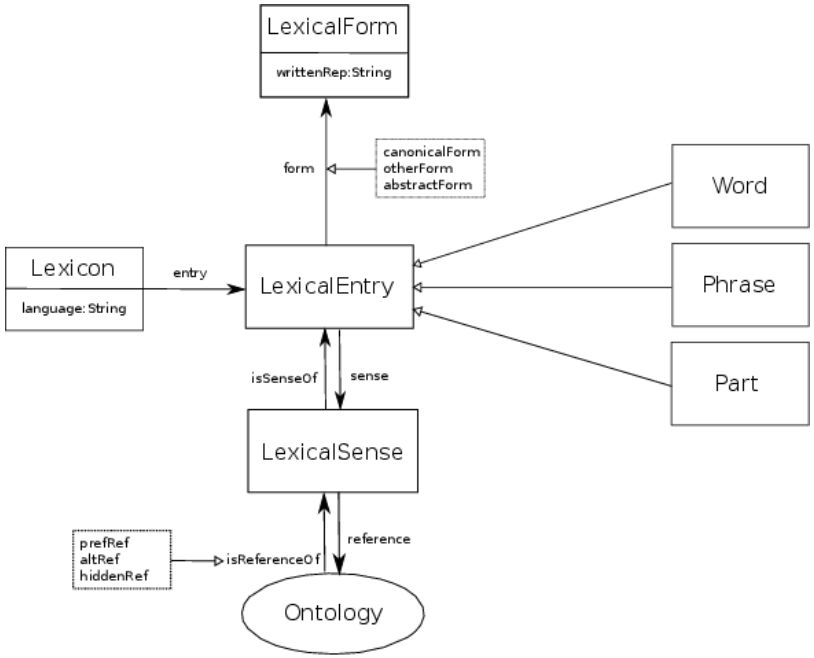
- http://nachhalt.sfb632.uni-potsdam.de/owl/
- Set of ontologies for annotation schemes used in NLP, linked together under a common reference model of linguistic concepts.
  - e.g. Penn Treebank
- Integrated into NIF

```
<#char=1478,1484> nif:oliaLink penn:NNS .
```



lemon

- http://lemon-model.net/
- Ontology for modelling lexicons information relative to ontology classes/properties..)
- Used in DBpedia, BabelNet, PreMOn
- Extended by Ontolex W3C group (work in progress)

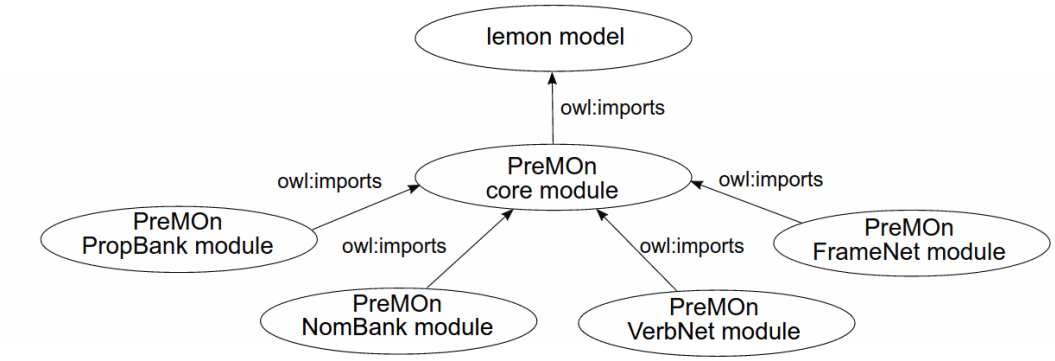



```
:tuberculosis lemon:canonicalForm [
          lemon:writtenRep "tuberculosis"@en ] ;
          lemon:sense [
               lemon:reference <http://dbpedia.org/resource/Tuberculosis> ] .

:consumption lemon:canonicalForm [
     lemon:writtenRep "consumption"@en ] ;
          lemon:sense [
               lemon:reference <http://dbpedia.org/resource/Tuberculosis> ;
               lemon:context :antiquated  ] ;
          lemon:sense [
               lemon:reference <http://dbpedia.org/resource/Consumption_(economics)> ] .
```

PreMOn

- https://premon.fbk.eu/
- Set of RDF datasets for predicate lexicons encoded and mapped according to an ontology for modelling semantic classes and roles
- Lexicons covered:
  - PropBank
  - NomBank
  - VerbNet
  - FrameNet
  - Semlink
- Uses Lemon, NIF.




```
# VerbNet classs
pm:vn32-wipe_manner-10.4.1   a                 pmovn:VerbClass  .
pm:vn32-wipe_manner-10.4.1   rdfs:label        "wipe_manner-10.4.1" .

#subclass:
pm:vn32-wipe_manner-10.4.1-1 a                 pmovn:VerbClass  .
pm:vn32-wipe_manner-10.4.1-1 skos:broader      pm:vn32-wipe_manner-10.4.1

#x member:
pm:v-trim                    ontolex:evokes    pm:vn32-wipe_manner-10.4.1 .
```

Models Overview

Ontology	Information modelled	Applications
NIF Core Ontology	Stand-off annotations	• Exchange of information between text analysis services. • Production of LOD corpora
EARMARK		
ITS + NIF + NERD	Annotation of NEs and concepts	
PROV-O + NIF	Provenance and confidence	
OLiA	NLP annotation tagsets	
lemon	Non-predicative Lexicons	• Extend ontologies and LD with language-related information.
PreMOn	Predicative lexicons	• Exchange of information between text analysis services • Production of LOD corpora

Topics of this part

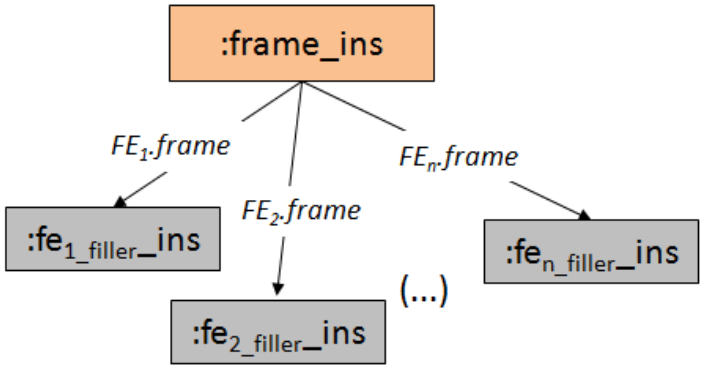
- Porting linguisitic data to RDF
  - Models
    - NIF, EARMARK
    - PROV-O, ITS and NERD, OLiA
    - lemon
    - PreMOn
- Mapping linguistic structures to SW representations
  - Some relevant background
    - DUL DnS, c.DnS, LMM
  - Tools
    - LODifier
    - talnSRL
    - PIKES
    - FRED

Mapping linguistic structures to OWL

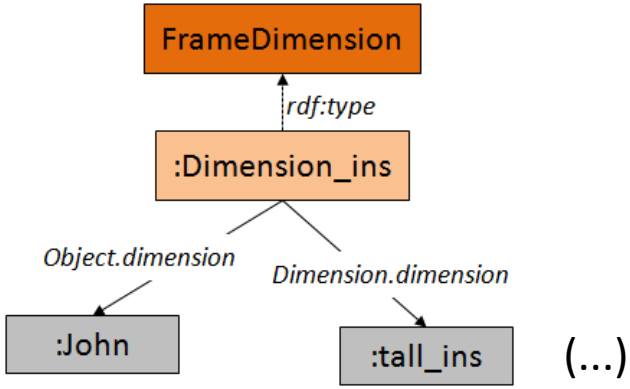
- Common translation practises:
  - NEs -> owl:NamedIndividual
  - Entity resolution/coreference -> owl:sameAs
  - Term -> owl:Class or owl:ObjectProperty / owl:DatatypeProperty
  - Sense Disambiguation -> owl:equivalentClass
  - Binary relation/semantic role -> owl:ObjectProperty or owl:DatatypeProperty
  - Taxonomy -> rdfs:SubClassOf
- But that's just half of the story...

Leveraging linguistic and ontology semantics (I)

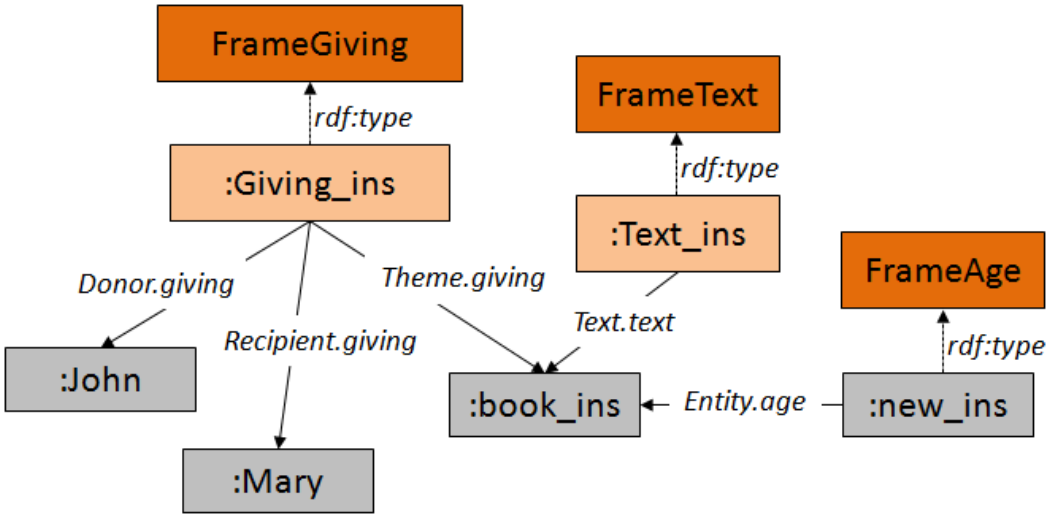
- E.g. uniform reified modelling
  - frame -> owl:NamedIndividual & Frame class
  - frame element -> objectProperty
    - connects frame and frame element filler instances



- *John is tall.*



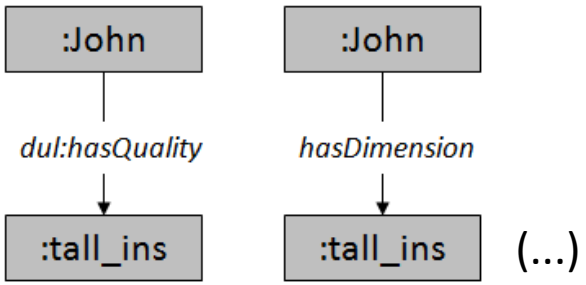
- *John gave Mary a new book.*



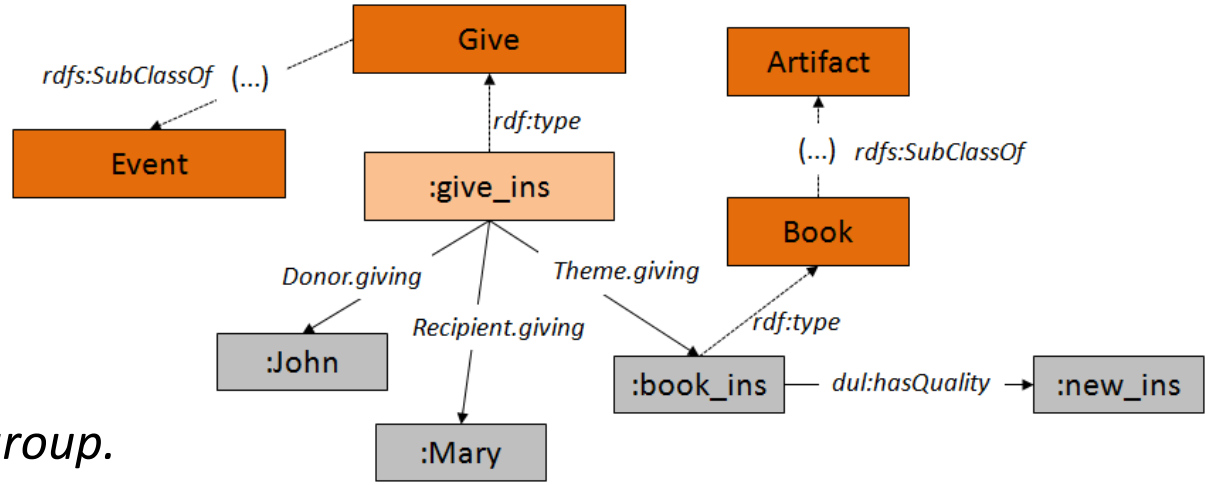
Leveraging linguistic and ontology semantics (II)

- E.g. distinguish between frames' ontological types
  - frame categories: events, artifacts, attributes, etc.
  - frame-category tailored element representation

- *John is tall.*



- *John gave Mary a new book.*



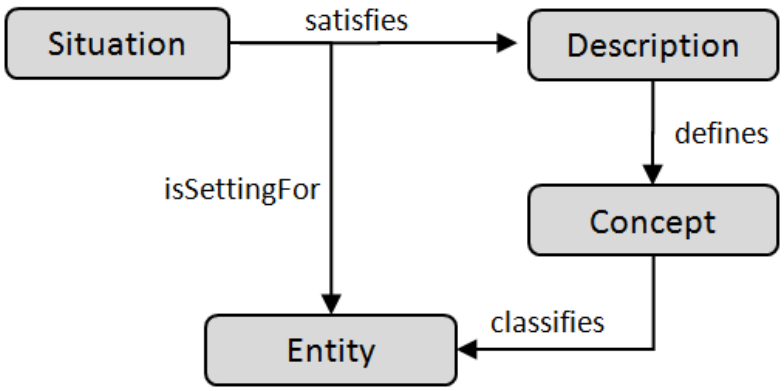
- *BUT John is the leader of the group.*

Leveraging linguistic and ontology semantics

- Challenges
  - gap between NLP structures and formal semantics adopted by SW ontologies
  - variety of requirements and application domains motivate different choices/translation practises from NLP to RDF/OWL representations

Some background

- *Description & Situation* (DnS) pattern in DOLCE+DnS Ultralite (DUL)
  - "relational context" based on a "descriptive context" (e.g. Diagnosis, Plan)
  - reified n-ary relation representation



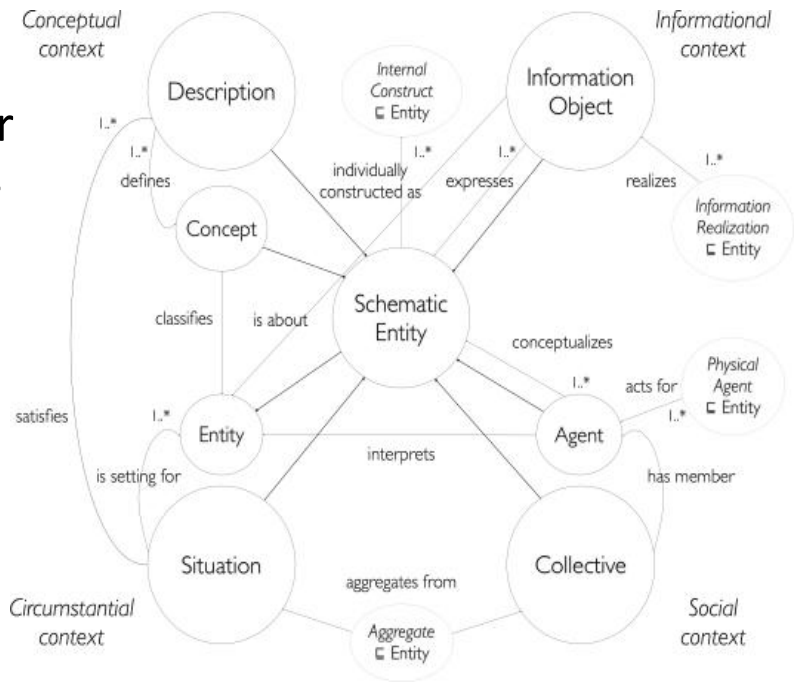
- *E.g. a DiagnosedSituation is a context of observed entities on the basis of a Diagnosis (Description)*
- *E.g. a SensorObservation is a context of a sensor, sensing method, observed feature and observation value, etc. satisfying a respective Observation (Description)*

c.DnS

- Semiotic triangle realisation
  - informational context -> expression layer
  - circumstantial context -> reference layer
  - social context -> interpreter layer
  - all contexts -> meaning layer

- Application to FrameNet
  - Frames: meanings & descriptions
  - Frame Elements: meanings & concepts
  - Lexical Units: meanings & descriptions

- Also for WordNet, VerbNet

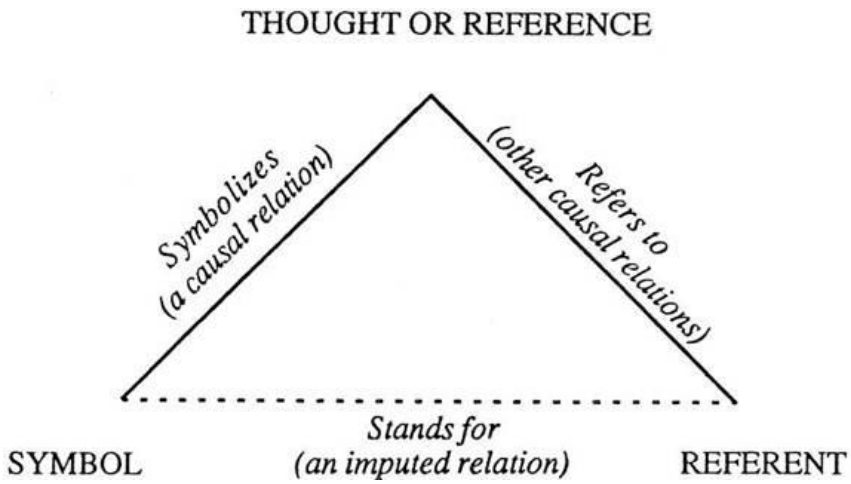


(Gangemi, 2009)

Linguistic Meta-Model (LMM)

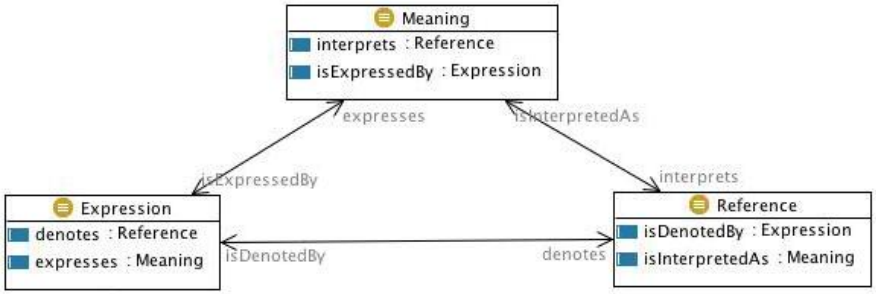
- OWL-DL ontology for semiotic-cognitive representation of lexical knowledge
  - aligned with DOLCE Ultra Light (DnS)
  - multilingual

- LMM1
- LMM2
- LMM Alignments



Linguistic Meta-Model (LMM)

- LMM1 formalises the semiotic triangle



(Picca et al., 2008)

  - Reference: all entities belonging to the universe of discourse
    - dul:PhysicalArtifact, lmm2:IndividualReference, lmm2:MultipleReference, dul:Situation
  - Meaning:
    - dul:Description, dul:Collection, dul:Situation, dul:Concept
  - Expression: social objects produced by agents

Linguistic Meta-Model (LMM)

- LMM2
  - extends LMM1 capturing specific linguistic constructs and references

- LMM Alignments
  - LMM1, LMM2, DUL
  - WordNet, FrameNet, SKOS...
    - specialising dul:Concept (e.g. WNsynset, FE)
    - specialising dul:specialiszes (e.g. WNhypernym)

Outline

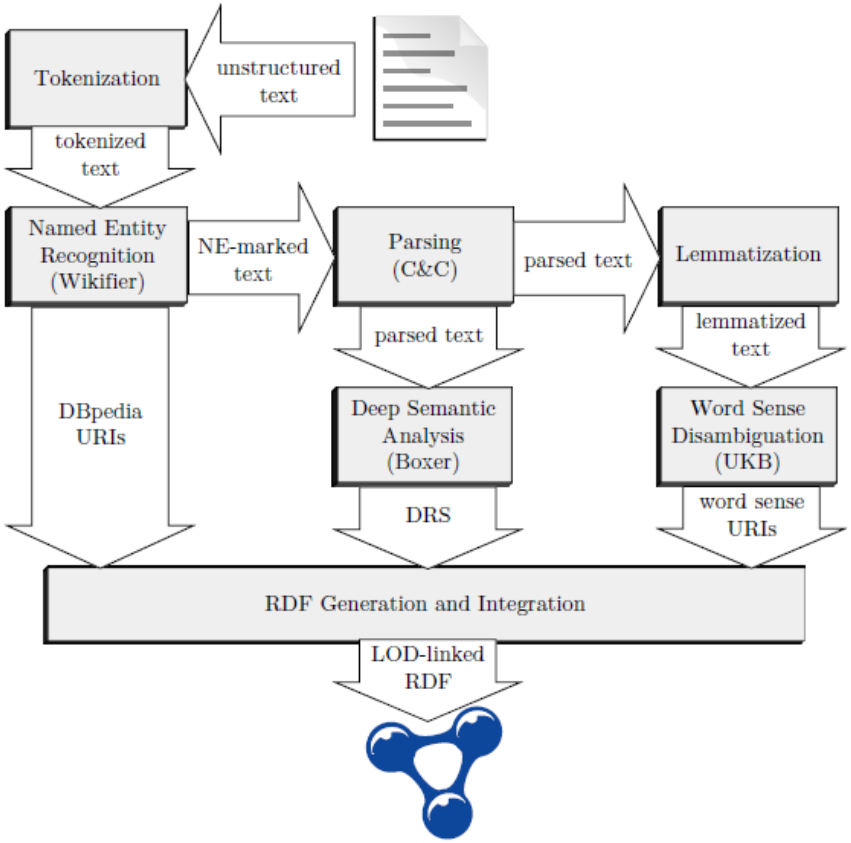
- Porting linguistic data to RDF
  - Models
    - NIF, EARMARK
    - PROV-O, ITS and NERD, OLiA
    - lemon
    - PreMOn
- Mapping linguistic structures to SW representations
  - Some relevant background
    - DUL DnS, c.DnS, LMM
  - Approaches
    - LODifier
    - talnSRL
    - PIKES
    - FRED

LODifier

- Generate Linked Data from unstructured text
  - RDF graphs linked to DBpedia and WordNet
- **Goal**: capture information from NL text as Linked Data
- Design principles
  - CCG & DRS
  - Linking to LOD cloud

LODifier

- NLP and KE tasks
  - Syntactic parsing: CCG
  - Semantic parsing and coreference resolution: Boxer
  - NER/EL: Wikifier
  - WSD: UKB
- NL to SW
  - Porting of DRSs to RDF



LODifier

- DRSs to RDF mapping
  - discourse referent -> blank node
  - unary condition -> RDF typing
  - binary condition -> RDF triple <s, **p**, o>
  - embedded DRSs -> RDF reification

- Vocabularies
  - closed: Boxer (event, agent, named, etc.)
  - open: WordNet senses, DBpedia URIs

LODifier

The New York Times reported that John McCarthy died.
He invented the programming language LISP.

```
| x0 x1 x2 x3 | | x4 x5 x6 |
|.............................| |.....................|
(| male(x0) |+| event(x4) |)
| named(x0,john_mccarthy,per) | | invent(x4) |
| programming_language(x1) | | agent(x4,x0) |
| nn(x1,x2) | | patient(x4,x2) |
| named(x2,lisp,nam) | | event(x5) |
| named(x3,new_york_times,org) | | report(x5) |
|_____________________________| | agent(x5,x3) |
| theme(x5,x6) |
| proposition(x6) |
| _____________ |
| | x7 | |
| x6:|..............| |
| | event(x7) | |
| | die(x7) | |
| | agent(x7,x0) | |
| |_____________| |
|_______________________|
```

```
_:var0x0 drsclass:named ne:john_mccarthy ;
rdf:type drsclass:male , foaf:Person ;
owl:sameAs dbpedia:John_McCarthy_(computer_scientist) .
_:var0x1 rdf:type class:programming_language ;
owl:sameAs dbpedia:Programming_language .
_:var0x2 drsrel:nn _:var0x1 .
_:var0x2 drsclass:named ne:lisp ;
owl:sameAs dbpedia:Lisp_(programming_language) .
_:var0x3 drsclass:named ne:the_new_york_times ;
owl:sameAs dbpedia:The_New_York_Times .
_:var0x4 rdf:type drsclass:event , wn30:wordsense-invent-verb-2 .
drsrel:agent _:var0x0 ; drsrel:patient _:var0x2 .
_:var0x5 rdf:type drsclass:event , wn30:wordsense-report-verb-3 ;
drsrel:agent _:var0x3 ; drsrel:theme _:var0x6 .
_:var0x6 rdf:type drsclass:proposition , reify:proposition , reify:conjunction ;
reify:conjunct [ rdf:subject _:var0x7 ;
rdf:predicate rdf:type ;
rdf:object drsclass:event . ]
reify:conjunct [ rdf:subject _:var0x7 ;
rdf:predicate rdf:type ;
rdf:object wn30:wordsense-die-verb-1 . ]
reify:conjunct [ rdf:subject _:var0x7 ;
rdf:predicate drsrel:agent ;
rdf:object _:var0x0 . ]
```

talnSRL

- Frame-based relation extraction from NL sentences
  - RDF graphs generates Linked Data graphs

- **Goal**: capture information from NL text as Linked Data, including NLP information

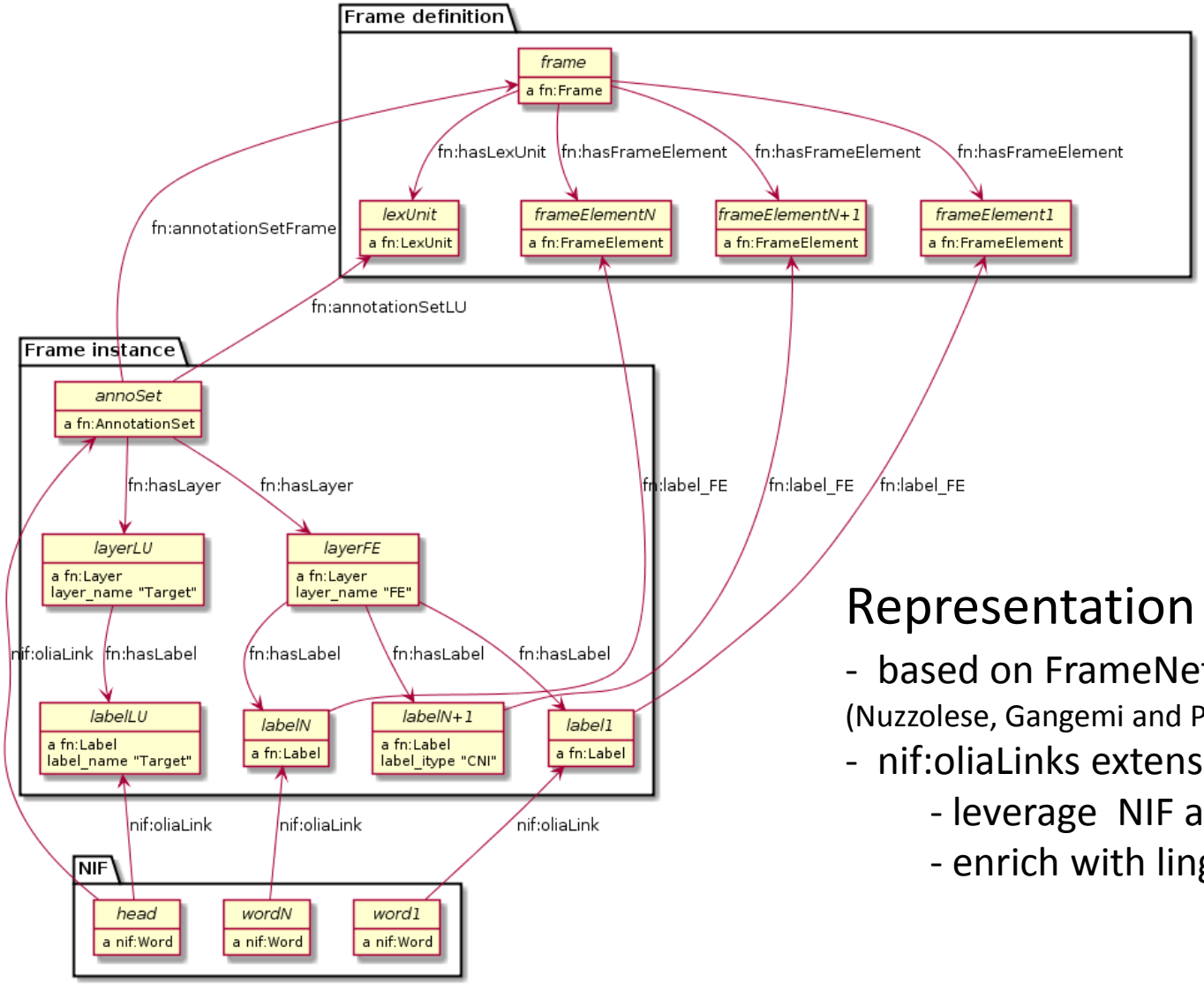
- Design principles
  - Dependency parsing
  - Linguistic frames
  - Linking to LOD cloud
  - Anchor textual annotations

- http://dparse.multisensor.taln.upf.edu/demoMS/relExtr/
- https://github.com/talnsoftware/FrameSemantics_Parser

talnSRL

- NLP and KE tasks
  - Dependency parsing & Semantic parsing : Mate Tools and SemLink mappings
  - NER, EL & WSD: Babelfy
- Vocabularies
  - OLIA
  - ITS, NERD, SKOS
  - FrameNet
  - BabelNet
  - NIF
- NL to SW
  - Porting to RDF

talnSRL



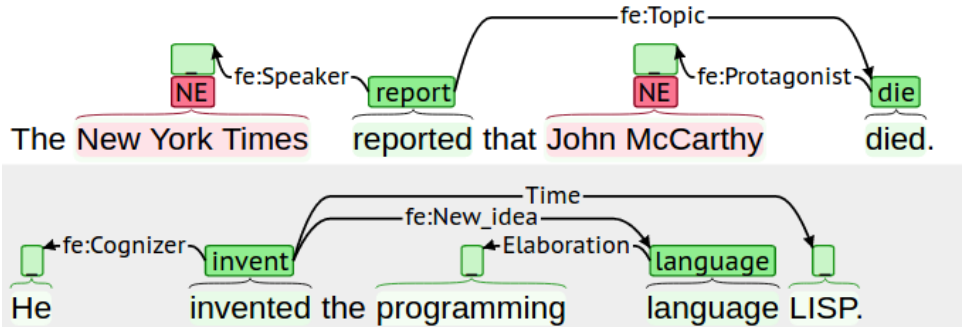
Representation model

- based on FrameNet v1.5 RDF

(Nuzzolese, Gangemi and Presutti 2011)

- nif:oliaLinks extensions
  - leverage NIF and OFN
  - enrich with linguistic information

talnSRL




```
:char= 20,28 a nif-core:Word ;
    nif-annotation:taIdentProv <http://babelfy.org/> ;
    nif-core:anchorOf "reported" ;
    nif-core:beginIndex "20" ;
    nif-core:endIndex "28" ;
    nif-core:referenceContext :char=0,97 ;
    nif-core:literalAnnotation "surf=spos=VV" , "deep=spos=VV" ... ;
    nif-core:oliaLink penn:VBD , penn-dep-syntax:root , taln:ROOT ,
                    :char=20,28_annoSet ;
    its:taClassRef ms:GenericConcept ;
    its:taIdentRef <http://babelnet.org/rdf/s00092823v> ;

:char= 20,28_annoSet a ofn-tbox:AnnotationSet ;
    ofn-tbox:annotationSetFrame ofn-frame:Statement ,
                                taln-frame:Linguistic_situation ;
    ofn-tbox:annotationSetLU ofn-abox-lu:report ;
    ofn-tbox:hasLayer :char=20,28_FEs ;

:char= 20,28_FEs a ofn-tbox:Layer ;
    ofn-tbox:layer_name "FE" ;
    ofn-tbox:hasLabel :char=5,19_fe , char=48,52_fe ;

:char=5,19_fe  a ofn-tbox:Label ;
    ofn-tbox:label_FE taln-FE:Argument0.linguistic_situation ,
                      ofn-fe:Speaker.statement ;
```

PIKES

- Frame-based KE for English texts; two phases
  - linguistic features extraction
  - knowledge distillation

- **Goal**: Decouple NLP from knowledge distillation

- Design principles
  - Dependency parsing
  - FrameNet frames
  - Three-layer representation
    - Anchor textual annotations

- http://pikes.fbk.eu/

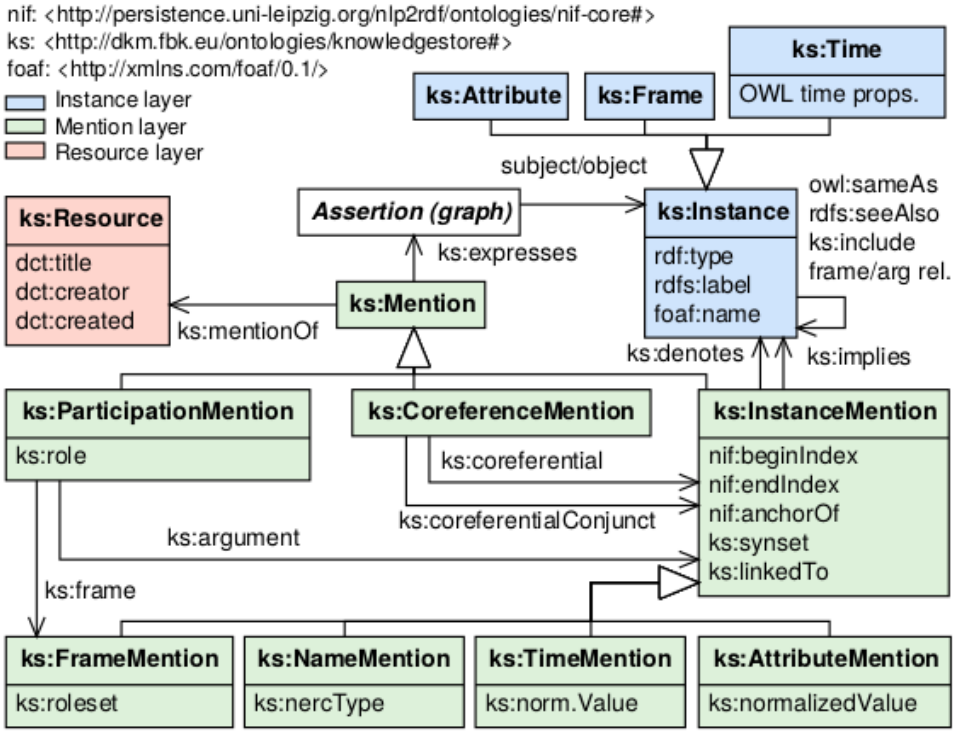
PIKES

- NLP and KE tasks:
  - Dependency parsing & SRL: Mate Tools and Predicate Matrix mappings
  - NER and coreference resolution: Stanford CoreNLP
  - EL: DBpedia Spotlight
  - WSD: UKB
- Vocabularies
  - PropBank, NomBank, VerbNet, FrameNet
  - SEM, SUMO, FOAF, KnowledgeStore
  - DBpedia, YAGO
  - NIF
- NL to SW
  - Re-engineering using SPARQL-like mapping rules
  - Post-processing for materialising and compacting the resulting KB

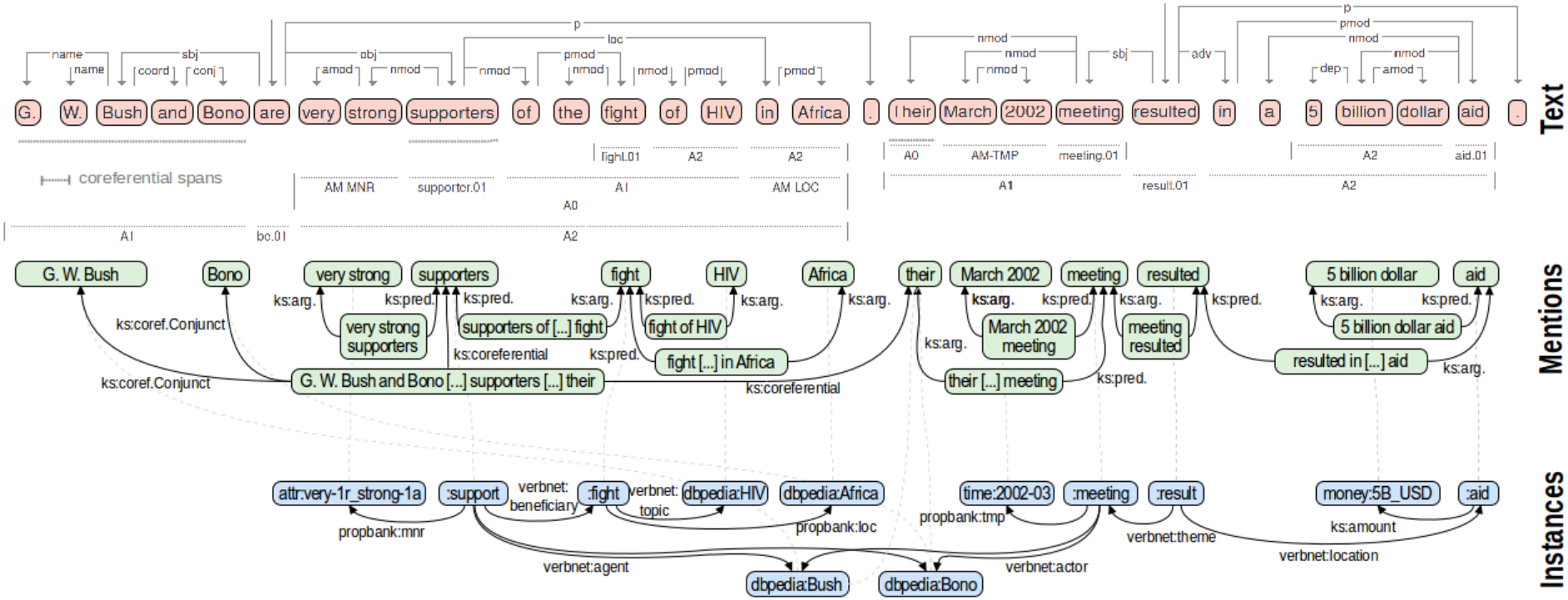
PIKES

- Three layer representation model

- Inter-layer relations
  - denotes
  - implies
  - expresses

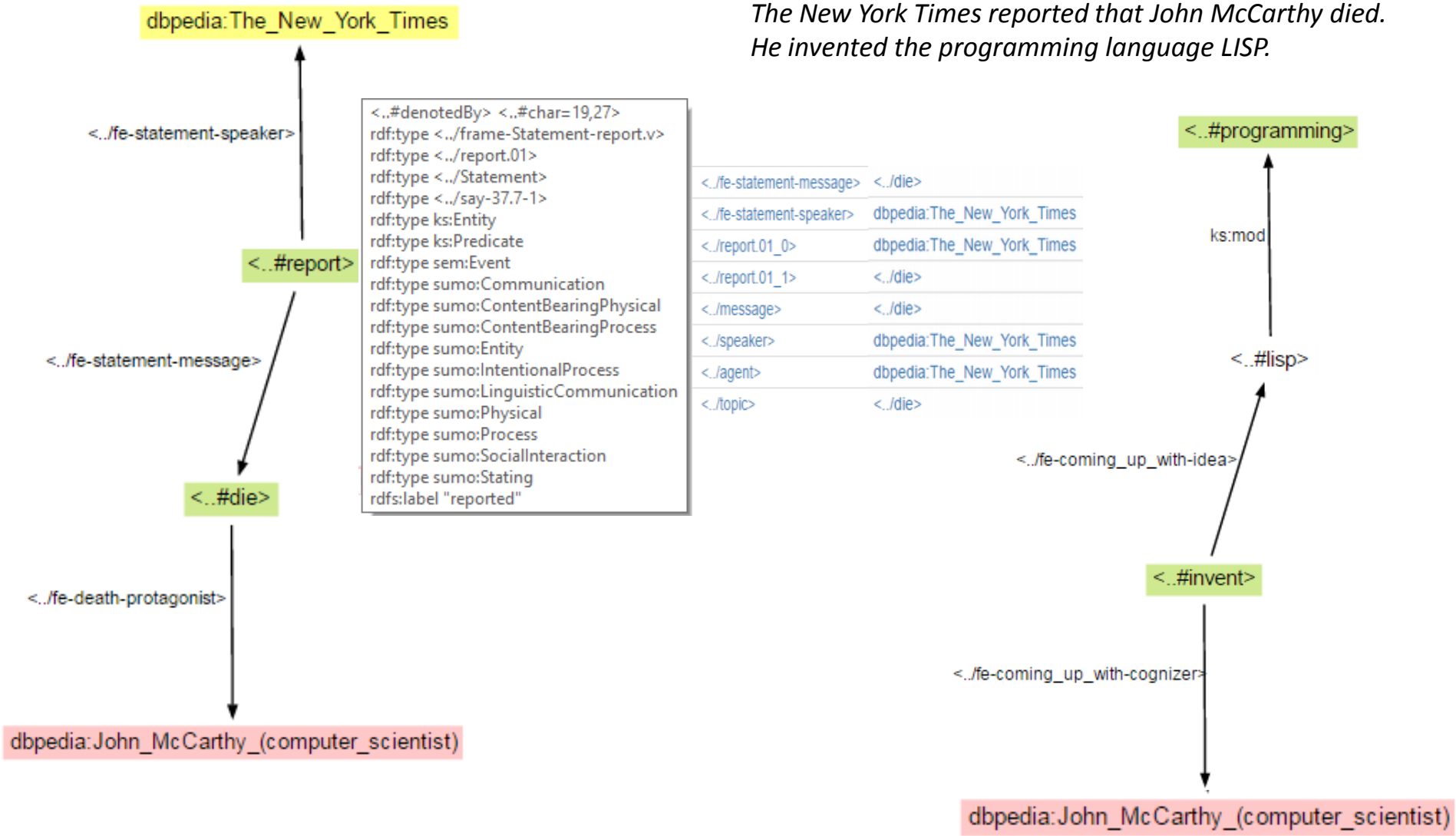


PIKES


Text
Mentions
Instances
G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa . their March 2002 meeting resulted in a 5 billion dollar aid .
coreferential spans
G. W. Bush
Bono
very strong
supporters
fight
HIV
Africa
their
March 2002
meeting
resulted
5 billion dollar
aid
very strong supporters
supporters of [...] fight
fight of HIV
fight [...] in Africa
G. W. Bush and Bono [...] supporters [...] their
March 2002 meeting
meeting resulted
5 billion dollar aid
resulted in [...] aid
their [...] meeting
ks:coref.Conjunct
ks:coreferential
ks:arg.
ks:pred.
attr:very-1r_strong-1a
:support
:fight
dbpedia:HIV
dbpedia:Africa
time:2002-03
:meeting
:result
money:5B_USD
:aid
dbpedia:Bush
dbpedia:Bono
verbnet: beneficiary
verbnet: topic
propbank:mnr
propbank:loc
propbank:tmp
verbnet:theme
verbnet:location
ks:amount
verbnet:agent
verbnet:actor

PIKES

The New York Times reported that John McCarthy died.
He invented the programming language LISP.

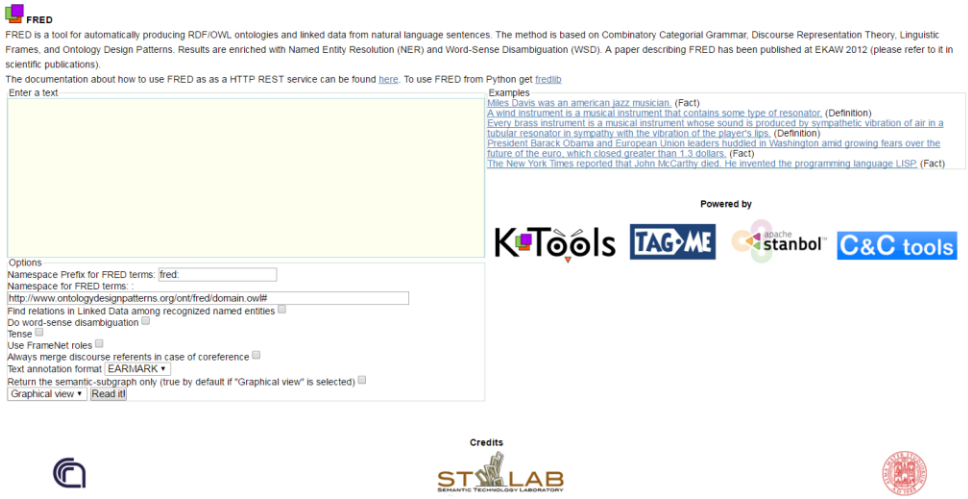


FRED

- Machine reader for the SW
  - produces RDF/OWL ontologies and Linked Data from natural language sentences

- ***Goal**: maximise modelling choices wrt Semantic Web practices*

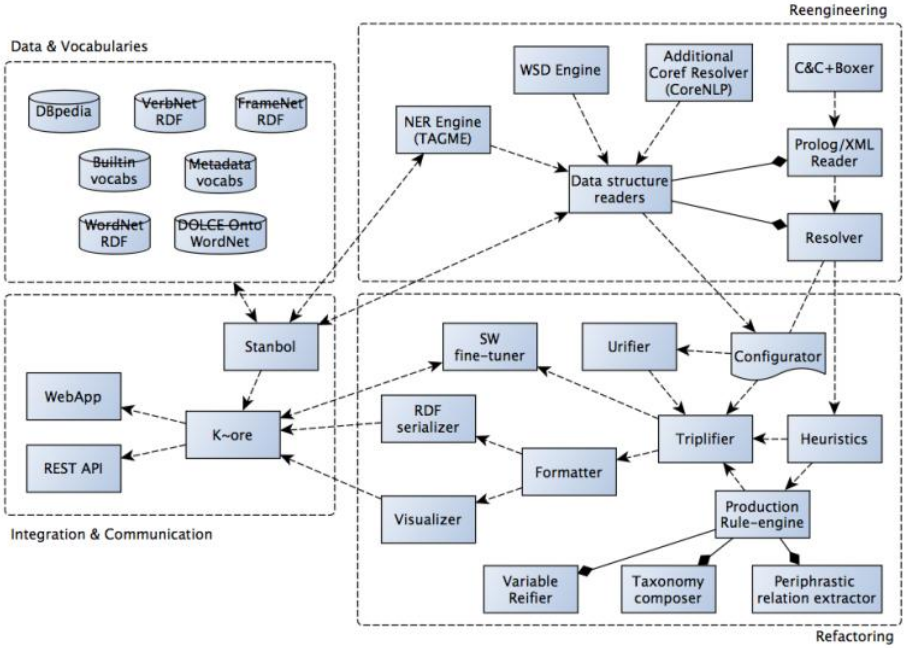
- Design principles
  - CCG & DRS
  - FrameNet frames
  - Ontology design patterns
  - Linking to LOD cloud
  - Anchor textual annotations



- http://wit.istc.cnr.it/stlab-tools/fred

FRED

- NLP and KE tasks
  - Syntactic parsing: CCG
  - Semantic parsing: Boxer + SemLink
  - Coreference resolution: CoreNLP
  - NER & EL: Stanbol, TagMe
  - WSD: UKB
- Vocabularies
  - LMM, DUL
  - VerbNet and FrameNet
  - DBpedia, WordNetRDF, OntoWordNet
  - NIF, EARMAR for annotation anchoring
  - POS, quantifiers, Boxer & FRED built-ins
- NL to SW
  - re-engineering & refactoring (*Semion*)

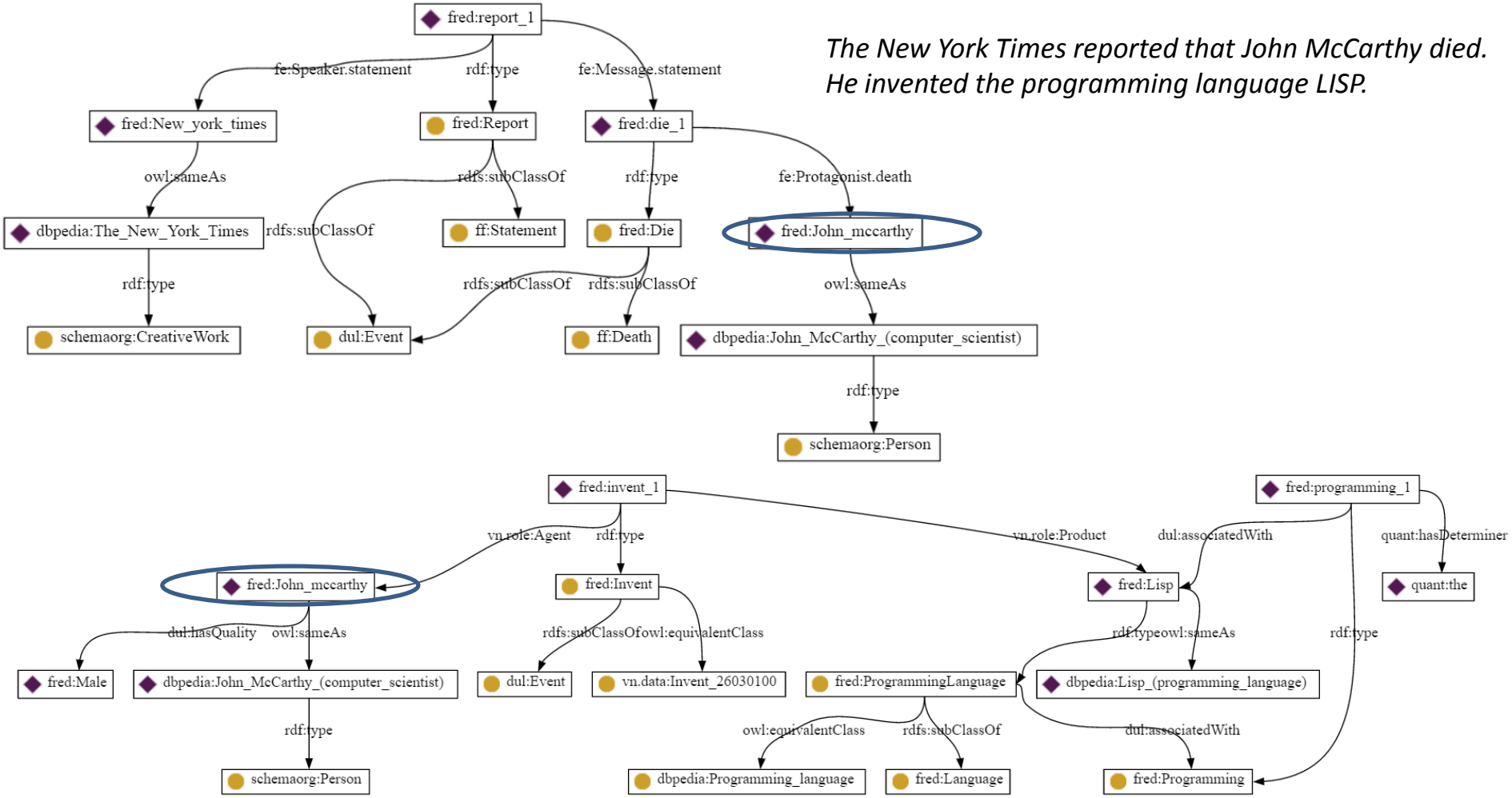


FRED

- N-ary logical patterns
  - events (role labelling) -> dul:Event | boxing -> dul:Situation
- Compositional semantics, taxonomy induction
  - “The programming language LISP...” -> ProgrammingLanguage(LISP) & subclassOf(ProgrammingLanguage,Language)
  - “John, the mayor of the city, ...” -> mayorOf(John, city)
- RDF-oriented representation for modality and negation
  - “John doesn’t like apples” -> hasTruthValue(like1, false)

FRED

The New York Times reported that John McCarthy died.
He invented the programming language LISP.



Tools Overview

	LODifier	talnSRL	PIKES	FRED
Modelling	porting to RDF	porting to RDF	re-engineering	re-engineering, maximises
Textual annotations grounding	-	NIF	NIF	NIF, EARMARK
Semantic parsing	DRS with Boxer	Dependencies with MTT DSynt + SemLink mappings	Dependencies with Mate tolos SLR + Predicate Matrix mappings	DRS with Boxer + SemLink mappings
Vocabularies	• WordNet, DBpedia	• OLiA, ITS, NERD • SKOS • OFN • BabelNet	• SUMO, YAGO2, SEM, DCMI, FOAF • PreMOn • DBpedia	• DUL, LMM, • OFN • DBpedia, WordNet, OntoWN
Coreference	√	-	√	√
Multi-linguality	only English	only English	only English	√

Summary

- Two perspectives
  - porting linguistic resources and annotations
    - interlinking, interoperability, sharing & reuse
  - capturing distilled knowledge
    - automated reasoning (OL&P, QA, semantic search/retrieval, etc.)

- Translating linguistic representations to ontological ones involves non-trivial choices
  - linguistic vs knowledge engineering considerations
  - dependencies on application context

PART IV – Applications & Evaluation

Outline

- Applications
  - Question Answering over SW data
  - Abstractive summarization
- Evaluation
  - Intrinsic vs Extrinsic
  - KE tools comparison (Gangemi 2013)
  - Individual Tool Evaluation
    - LODifier, FRED, PIKES, SMATCH

Applications

- Capturing information from NL texts as structured representations is relevant to a plethora of applications
  - Ontology population & learning
  - Semantic search & retrieval
  - Linked Data publishing
  - Opinion and sentiment analysis
  - **Question Answering**
  - Natural Language generation-oriented tasks
    - ontology (axioms) verbalisation
    - delivery of RDF/OWL content
    - **abstractive summarisation of RDF/OWL content**
  - ....

Question Answering (I)

- Structured knowledge made available in RDF/OWL KBs keeps growing fast!!!
  - LOD cloud > 30 billion RDF triples
  - public knowledge bases (e.g. DBpedia, Freebase)
  - proprietary KBs (clinical, daily activities, smart homes, etc.)
- NL interfaces for Question Answering
  - intuitive paradigms for accessing & querying
  - hide the complexity of formal representation/query languages

Question Answering (II)

- **Goal:** translate the NL user questions into structured queries and interpret them against the underlying KB

- **Challenge:** bridge the gap between the way users communicate with the system and the way domain knowledge is captured
  - formalise and interpret user questions
    - conceptual granularity mismatches
    - meaning variations

Question Answering (III)

- Most systems focus primarily on simple factoid questions
  - *"Who is the daughter of Robert Kennedy married to?"*
  - *"Who created Goofy?"*

and (more recently) questions including superlatives, quantification, etc.
  - "What is the second *highest* mountain on Earth?"
  - "Give me *all* cities in Germany."
  - "Which countries have *more than two* official languages?"

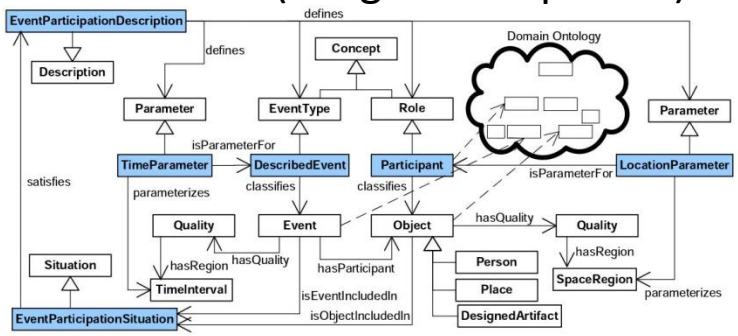
- I.e. questions with **binary relations** & **"light" linguistic constructions**
  - the more complex the linguistic constructions, the more challenging the question analysis (syntactic attachment & scope ambiguities, etc.)

- Example systems
  - triple-based serialisation: AquaLog, PowerAqua, NLP-reduce, FREyA, etc.
  - compositional semantics: Pythia, etc.

Question Answering (IV)

- Little support for QA over conceptually rich KBs capturing n-ary relational contexts
  - e.g. DOLCE + DnS ontology design patterns

(Ansgar Scherp et al.)



- Example domains
  - *event modelling*

    e.g. Event-Model-F patterns for participation, mereology, causality, correlation, etc.
  - context-aware activity/behaviour modelling

Question Answering - Evaluation

- Question Answering over Linked Data (QALD)
  - http://qald.sebastianwalter.org
  - set of questions along with corresponding answers and manually specified SPARQL queries
  - evaluation metrics: precision & recall

Abstractive summarisation

- Extractive summarization
  1. Splits source texts into fragments, e.g. sentences.
  2. Evaluates relevance of each.
  3. Produces summary by selecting most relevant fragments.
- Abstractive summarization
  1. Analyses source texts using NLP and IE methods.
  2. Creates an intermediate **representation** of contents.
  3. Uses **natural language generation** methods to produce a summary

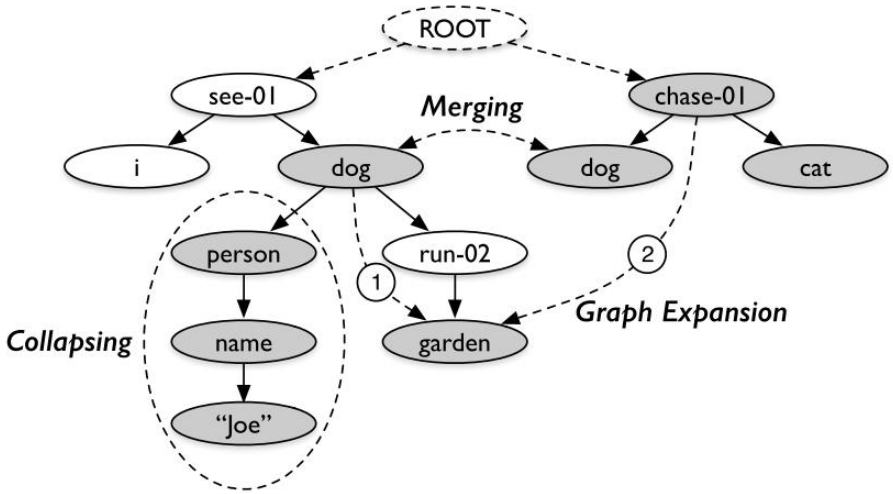
Abstractive summarization

- Intermediate representations have different degrees of abstraction:
  - Abstract Meaning Representation, e.g. Liu et al. 2015
  - Discourse trees, e.g. Gerani et al. 2014
  - Dependency trees, e.g. Cheung and Penn 2014
  - Word co-occurrence graphs, e.g. Ganesan et al. 2010
- Most abstractive summarizers use linguistic semantic structures
  - Yet ontological representations open the door to reasoning!
    - E.g. Temporal reasoning can be used to order events in the summary.
    - E.g. Reasoning about models of user profiles and preferences can be used to create tailored summaries.

Abstractive summarization with AMR structures (Liu et al. 2015)

- Resources:
  - Statistical AMR parser JAMR trained on AMR Bank.
  - Dataset of documents, summaries and gold-standard AMR annotations of both (AMR Bank).

- AMR sentence graphs are merged into a single document graph by collapsing common NE nodes.
- Summary is a subgraph of the document graph.
- Summarization addressed through features on the graph with scores estimated from the dataset using ILP.



Sentence A: I saw Joe's dog, which was running in the garden.

Sentence B: The dog was chasing a cat.

Outline

- Applications
  - Question Answering over SW data
  - Abstractive summarization
- Evaluation
  - Intrinsic vs Extrinsic
  - KE tools comparison (Gangemi 2013)
  - Individual Tool Evaluation
    - LODifier, FRED, PIKES, SMATCH

Evaluation

- How to evaluate a relation extraction system?
- Intrinsic vs extrinsic evaluation
  - Extrinsic evaluation
    - performance of downstream application as indicator of quality
      - e.g. question answering, ontology population, summarization.
  - Intrinsic evaluation
    - compare extracted relations to a manually annotated gold standard in terms of precision and recall

KE tools comparison (Gangemi 2013)

- Unified framework for comparing multiple KE tools
- A reference knowledge space is created by converting the results of each tool to RDF/OWL expressions and merging the results.
- Evaluation metrics: precision, recall, f-score and accuracy

Tool	Topics	NER	NE-RS	TE	TE-RS	Senses	Tax	Rel	Roles	Events	Frames
AIDA	–	+	+	–	–	+	–	–	–	–	–
Alchemy	+	+	–	+	–	+	–	+	–	–	–
Apache Stanbol	–	+	+	–	–	+	–	–	–	–	–
CiceroLite	–	+	+	+	+	+	–	+	+	+	+
DB Spotlight	–	+	+	–	–	+	–	–	–	–	–
FOX	+	+	+	+	+	+	–	–	–	–	–
FRED	–	+	+	+	+	+	+	+	+	+	+
NERD	–	+	+	–	–	+	–	–	–	–	–
Open Calais	+	+	–	–	–	+	–	–	–	+	–
PoolParty KD	+	–	–	–	–	–	–	–	–	–	–
ReVerb	–	–	–	–	–	–	–	+	–	–	–
Semiosearch	–	–	+	–	+	–	–	–	–	–	–
Wikimeta	–	+	–	+	+	+	–	–	–	–	–
Zemanta	–	+	–	–	–	–	–	–	–	–	–

Table 2: Summary of featured basic tasks (as obtained from testing)

Topic	<Document> dc:subject <Topic>
Named entity	owl:NamedIndividual
Entity resolution (NE)	owl:sameAs
Entity coreference	owl:sameAs
Term	owl:Class || owl:ObjectProperty || owl:DatatypeProperty
Sense tag	owl:NamedIndividual rdf:type owl:Class
Sense disambiguation (classes)	owl:equivalentClass
Taxonomy (subclasses)	owl:subClassOf
Extracted (binary) relation	owl:ObjectProperty || owl:DatatypeProperty
Semantic role	owl:ObjectProperty || owl:DatatypeProperty
Event	<Event> rdf:type <Event.type> . <Event> <semrole$_i$> <Entity$_j$>
Frame	<Event.type> owl:subClassOf <Frame>
Restriction	owl:Restriction
Linked entities	owl:sameAs || owl:differentFrom
Conjunct of individuals	owl:NamedIndividual
Disjunction of individuals	owl:NamedIndividual
Factual entailment	<Event$_1$> <dependency> <Event$_2$>

Table 1: Translation table, used when default assumptions are to be applied on the results of a tool. The output of basic tasks not listed here is trivially translated according to model-theoretic semantics (e.g. "union of classes").

Comparison of KE tools (Gangemi 2013)

- Comparison of n-ary relations as overlap of unlabeled binary relations.

"it plans to blacklist the Nusra Front as a terrorist organization"

– (ReVerb): no extraction
– (Alchemy): `plans to blacklist(it,the Nusra Front as a terrorist organization)`
– (CiceroLite): `plans to blacklist(it,front,(AS) a terrorist organization)`
– (FRED): `experiencer(plan_3,United_States) ; theme(plan_3,blacklist_3) ; agent(blacklist_3,United_States) ; patient(blacklist_3,NusraFront) ; as(blacklist_3,organization_3) ; TerroristOrganization(organization_3)`

RelEx Tool	p	r	$F1$	a
Alchemy	.69	.25	.37	.30
CiceroLite	.90	.20	.33	.25
FRED	.84	.82	.83	.82
ReVerb	.67	.23	.34	.27

Table 9: Comparison of relation extraction tools.

- "There are important cohesion aspects here that are hardly caught by means of simple triple patterns"

Evaluation of LODifier

- Extrinsic evaluation: document similarity task
  - subset of 183 document pairs belonging to same topic, taken from TDT-2 benchmark dataset
  - baseline: set of similarity measures between documents that ignore relations: random, bag-of-words, bag-of-URI (NEs and WordNet)

Evaluation of LODifier

- Similarity measures based on similarity between graphs extracted by LODifier for each document
  - graph isomorphism too computationally expensive
  - approximated similarity using overlap in set of shortest paths between relevant entities

Model	normal	extended
Similarity measures without structural knowledge		
Random Baseline	50.0	–
Bag of Words	63.0	–
Bag of URIs (Variant 1)	61.6	75.1
Bag of URIs (Variant 2)	70.6	76.0
Bag of URIs (Variant 3)	**73.4**	**76.4**
Similarity measures with structural knowledge		
$\text{proSim}_{\text{cnt}}$ (k=6, Variant 1)	79.0	78.9
$\text{proSim}_{\text{cnt}}$ (k=6, Variant 2)	80.3	80.3
$\text{proSim}_{\text{cnt}}$ (k=6, Variant 3)	81.6	81.6
$\text{proSim}_{\text{cnt}}$ (k=8, Variant 1)	77.7	77.6
$\text{proSim}_{\text{cnt}}$ (k=8, Variant 2)	79.2	79.0
$\text{proSim}_{\text{cnt}}$ (k=8, Variant 3)	**82.1**	**81.9**
$\text{proSim}_{\text{len}}$ (k=6, Variant 3)	81.5	81.4
$\text{proSim}_{\text{len}}$ (k=8, Variant 3)	80.3	80.1
$\text{proSim}_{\text{len}}$ (k=10, Variant 3)	80.0	79.8
$\text{proSim}_{\text{sqlen}}$ (k=6, Variant 3)	80.4	80.4
$\text{proSim}_{\text{sqlen}}$ (k=8, Variant 3)	81.1	80.9
$\text{proSim}_{\text{sqlen}}$ (k=10, Variant 3)	80.5	80.4

Evaluation of FRED

- Intrinsic evaluation
  - gold standard of 1214 sentences randomly selected from FrameNet annotated corpus; one frame per sentence; only verbs
  - given a sentence, when a predicted frame matches gold-standard full score (1), when semantically close partial score (0.8)
  - metrics: precision & recall of extracted frames
    - comparison against *Semafor*

Tool	Precision	Recall	F-Score
Boxer	75.320	57.519	65.227
Semafor	75.325	74.797	75.060

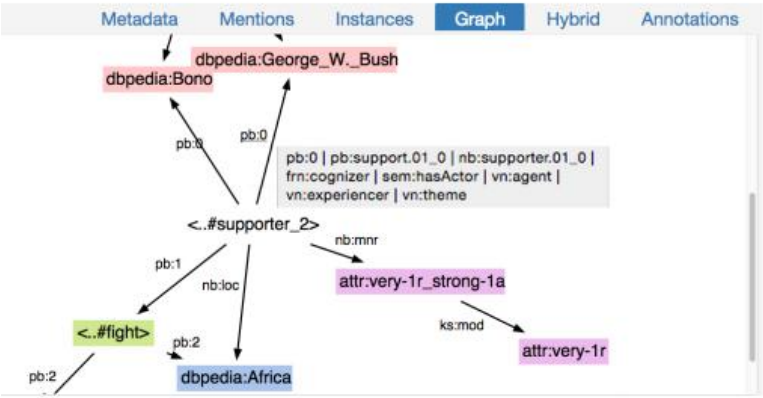
Evaluation of PIKES

- Intrinsic evaluation
  - gold standard of 8 sentences manually annotated by two annotators
  - metrics: precision and recall in extracted entities, frames attributes, and their relations

  - Gold and system outputs seen as a graph G where:
    - nodes: entities, frames, attributes annotated and matched against entries in DBpedia and types in VerbNet, FrameNet and PropBank
    - edges: coreference relations, instance-attribute associations, and frame-argument participation with VN/FN/PB/NB thematic roles

Evaluation of PIKES against gold

- Evaluations for each component of G and G_S:
  - Nodes/Instances
  - (unlabeled) Edges
  - (labeled) Triples
- Triples are considered divided by category:
  - Links to DBpedia, VN/FN/PB/NB classes, VN/FN/PB/NB participation relations, coreference relations
- True positives, false positives, true negatives and false negatives over sets of components



Component	N_G	P	R	F_1
Instances	153	.943	.967	.955
Triples	596	.713	.508	.594
DBpedia links	18	.700	.778	.737
types (VN)	44	.706	.545	.615
types (FN)	53	.731	.358	.481
types (PB)	53	.844	.717	.776
types (NB)	37	.690	.784	.734
roles (VN)	94	.742	.489	.590
roles (FN)	108	.500	.259	.341
roles (PB)	119	.829	.571	.677
roles (NB)	55	.627	.582	.604
`owl:sameAs`	15	.714	.333	.455
Edges	171	.891	.766	.824

(b)

Evaluation of PIKES against FRED

- A modified gold graph G' is produced without the information that FRED doesn't produce:
  - PropBank and NomBank types and roles
  - FrameNet roles
  - Nominal predicates
- An additional graph G'' is produced by merging the results of FRED and PIKES

Component	$N_{G'}$	FRED vs G'			PIKES vs G'			$N_{G''}$	FRED vs G''			PIKES vs G''		
		P	R	F_1	P	R	F_1		P	R	F_1	P	R	F_1
Instances	137	.930	.869	.898	.937	.978	.957	136	.930	.875	.902	.937	.985	.961
Triples	166	.543	.416	.471	.713	.554	.624	115	.543	.600	.570	.713	.800	.754
DBpedia links	18	.615	.444	.516	.700	.778	.737	14	.615	.571	.593	.700	1.000	.824
types (VN)	31	.593	.516	.552	.667	.581	.621	27	.593	.593	.593	.667	.667	.667
types (FN)	26	.550	.423	.478	.762	.615	.681	17	.550	.647	.595	.762	.941	.842
roles (VN)	76	.547	.382	.450	.722	.513	.600	51	.547	.569	.558	.722	.765	.743
`owl:sameAs`	15	.357	.333	.345	.714	.333	.455	6	.357	.833	.500	.714	.833	.769
Edges	155	.869	.555	.677	.937	.768	.844	134	.869	.642	.738	.937	.888	.912

(c)

SMATCH

- AMR evaluation metric as degree of overlap between two whole-sentence semantic structures
- Measures precision, recall, and f-score of the triples in the second AMR against the triples in the first AMR, i.e., the amount of propositional overlap
- Conjunction of logical propositions, or triples:
  - *Maximum f-score obtainable via a one-to-one matching of variables between the two AMRs*

```
instance(a, want-01)  ∧
instance(b, boy)  ∧
instance(c, go-01)  ∧
ARG0(a, b)  ∧
ARG1(a, c)  ∧
ARG0(c, b)
```

```
instance(x, want-01)  ∧
instance(y, boy)  ∧
instance(z, football)  ∧
ARG0(x, y)  ∧
ARG1(x, z)
```

	M	P	R	F
x=a, y=b, z=c:	4	4/5	4/6	0.73
x=a, y=c, z=b:	1	1/5	1/6	0.18
x=b, y=a, z=c:	0	0/5	0/6	0.00
x=b, y=c, z=a:	0	0/5	0/6	0.00
x=c, y=a, z=b:	0	0/5	0/6	0.00
x=c, y=b, z=a:	2	2/5	2/6	0.36
smatch score:				0.73

"the boy wants to go" "the boy wants the football"

Conclusions

- NLP and SW are starting to get along better
  - yet, still a long way to go.....

- Challenges
  - Performance of NLP tools
    - error propagation
  - No principled way of converting NL structures to OWL
    - no straightforward mappings between different conversions
  - Evaluation is not straightforward

QUESTIONS?

This tutorial is supported by the European commission under the contract numbers
FP7-ICT-610411 and H2020-645012-RIA

References (I)

- Androutsopoulos, I., Lampouras, G., and Galanis, D. (2013). Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System. In J. Artif. Intell. Res. (JAIR) 48, pp. 671-715.
- Augenstein, I., Padó, S., and Rudolph, S. (2012). LODifier: Generating Linked Data from Unstructured Text. In Extended Semantic Web Conference (ESWC'12), pp. 210-224.
- Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In Proceedings of the 17th international conference on Computational linguistics-Volume 1, pp. 86-90. Association for Computational Linguistics.
- Ballesteros, M., Bohnet, B., Mille, S., Wanner, L. (2015). Data-driven deep dependency parsing. In Natural Language Engineering, 1-36.
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., and Schneider, N. (2012). Abstract meaning representation (AMR) 1.0 specification. In Parsing on Freebase from Question-Answer Pairs." In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: ACL (pp. 1533-1544).
- Barabucci, G. Di Iorio, A., Peroni, S., Poggi, F., and Vitali. F. (2013). Annotations with EARMARK in practice: a fairy tale. In Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE '13). ACM, New York, NY, USA, , Article 11 , 8 pages.
- Björkelund, A., Bohnet, B., Hafdell, and L., Nugues, P. (2010). A high-performance syntactic and semantic dependency parser. In Coling 2010: Demonstration Volume, pp. 33-36, Beijing.
- Bontcheva, K. (2005). Generating tailored textual summaries from ontologies. In European Semantic Web Conference (ESWC'05), pp. 531-545.
- Bos, J. (2008). Wide-coverage semantic analysis with boxer. In Proceedings of the 2008 Conference on Semantics in Text Processing (pp. 277-286). Association for Computational Linguistics.
- Bouayad-Agha, N., G. Casamayor, S. Mille, M. Rospocher, H. Saggion, L. Serafini and L. Wanner. (2012). From Ontology to NL: Generation of Multilingual User-Oriented Environmental Reports. In Natural Language Processing and Information Systems, Springer.
- Bouayad-Agha, N., Casamayor, G., Díez, F., Mille, S. and Wanner, L. (2012). Perspective-based Generation of Football Matches Summaries: Old Tasks, New Challenges. In ACM Transactions on Speech and Language Processing, Vol.9 Issue 2, July 2012.
- Cai, S., & Knight, K. (2013, August). Smatch: an Evaluation Metric for Semantic Feature Structures. In Proceedings of the American Computational Lingusitics Conference (ACL 2013), p. 748.

References (II)

- Clark, S., and Curran, J. R. (2007). Wide-coverage efficient statistical parsing with CCG and log-linear models. Computational Linguistics, 33(4), pp. 493-552.
- Cheung, J. C. K., & Penn, G. (2014). Unsupervised Sentence Enhancement for Automatic Summarization. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 775–786). Doha, Qatar: Association for Computational Linguistics.
- Christian Chiarcos, Sebastian Hellmann and Sebastian Nordhoff. 2012. Linking linguistic resources: Examples from the Open Linguistics Working Group, In: Christian Chiarcos, Sebastian Nordhoff and Sebastian Hellmann (eds.), Linked Data in Linguistics. Representing Language Data and Metadata, Springer, Heidelberg, p. 201-216.
- Cimiano, P., and Reyle, U. (2006). Towards Foundational Semantics - Ontological Semantics Revisited. In International Conference on Formal Ontology in Information Systems (FOIS'06), pp. 51-62.
- Cimiano, P., Haase, P., and Heizmann, J. (2007). Porting natural language interfaces between domains: an experimental user study with the ORAKEL system. In Inter. Conference on Intelligent User Interfaces (IUI'07), pp. 180-189.
- Corcoglioniti, F., Rospocher, M., and Aprosio., A. P. (2016). A 2-phase frame-based knowledge extraction framework. In Proc. of ACM Symposium on Applied Computing (SAC'16), pp. 354-361.
- Corcoglioniti, F., Rospocher, M., Aprosio, A. P., & Tonelli, S. (2016). PreMOn: a Lemon Extension for Exposing Predicate Models as Linked Data. In Proceedings of the 10th language resources and evaluation conference (LREC2016), Portoroz (Slovenia).
- Das, D., Schneider, N., Chen, D., & Smith, N. A. (2010). SEMAFOR 1.0: A probabilistic frame-semantic parser. Language Technologies Institute, School of Computer Science, Carnegie Mellon University.
- De Lacalle, M. L., Laparra, E., and Rigau, G. (2014). Predicate Matrix: extending SemLink through WordNet mappings. In LREC'14, pp. 903-909.
- De Marneffe, M. C., and Manning, C. D. (2008). The Stanford typed dependencies representation. In Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation, pp. 1-8. Association for Computational Linguistics.
- Draicchio, F., Gangemi, A., Presutti, V., & Nuzzolese, A. G. (2013). FRED: from natural language text to RDF and OWL in one click. In The Semantic Web: ESWC 2013 Satellite Events (pp. 263-267). Springer Berlin Heidelberg.
- Dyer, C., Ballesteros, M., Ling, W., Matthews, A., and Smith, N. A. (2015). Transition-based dependency parsing with stack long short-term memory. arXiv preprint arXiv:1505.08075.

References (III)

- Filip, D., McCance, S., Lewis, D., Lieske, C., Lommel, A., Kosek, J., & Savourel, Y. (2013). Internationalization Tag Set (ITS) Version 2.0, W3C Proposed Recommendation 24 September 2013.
- Franconi, E., Gardent, C., Juarez-Castro, X. I., and Perez-Beltrachini, L. (2014). Quelo Natural Language Interface: Generating queries and answer descriptions. In Natural Language Interfaces for Web of Data Workshop (NLIWoD'14).
- Ganesan, K., Zhai, C., & Han, J. (2010). Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10).
- Gangemi, A. (2005). Ontology Design Patterns for Semantic Web Content. In Internationl Semantic Web Conference (ISWC'05), pp. 262-276.
- Gangemi, A. (2010). What's in a Schema? In C. Huang, N.Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prevot, editors, Ontology and the Lexicon. Cambridge University Press, 2010.
- Gangemi, A. (2013). A Comparison of Knowledge Extraction Tools for the Semantic Web. In Extended Semantic Web Conference (ESWC'13), pp. 351-366.
- Hellmann, S., Lehmann, J., Auer, S., and Brümmer, M. (2013). Integrating NLP using Linked Data . 12th International Semantic Web Conference, 21-25 October 2013, Sydney, Australia.
- Johansson, R., and Nugues, P. (2007). Extended constituent-to-dependency conversion for English. In 16th Nordic Conference of Computational Linguistics, pp. 105-112.
- Kamp, H., Van Genabith, J., and Reyle, U. (2011). Discourse representation theory. In Handbook of philosophical logic, pp. 125-394.
- Kingsbury, P., and Palmer, M. (2002). From TreeBank to PropBank. In proceedings of LREC'02.
- Klein D. and Manning, C. D. (2003). Accurate Unlexicalized Parsing. In Proceedings of the 41st Meeting of the Association for Computational Linguistics (ACL'03), pp. 423-430.

References (IV)

- Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., & Cheney, J. (2013). PROV-O: The PROV Ontology. W3C Recommendation, 30 April 2013. World Wide Web Consortium.
- Flanigan, J., Liu, F., Sadeh, N.M., Smith, N.A., & Thomson, S.. (2015). Toward Abstractive Summarization Using Semantic Representations. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2015), Denver, Colorado.
- Lopez, V., Uren, V., Sabou, M., and E. Motta (2011). Is question answering for the Semantic Web?: A survey. In Semantic Web, 2(2), pp. 125-155.
- Lopez, V., Fernández, M., Motta, E., and Stieler, N. (2012). PowerAqua: Supporting users in querying and exploring the Semantic Web. In Semantic Web 3(3), pp. 249-265.
- Martins, A. F., Smith, N. A., Xing, E. P., Aguiar, P. M., and Figueiredo, M. A. (2010). Turbo parsers: Dependency parsing by approximate variational inference. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (pp. 34-44). Association for Computational Linguistics.
- McCrae, J. P., Spohr, D., Cimiano, P., (2011). Linking Lexical Resources and Ontologies on the Semantic Web with Lemon. In Extended Semantic Web Conference (ESWC'11), pp. 245-259.
- Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The NomBank project: An interim report. In HLT-NAACL 2004 workshop: Frontiers in corpus annotation, pp. 24-31.
- Moro, A., Raganato, A., Navigli, R., Informatica, D., & Elena, V. R. (2014). Entity Linking meets Word Sense Disambiguation : a Unified Approach. ACL, 2, 231–244.
- Nuzzolese, A. G., Gangemi, A., Presutti, V. (2011). Gathering lexical linked data and knowledge patterns from FrameNet. In International Conference on Knowledge Capture (K-CAP'11), pp. 41-48.
- Ovchinnikova, E., Vieu, L., Oltramari, A., Borgo, S., and Alexandrov, T. (2010). Data-driven and ontological analysis of FrameNet for natural language reasoning. In LREC'10.

References (V)

- Palmer, M. (2009). Semlink: Linking PropBank, VerbNet and frameNet. In Proceedings of the Generative Lexicon Conference (pp. 9-15).
- Picca, D., Gliozzo, A. M., and Gangemi, A (2008). LMM: an OWL-DL MetaModel to Represent Heterogeneous Lexical Knowledge. In LREC'08.
- Presutti, V., Draicchio, F., Gangemi, A. (2012). In Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames (EKAW 2012), pp. 114-129.
- Rizzo, G., & Troncy, R. (2012, April). NERD: a framework for unifying named entity recognition and disambiguation extraction tools. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 73-76). Association for Computational Linguistics.
- Rouces, J., de Melo, G. and Hose, K. (2015). FrameBase: Representing N-Ary Relations Using Semantic Frames. In Extended Semantic Web Conference (ESWC'15), pp. 505-521.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Scheffczyk, J. (2010). FrameNet II: Extended Theory and Practice.
- Schuler, K. K. (2005). VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon. Ph.D. thesis, Univ. of Pennsylvania.
- Soler, J., Ballesteros, M., Bohnet, B., Mille, S., Wanner, L. (2015). Visualizing Deep-Syntactic Parser Output. In Proceedings of NAACL-HLT-DEMOS, Denver, CO, USA.
- Unger, C., and Cimiano, P. (2011). Pythia: Compositional Meaning Construction for Ontology-Based Question Answering on the Semantic Web. In NLDB 2011, pp. 153-160.
- Unger, C., Freitas, A., and Cimiano, P. (2014). An Introduction to Question Answering over Linked Data. In Reasoning Web, pp. 100-140.
- Chiarcos, C., Sukhareva, M. (2015). OLiA–Ontologies of Linguistic Annotation. Semantic Web Journal: Multilingual Linked Open Data, vol. 6, number 4, pp. 379-386.