Brat2Viz: a Tool and Pipeline for Visualizing Narratives from Annotated Texts

ECIR 2021

Evelin Amorim, Alexandre Ribeiro, Inês Cantante, Alípio Jorge, Brenda Santana, Sérgio Nunes, Purificação Silvano, António Leal, Ricardo Campos

INESCTEC
U.PORTO FLUP FACULDADE DE LETRAS UNIVERSIDADE DO PORTO
Centro de Linguística da Universidade do Porto
U.PORTO FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO
ipt Instituto Politécnico de Tomar

Narratives

- **Narratives texts** are characterized by a chronological succession of **events**, **causality relations** between the events, and **actors** [Adams 92]
- Narratives are central in **many domains**:
  - Literature
  - Journalism
  - Health
  - Finance
- Automatic **Narrative Extraction from Text** is challenging and relevant
- **Huge lack** of densely **annotated data**
- Narrative **annotation effort** is **complex** and **hard to validate**
- We propose a **pipeline** for **narrative visualization** and labeling validation

Narratives

Narrative Texts

characterized by the chronological succession of events, causality relations between the events and actors

Narratives Main Domains

- Literature
- Journalism
- Health
- Finance

Automatic **Narrative Extraction from Text**

Challenging and relevant task

Huge Lack

Densely annotated data

Narrative Annotation Efforts

complex and **hard to validate**

Narrative Extraction from Text

- **Narrative Extraction from Text** - identify:
  - **Narrative elements** (actors, events, times)
  - Their **attributes**
  - **Semantic links** between them (temporal, referential, semantic roles).
- **Result**
  - **Structure** that can then be **represented** in a formal language
  - From this structure we can produce different representations such as diagrams or texts
- **Existing approaches** benefit from **annotated corpora**
  - Training
  - Validation

Narrative Extraction from Text

Identification of

- Narrative elements (actors, events, times)
- Their attributes
- Semantic links between them
  - temporal,
  - referential,
  - semantic roles

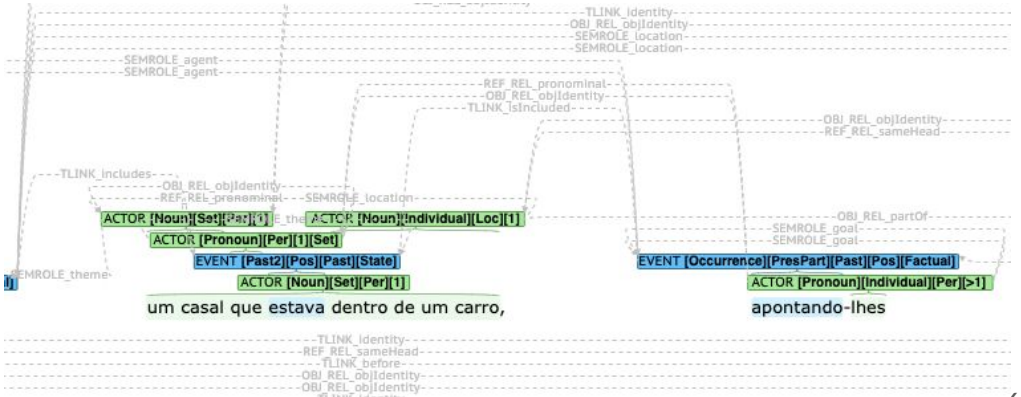
Result

Structure that can then be **represented** in a formal language

From this structure we can produce different representations such as diagrams or texts

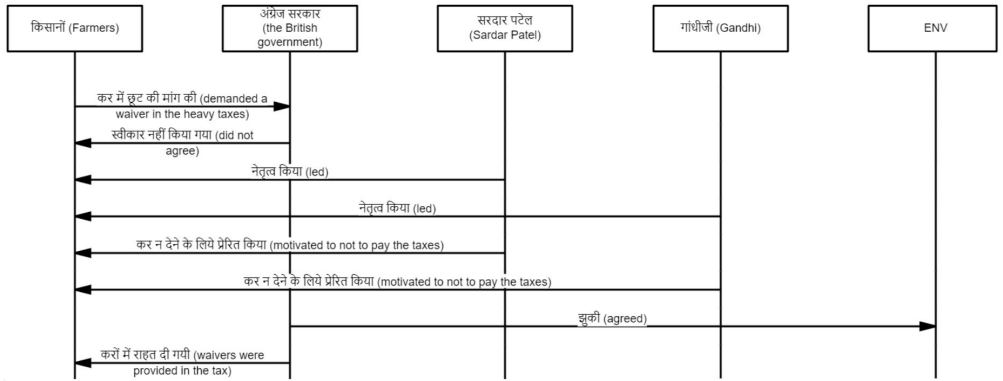
The annotation effort

- **The task of corpus annotation** is hard and **laborious**
  - Choose a **semantic annotation framework**
  - Adapt to the **target language** (e.g. Portuguese)
  - Avoid overloaded annotation
  - Check and i**nspect annotation**
- **Existing annotation tools**
  - Brat, Prodigy, etc.
  - Easy to use
  - **Do not enable easy verification**



Visualization of Narratives

- A **final product** of Narrative Extraction
- In the **annotation process**
  - Enable visualization of the result of the annotation
  - **Help** the annotator



(Bhattacharyya et al 2020)

Our Proposal: Brat2Viz

A tool and pipeline to produce a visualization of narratives from annotations, which enables the human annotator to validate the labels assigned

Brat2Viz and Text2Story

- A Narrative Annotation Visualization tool: **Brat2Viz**
  - Supports the **high level visualization** of narrative annotation done with Brat
  - Implements a **pipeline**
    - From text annotation to visualization
- **Text2Story** is a project (https://text2story.inesctec.pt/)
  - Currently annotating a corpus of news stories in **European Portuguese**

text2story

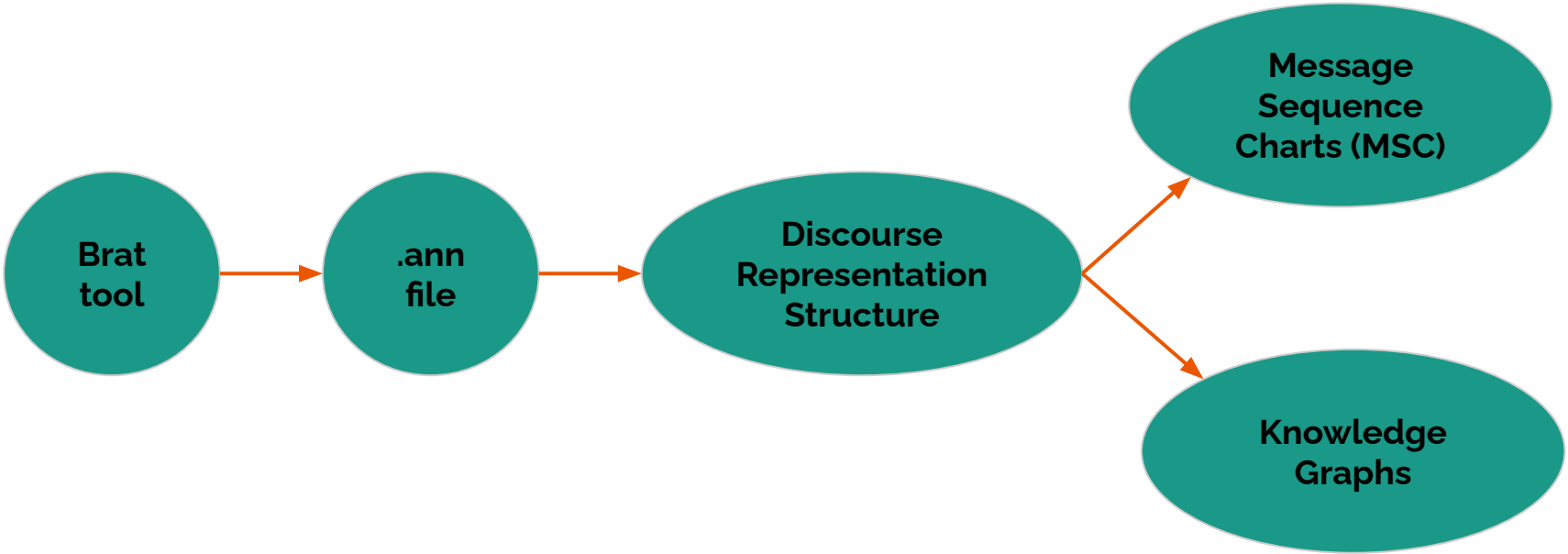
Main contributions

1. An **extensible framework** to generate visual representations of narratives from annotated texts
2. The use of a **formal logic-based language**, Discourse Representation Structure (DRS), as an intermediate language to the visual representation
3. An online interactive **demonstration** of our pipeline
4. Available **code**

Related Work

- Narrative extraction
  - ROCStories
  - Ontologies
- Narrative visualization
  - Communication Design and Journalism
  - MSC

The Narrative Annotation Visualization Pipeline



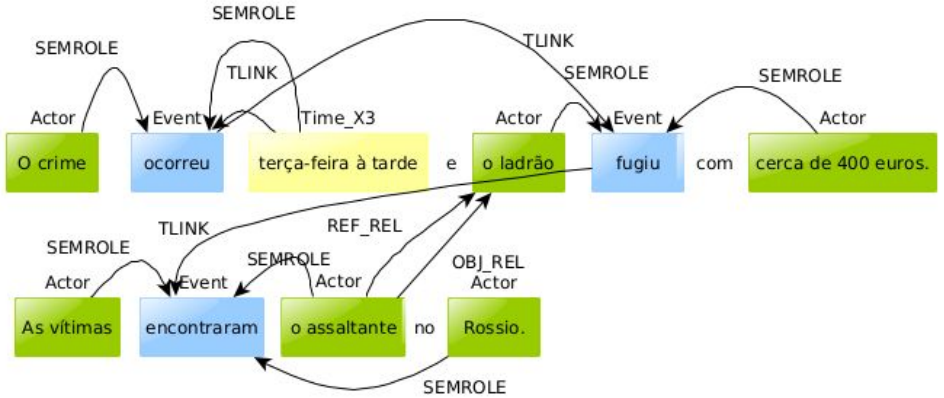
Annotation Scheme

- Annotation in **three levels**:
  - Referential
  - Temporal
  - Semantic role labeling
- **Annotation scheme** used:
  - **ISO 24617-1/9** for the first two levels
  - **L**inguistic **I**nf**R**astructure for **I**nteroperable Resour**C**es and **S**ystems (**LIRICS - ISO 24617-4**) for the third
- Adaptations of **number of tags** and **types of attributes** due to
  - Multilayer annotation
  - Properties of Portuguese
  - Genre of the corpus (news)

Annotation Scheme

- **Three types of tags**:
  - **Actors**, to annotate characters in the story
    - (e.g., “um homem” – “a man”)
  - **Events**, for events
    - (e.g., “assaltou” – “robbed”)
  - **Timex3**, for temporal expressions.
    - (e.g., “sexta-feira” - “Friday”)
- **Attributes** (with different values)

Annotation Scheme Linkage



- Between the **actors**
  - REF_REL
    - Synonymy, antonymy, hyponymy, etc.
  - OBJ_REL
    - Relations between linguistic units, from a discourse point of view
- Between **events** and **actors**
  - SEM_ROLES_LINKS
    - Semantic relations between events and actors
- Between **events**, **temporal references** or **locations**:
  - TLINKS
    - Temporal links (TLINKS) account for the **chronological ordering of the event**

Structure of Brat2Viz

- **Brat2DRS** module:
  - **From annotation files** generated by Brat
  - parses them, and creates **a DRS representation** (Kamp & Reyle 93) for each news story
- **DRS2Viz** module:
  - Then, the **DRS2Viz deploys a web application** that produces visualizations of the original newstext

Brat2DRS module

- Parse the "**.ann**" file
- Build a **dictionary**
  - with the linguistic elements found
- Assign a **symbolic variable** to each event
- **Transcribe** to
  - First Order Logic
  - DRS (**Discourse Representation Structure**)

```
T3 ACTOR 0 8 Um homem
A1 Lexical_Head T3 Noun
A2 Individuation T3 Individual
A3 Actor_Type T3 Per
A4 Involvement T3 1
T9 EVENT 9 17 assaltou
E1 EVENT:T9
A32 Class E1 Occurrence
A33 Tense E1 Past1
A34 Polarity E1 Pos
A35 Time E1 Past
```

```
event(a), eventClass(a,Occurrence), Tense(a,Past1),
Polarity(a,Pos), Time(a,Past), relationRole(agent,T3)
```

DRS - Discourse Representation Structure

- **DRS statements** are generated in a textual format declaring:
  - Events' properties
  - Actors
  - Time expressions
  - Relations between them
- High-level abstract representation of the narrative
  - Can be used by **subsequent operations**
  - **No need to go back** to the original text
  - Can be used for reasoning
- DRS declaring one event named 'a' and some attributes of the event

```
event(a), eventClass(a,Occurrence), Tense(a,Past1),
Polarity(a,Pos), Time(a,Past), relationRole(agent,T3)
```

DRS2Viz module

- **Module**:
  - **Parser** component
  - **Visualization** engine
- **Input**
  - DRS
- **Output**:
  - Visualizations in the web browser
- **Current visualizations**
  - MSC and Knowledge Graphs
- **Actors** are represented as **nodes**
- **Events** and **relations** are represented as **links** between nodes

DRS2Viz module

- The **parser** uses the DRS and extracts
  - actors
  - events
  - relations
- Into independent data structures
- **Structures** keep track of **identifiers**, and the lexical items that represent them

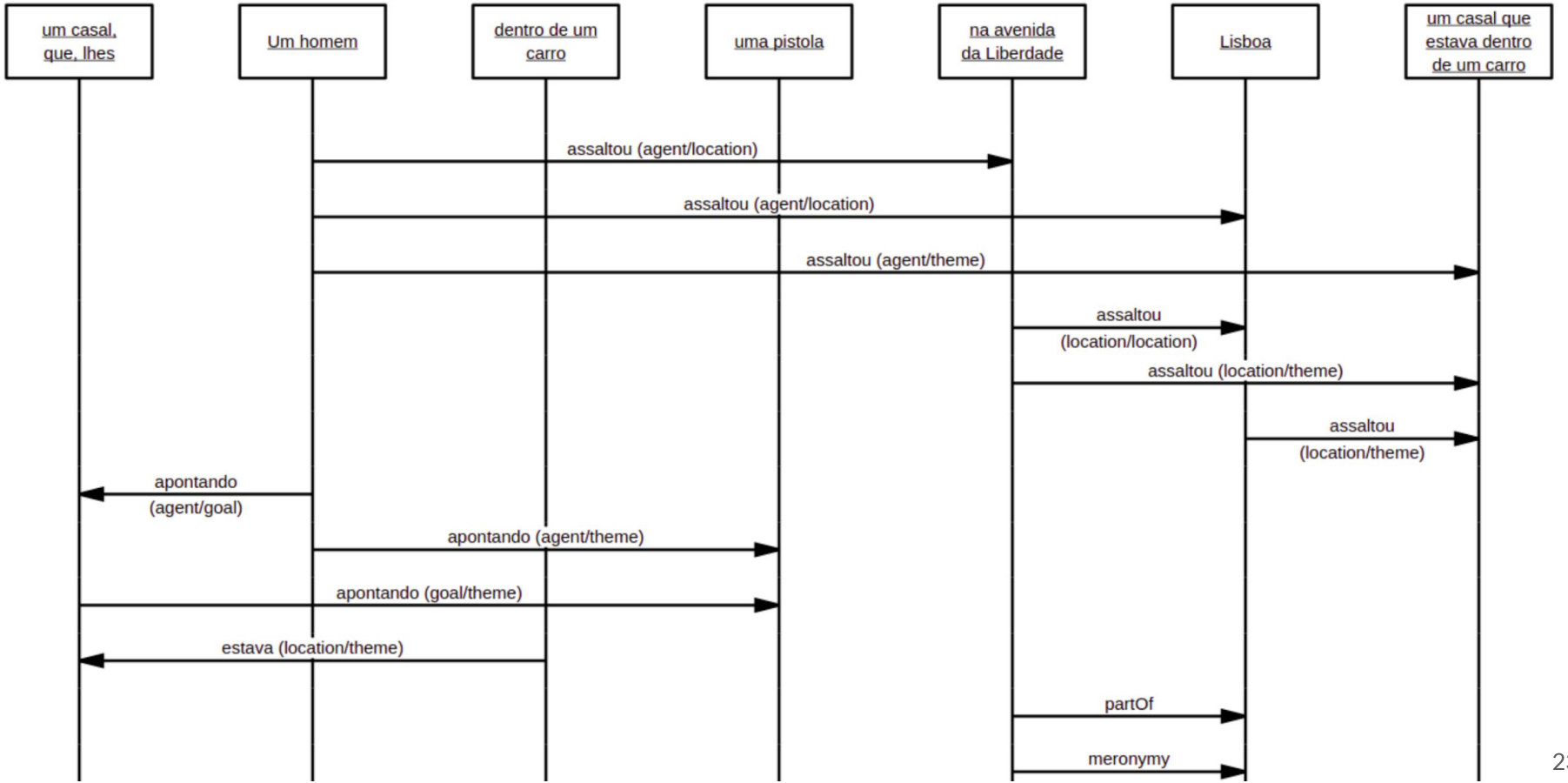
```
event(a), eventClass(a,Occurrence), Tense(a,Past1),
Polarity(a,Pos), Time(a,Past), relationRole(agent,T3)
```

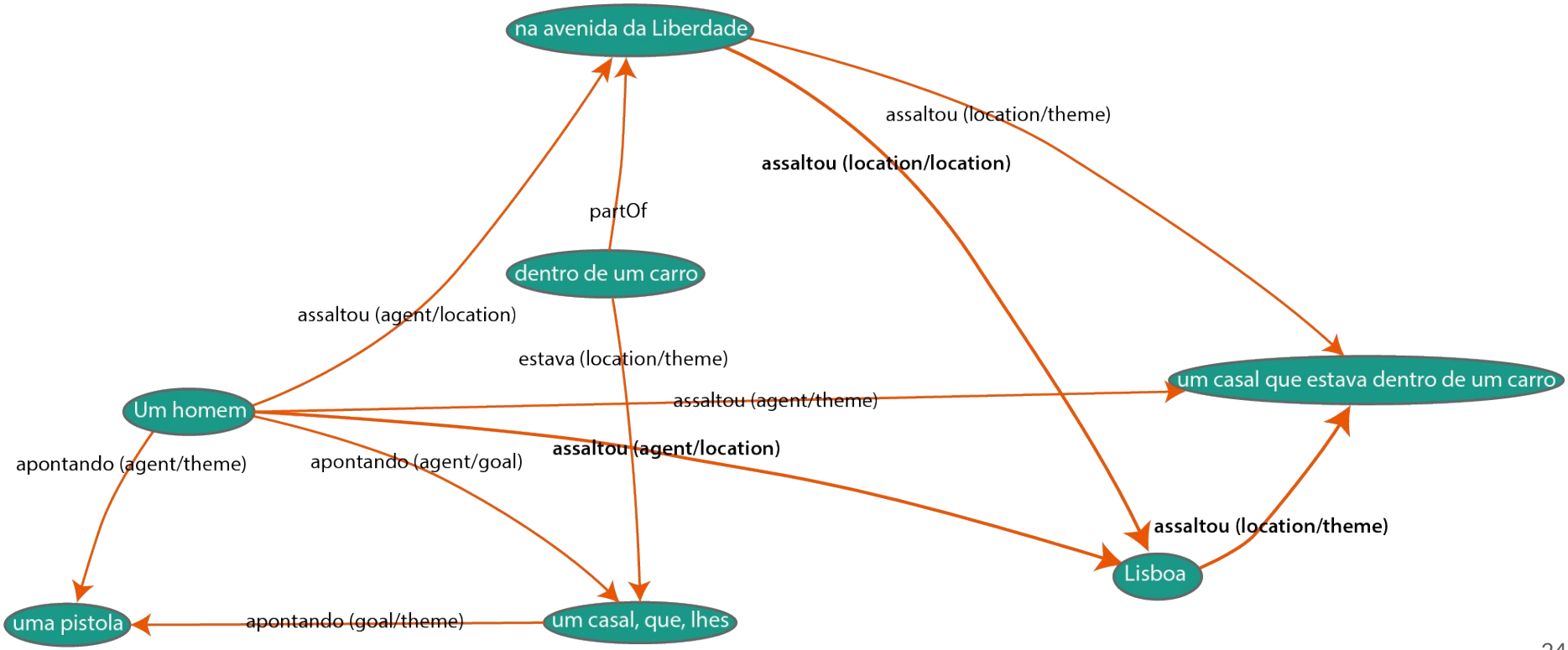
DRS2Viz module

- The **actor** is represented as T3: “Um homem”
- The **event** is: “assaltou”
- **Links** occur between actors
- Identify **pairs of actors** linked by an event
- **Group** references by actor
  - ‘Thief’, ‘man’, ‘assailant’

```
event(a), eventClass(a,Occurrence), Tense(a,Past1),
Polarity(a,Pos), Time(a,Past), relationRole(agent,T3)
```

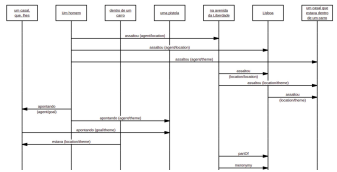
Visualization

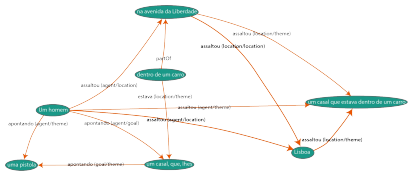

um casal, que, lhes
Um homem
dentro de um carro
uma pistola
na avenida da Liberdade
Lisboa
um casal que estava dentro de um carro
assaltou (agent/location)
assaltou (agent/location)
assaltou (agent/theme)
assaltou (location/location)
assaltou (location/theme)
assaltou (location/theme)
apontando (agent/goal)
apontando (agent/theme)
apontando (goal/theme)
estava (location/theme)
partOf
meronymy


na avenida da Liberdade
assaltou (location/theme)
assaltou (location/location)
partOf
dentro de um carro
assaltou (agent/location)
estava (location/theme)
um casal que estava dentro de um carro
Um homem
assaltou (agent/theme)
assaltou (agent/location)
apontando (agent/theme)
apontando (agent/goal)
assaltou (location/theme)
Lisboa
uma pistola
apontando (goal/theme)
um casal, que, lhes

Conclusion / Summary

- **Brat2Viz** tool and pipeline
  - Visualize narratives from annotations produced in Brat
- Two-step **modular pipeline**
  - Narrative **annotations** to the **DRS formal language**
  - **DRS** to **visual** representations
- Currently, we **visualize** the narratives as
  - **Message Sequence Charts**
  - **Knowledge Graphs**





If you want to try or extend Brat2Viz

- Online **demo**
  - https://nabu.dcc.fc.up.pt/brat2viz
- Check out our **project** web page and get in touch
  - https://text2story.inesctec.pt/
- Github **repo**
  - https://github.com/LIAAD/brat2viz

text2story

Future Work

- **Modular** pipeline enables
  - **Extension** and adaptation to other scenarios
  - **Other visualization**s from DRS, such as **timelines**, **story trees** and **icon-strips**
  - Other types of representations (e.g., simplified **textual narratives**)
  - **The annotation scheme** can also be adapted to other needs
  - **Narrative extraction algorithms** may also be plugged as automatic annotators
    - **Narrative Extraction Visualization Pipeline**

Thank You!

text2story