

# DESAFIOS PARA A ANOTAÇÃO SEMÂNTICA DE TEXTOS JURÍDICOS: LIMITES NO USO DA FRAMENET E ROTAS ALTERNATIVAS

Anderson Bertoldi  
Rove Chishman

Programa de Pós-Graduação em Linguística Aplicada  
Universidade do Vale do Rio dos Sinos



## Objetivo:

- Este trabalho tem como objetivo analisar em que medida as etiquetas semânticas da FrameNet podem ocasionar divergências no que tange a sua aplicação para anotação de *corpus* em línguas diversas do inglês.
- Parte-se das unidades lexicais evocadoras do *frame* `Criminal_process` na FrameNet e se analisa:
  - (i) equivalência das unidades lexicais,
  - (ii) correspondência dos *frames* (eventos jurídicos) e
  - (iii) correspondência dos elementos de *frame* (participantes dos eventos jurídicos).

# FrameNet e suas Aplicações:

- FrameNets:
  - Spanish FrameNet (Subirats, 2009)
  - Japanese FrameNet (Ohara, 2009)
  - FrameNet Brasil (Salomão, 2009)
- Anotação de *corpora*:
  - SALSA (Burchardt et al., 2009),
  - FrameCorp (Chishman et al., 2008)
- Criação de léxicos de domínio baseados em *frames*:
  - Kicktionary (Schmidt, 2009)
- Criação Automática de Recursos lexicais:
  - Padó e Lapata (2005)

# A FrameNet:

- Recurso lexical organizado segundo os princípios da Semântica de *Frames*. Segundo Fillmore et al., 2003, as unidades primárias de descrição na FrameNet são o **frame** e a **unidade lexical**.

- **Frame**: um sistema de conceitos relacionados de tal forma que, para se entender um conceito, é necessário a compreensão de todos os conceitos relacionados (Fillmore, 1982).

Arrest: AUTHORITIES, SUSPECT, OFFENSE, CHARGES.

- **Unidade Lexical**: A unidade lexical é compreendida como a soma de uma forma e um significado .

*apprehend.v, apprehension.n, arrest.n, arrest.v*



## Frame Index

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#)  
[Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

- [Abandonment](#)
- [Abounding with](#)
- [Absorb heat](#)
- [Abundance](#)
- [Abusing](#)
- [Access scenario](#)
- [Accompaniment](#)
- [Accomplishment](#)
- [Accoutrements](#)
- [Accuracy](#)
- [Achieving first](#)
- [Active substance](#)
- [Activity](#)
- [Activity abandoned state](#)
- [Activity done state](#)
- [Activity finish](#)
- [Activity ongoing](#)
- [Activity pause](#)
- [Activity paused state](#)
- [Activity prepare](#)
- [Activity ready state](#)
- [Activity resume](#)
- [Activity start](#)
- [Activity stop](#)

## Arrest

[Lexical Unit Index](#)

### Definition:

**Authorities** charge a **Suspect**, who is under suspicion of having committed a crime (the **Charges**), and take him/her into custody.  
 The police **ARRESTED** Harry on charges of manslaughter.

### FEs:

#### Core:

**Authorities [Auth]**      The **Authorities** charge the **Suspect** with committing a crime, and take him/her into custody.  
**Semantic Type: Sentient**      **The police** **ARRESTED** Harry on charges of manslaughter.

**Charges [Chrg]**      **Charges** identifies a category within the legal system; it is the crime with which the **Suspect** is charged.  
 The police **ARRESTED** Harry on charges of manslaughter.

**Offense [Off]**      **Offense** identifies the ordinary language use of the reason for which a **Suspect** is arrested.  
 They arrested Harry for shoplifting.

**Suspect [Susp]**      The **Suspect** is taken into custody, under suspicion of having committed a crime.  
 The police **ARRESTED** Harry on charges of manslaughter.



## Frame Index

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#)  
[Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

- [Abandonment](#)
- [Abounding with](#)
- [Absorb heat](#)
- [Abundance](#)
- [Abusing](#)
- [Access scenario](#)
- [Accompaniment](#)
- [Accomplishment](#)
- [Accoutrements](#)
- [Accuracy](#)
- [Achieving first](#)
- [Active substance](#)
- [Activity](#)
- [Activity abandoned state](#)
- [Activity done state](#)
- [Activity finish](#)
- [Activity ongoing](#)
- [Activity pause](#)
- [Activity paused state](#)
- [Activity prepare](#)
- [Activity ready state](#)
- [Activity resume](#)
- [Activity start](#)
- [Activity stop](#)

### Precedes: Arrangement

Is Preceded by:  
 Is Inchoative of:  
 Is Causative of:  
 See also:

### Lexical Units:

*apprehend.v, apprehension.n, arrest.n, arrest.v, book.v, bust.n, bust.v, collar.v, cop.v, nab.v, summons.v*

Created by 664 on 08/16/2001 04:47:24 PDT Thu

<u>Lexical Unit</u>	<u>LU Status</u>	<u>Lexical Entry Report</u>	<u>Annotation Report</u>	<u>Annotator ID</u>	<u>Created Date</u>
apprehend.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 03:51:06 PDT Mon
apprehension.n	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 04:28:20 PDT Mon
arrest.n	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 04:14:25 PDT Mon
arrest.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	01/23/2002 02:57:49 PST Wed
book.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	296	01/10/2003 01:21:57 PST Fri
bust.n	Needs_SCs	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 04:27:02 PDT Mon
bust.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 03:51:54 PDT Mon
collar.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 03:54:33 PDT Mon
cop.v	Insufficient_Attestations	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 03:55:49 PDT Mon
nab.v	Finished_Initial	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	664	04/29/2002 03:52:46 PDT Mon
summons.v	Needs_SCs	<a href="#">Lexical entry</a>	<a href="#">Annotation</a>	303	04/18/2005 04:44:14 PDT Mon

# Exemplos de sentenças anotadas pela FrameNet:

- Elizabeth Woodville , informed of the ARREST<sup>Target</sup> [<sub>Suspect</sub> of her brother and of her son by a former marriage , as well as of the young king 's faithful chamberlain] , was beside herself with anger and anxiety .
- The ARREST<sup>Target</sup> [<sub>Suspect</sub> of the ` Gang of Four " ] [<sub>Time</sub> in October 1976] eliminated them from the power struggle .
- Then came the [<sub>Manner</sub> dramatic] ARREST<sup>Target</sup> [<sub>Suspect</sub> of Roderick] [<sub>Time</sub> last August] when the Royal Navy frigate Argonaut intercepted his yacht 150 miles south west of Gibraltar .

# Equivalência de Unidades Lexicais:

- Ao usar a FrameNet para anotação de *corpora* em outras línguas que não seja a língua inglesa, é necessário levar-se em conta a equivalência das unidades lexicais.
- No caso de um *corpus* jurídico, é necessário levar em conta: (i) a equivalência das unidades lexicais e (ii) a equivalência dos conceitos jurídicos.
- A equivalência das unidades lexicais em inglês e português é um ponto crítico para a anotação de *corpus* jurídico, visto que os eventos jurídicos no Brasil e nos Estados Unidos não são sempre correspondentes.



# O caso do verbo *acusar*:

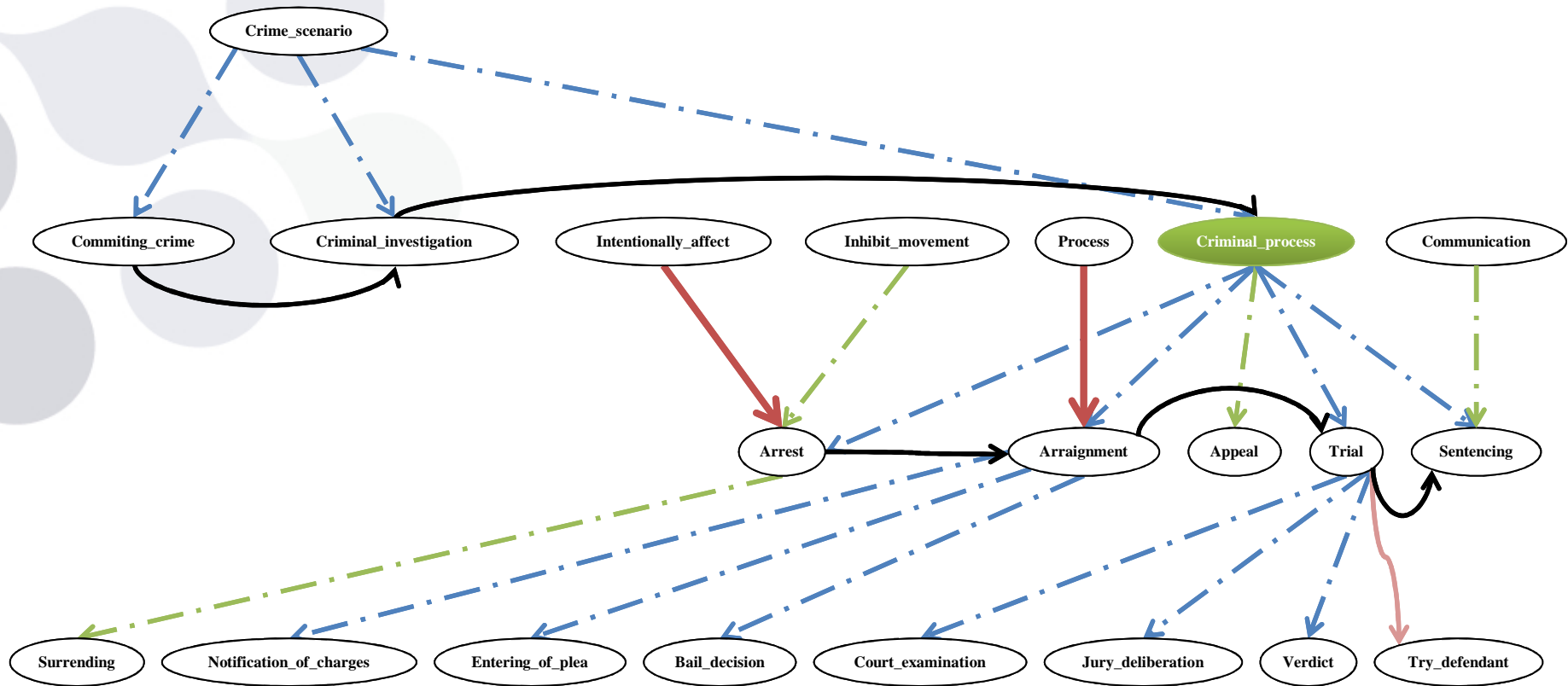
- Ao anotar um *corpus* jurídico, um anotador vai facilmente identificar o verbo ***to accuse*** como equivalente de tradução do verbo ***acusar***.
- O próximo passo é identificar na FrameNet o ***frame*** evocado pelo verbo ***to accuse***.
- Ao consultar a FrameNet, o anotador vai ver que o verbo ***to accuse*** evoca três ***frames*** diferentes na FrameNet :
  - Judgment\_communication
  - Judgment
  - Notification\_of\_charges

# O caso do verbo *denunciar*:

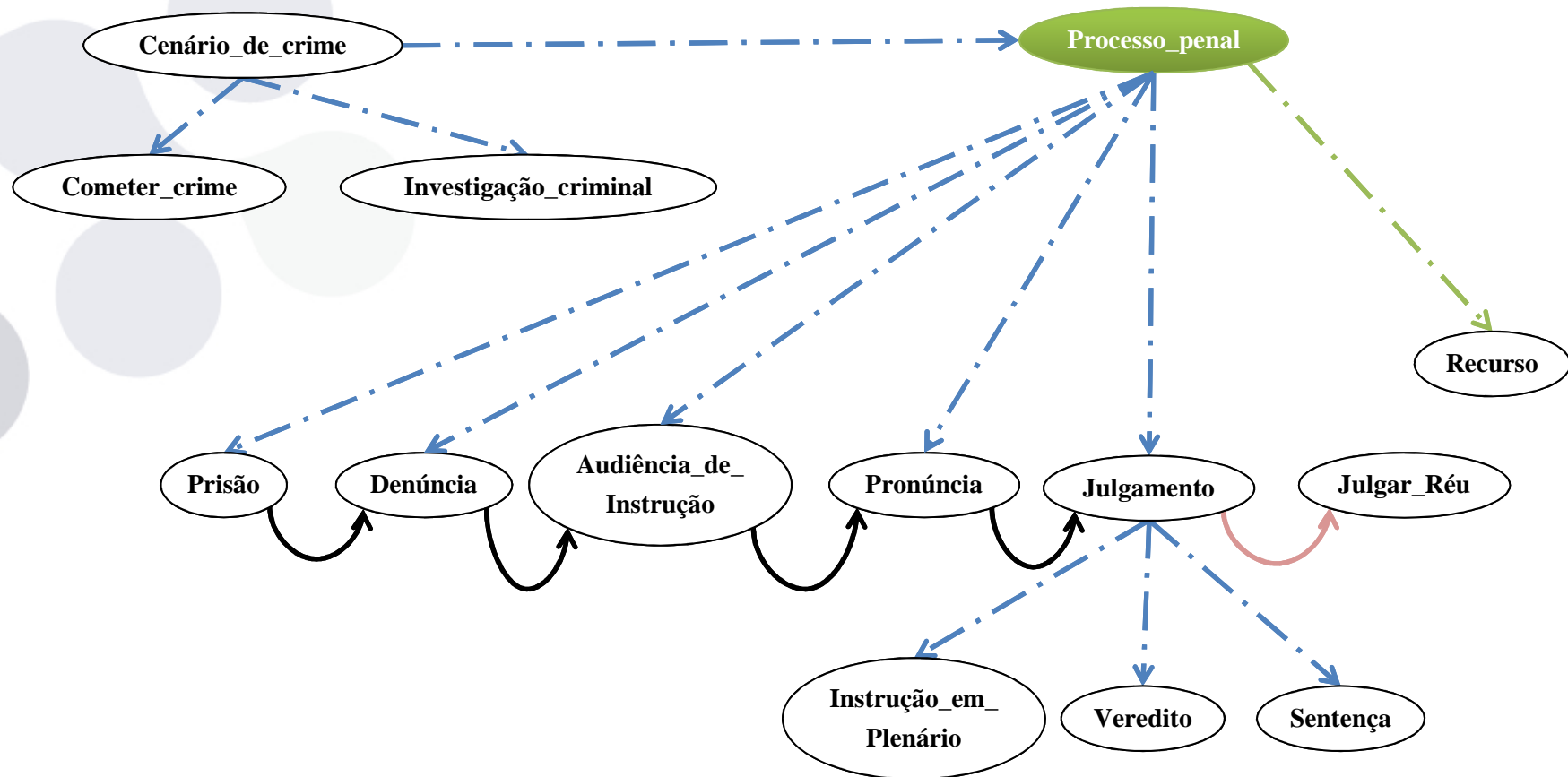
- No caso do verbo *denunciar*, o equivalente de tradução mais provável parece ser o verbo *to denounce*.
- No entanto, o anotador verá que na FrameNet o verbo *to denounce* evoca apenas o *frame* `Judgment_communication`.
- Assim, o anotador terá de encontrar um outro equivalente para o verbo *denunciar* e fazer uma nova busca na base de dados da FrameNet.
- Se o anotador procurar um equivalente para *denunciar* em um dicionário jurídico bilíngue (Goyos Jr., 1992), encontrará como equivalentes *to denounce; accuse; inform against; report; proclaim*.
- Isso só dificultará o trabalho do anotador, uma vez que o *frame* evocado pelo verbo *to accuse* não corresponde ao *frame* evocado pelo verbo *denunciar*.

# Correspondência de *frames*:

- Outro desafio para se utilizar os frames semânticos para anotação de *corpora* jurídicos é a correspondência de eventos jurídicos entre dois sistemas distintos, o norte-americano e o brasileiro.
- Os verbos ***acusar*** e ***denunciar*** possuem o mesmo equivalente em inglês: ***to accuse***. Enquanto o verbo ***to accuse*** evoca o *frame* `Notification_of_charges` na FrameNet, os verbos ***acusar*** e ***denunciar*** podem, ambos evocarem os *frames* `Investigação_criminal` e `Denúncia`, que não são correspondentes ao *frame* `Notification_of_charges`.
- Os *frames* `Investigação_criminal` e `Denúncia` não são correspondentes ao *frame* `Notification_of_charges`.
- A correspondência entre *frames* “jurídicos” envolve a correspondência entre sistemas jurídicos.



Legend		Parent → Child Relation Types:				
Parent Frame	→	Child Frame				
			→	Inheritance	→	Causative Of
			-.-.-→	Subframe	→	Inchoative Of
			→	Perspective On	→	See Also
			-.-.-→	Using		
			→		→	Precedes



Legend		Parent → Child Relation Types:				
Parent Frame	→	Child Frame				
				Inheritance		Causative Of
				Subframe		Inchoative Of
				Perspective On		See Also
				Using		
					<b>Parent → Child Relation Types:</b>	
						Precedes

# Correspondência de elementos de *frame*:

- Um vez que os eventos jurídicos variam entre os sistemas norte-americano e brasileiro, os *frames* e, conseqüentemente, os elementos de *frame* também variam.
- Exemplo: Se um anotador tiver que anotar uma sentença com o verbo **acusar**, ao procurar pelo equivalente **to accuse** na FrameNet, ele vai chegar ao *frame* `Notification_of_charges`.
- O *frame* `Notification_of_charges`, na FrameNet, descreve uma situação inicial de uma sessão de Arraignment e possui como elementos de *frame*: `ARRAIGN_AUTHORITY`, `ACCUSED` e `CHARGES`.
- O que deve o anotador fazer se o elemento de *frame* `ARRAIGN_AUTHORITY` não se ajusta à descrição do sistema processual brasileiro?

## FrameNet

Try\_defendant

### Legal event:

A **DEFENDANT** is tried by a JURY or **JUDGE** in a COURT for **CHARGES**. This frame perspectivalizes the general Trial frame.

### Frame elements:

**CHARGES**  
**DEFENDANT**  
**GOVERNING\_AUTHORITY**  
**JUDGE**  
**JURY**

### Lexical Units:

try.v



## Frame Jurídico Brasileiro

Julgar\_acusado

### Evento Jurídico:

O *frame* Julgar\_acusado descreve o julgamento de um **RÉU** que é acusa de um crime. Um **JÚRI** é responsável por avaliar as **ACUSAÇÕES** e decidir se o **RÉU** é inocente ou culpado.

### Elementos de *Frame*:

**ACUSAÇÕES**  
**RÉU**  
**AUTORIDADE\_GOVERNAMENTAL**  
**JUIZ**  
**JÚRI**

### Unidades Lexicais:

Julgar.v

# Comparação entre anotações:

- Three members of the 11-member court voted to give [the U.S. court JUDGE] the right to **try** [Pang DEFENDANT] [for first-degree murder . CHARGES]
- [The court JUDGE] ] would **try** [people DEFENDANT] [for war crimes , crimes against humanity and other serious human-rights violations . CHARGES]
- Anulado parcialmente o julgamento do Júri, [o acusado RÉU] voltará a ser **jugado** [pelo Tribunal Popular JÚRI] [somente em relação ao homicídio. ACUSAÇÕES]
- Para o governador, o fato de [os acusados RÉU] serem **jugados** [por um júri popular JÚRI] é muito positivo.



## FrameNet

Notification\_of\_charges

### Legal event:

The judge or other court officer (the **ARRAIGN\_AUTHORITY**) informs the **ACCUSED** of the **CHARGES** against him/her, i.e. the alleged actions and the relevant laws.

### Frame elements:

**ACCUSED**  
**ARRAIGN\_AUTHORITY**  
**CHARGES**

### Lexical Units:

*accuse.v.* →  
*charge.n.* →  
*charge.v.* →  
*indict.v.* →  
*indictment.n.* →

## Frame Jurídico Brasileiro

Denúncia

### Evento jurídico:

O *frame* Denúncia representa um evento jurídico em que a promotoria, **AUTORIDADE\_DE\_ACUSAÇÃO**, apresenta denúncia contra o **ACUSADO**.

### Elementos de *Frame*:

**ACUSADO**  
**AUTORIDADE\_DE\_ACUSAÇÃO**  
**ACUSAÇÕES**

### Lexical Units:

Acusar.v  
Acusação.n  
Acusar.v  
Pronunciar.v  
Pronúncia.n

# Comparação entre anotações:

- [Police<sub>ARRAIGN\_AUTHORITY</sub>] are expected to **charge** [a father<sub>ACCUSED</sub>] [with attempted murder after his fifteen month old son was found stabbed in his cot .  
CHARGES]
- [The police<sub>ARRAIGN\_AUTHORITY</sub>] **charged** [a young nurse<sub>ACCUSED</sub>] [with four murders .  
CHARGES]
- [A partir desses documentos, <sub>EVIDÊNCIAS</sub>] [o Ministério Público <sub>AUTORIDADE\_DE\_ACUSAÇÃO</sub>] **denunciou** [os bicheiros<sub>ACUSADO</sub>] novamente e ficou comprovado que eles mantinham suas atividades mesmo de trás das grades
- [O empresário<sub>ACUSADO</sub>] foi **denunciado** [em março<sub>TEMPO</sub>] [pelo procurador da República Artur Gueiros, <sub>AUTORIDADE\_DE\_ACUSAÇÃO</sub>] [com base na Lei 1.533/51, que cataloga crimes contra a economia popular. <sub>BASE\_LEGAL</sub>]

## FrameNet

Arraignment

### Legal event:

At a court hearing, a **DEFENDANT** is informed of the **CHARGES** against him or her, (usually) enters a plea, and a decision is made by a **JUDGE** as to the amount of bail, if any.

### Frame elements:

**CHARGES**

**DEFENDANT**

**JUDGE**

### Lexical Units:

*arraign.v.*

*arraignment.n*

## Frame Jurídico Brasileiro

Sem *frame* correspondente

### Evento Jurídico:

Sem evento jurídico correspondente no processo penal brasileiro.

### Frame Elementos de *Frame*:

Sem elementos de *frame* correspondentes

### Lexical Units:

Sem unidades lexicais equivalente

Sem unidades lexicais equivalente

# Conclusões:

- A falta de correspondência entre os sistemas jurídicos e os procedimentos jurídicos de cada país se reflete nos *frames* semânticos.
- Algumas unidades lexicais podem apresentar um equivalente na língua alvo da anotação, porém o cenário jurídico evocado pela unidade lexical equivalente não corresponde ao cenário evocado pela unidade lexical na FrameNet.
- A solução para o problema da anotação de *corpora* jurídicos em diferentes línguas pode ser a criação de uma base de dados lexicais específica para a linguagem jurídica de cada país.