

Annotating Arguments in a Corpus of Opinion Articles

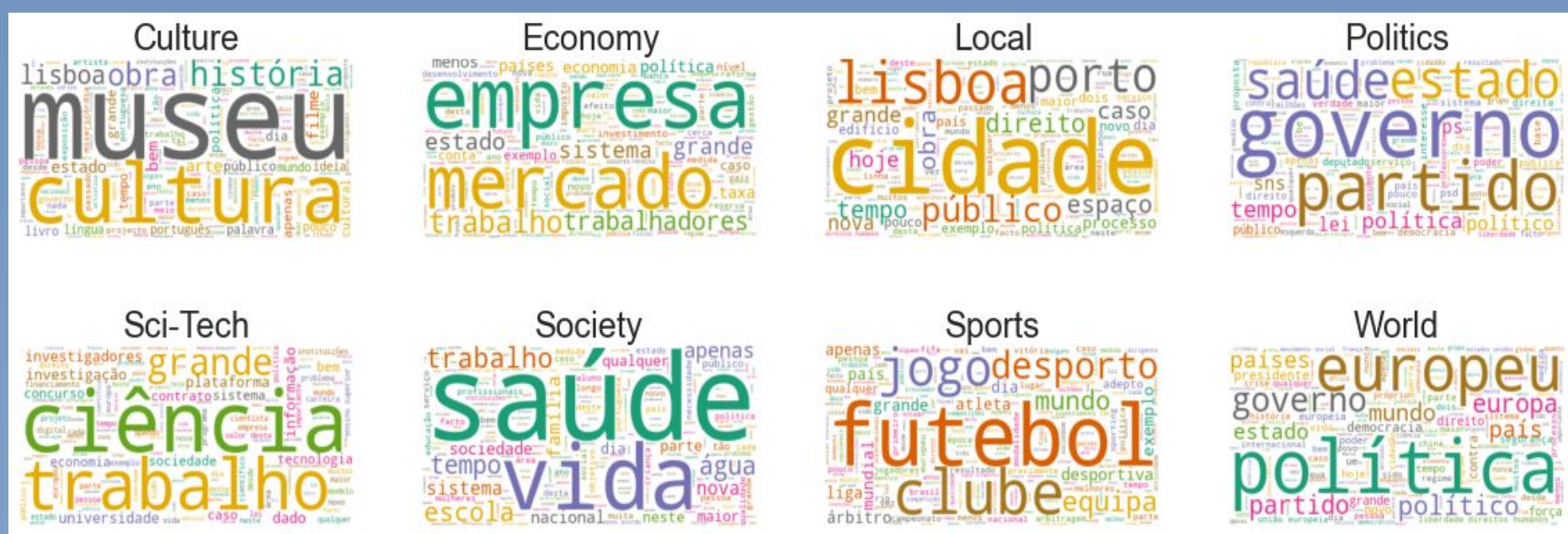
LREC 2022
Marseille

Gil Rocha, Luís Trigo, Henrique Lopes Cardoso, Rui Sousa-Silva
Paula Carvalho, Bruno Martins, Miguel Won

1. Motivation

- Increasing number of **argument annotated corpora**.
- Most focus on **English** texts with explicit **argumentative discourse markers**
 - persuasive essays, legal documents, etc
- Our works is the first **Portuguese argument annotation** project focused on **opinion articles**
- Opinion articles as argumentative text genre
 - be free and fluid
 - weak argumentative structure
 - almost no argumentative discourse markers

2. Corpus



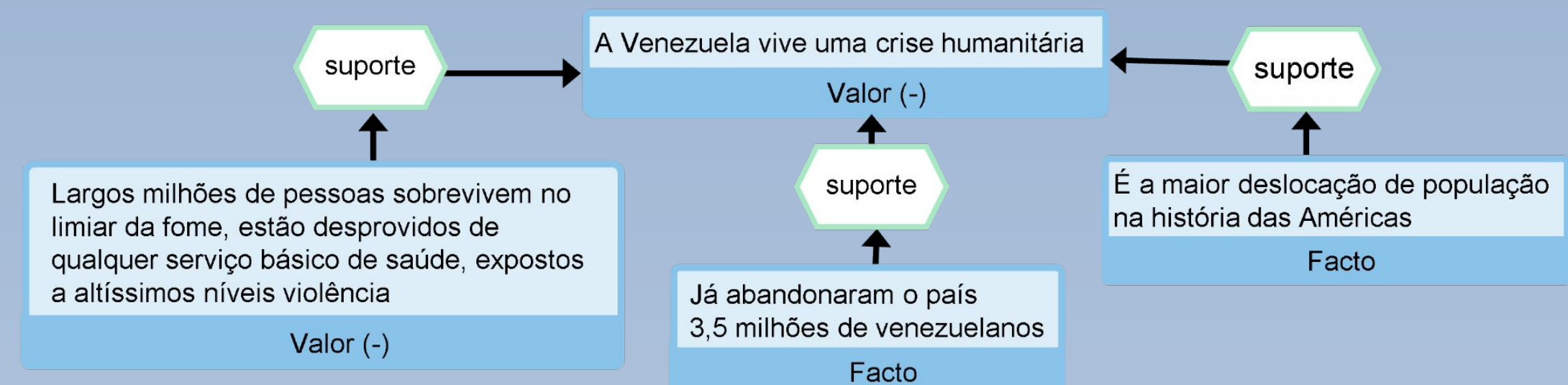
- 373 opinion articles
 - Median of 10 paragraphs and 189 tokens per article

from 8 different topics

- Median number of paragraphs between 9 and 11
- Median number of tokens between 161 and 203.

3. Annotation analysis

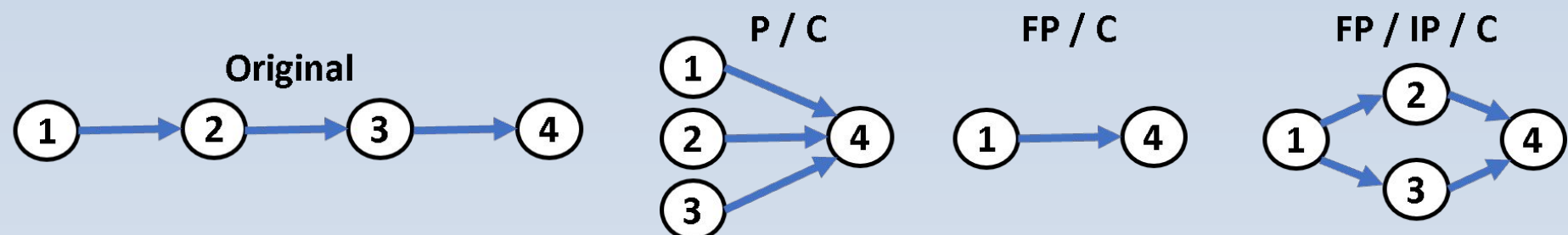
1. Primeiro, as primeiras coisas. A Venezuela vive uma crise humanitária. Não há como a esconder. Largos milhões de pessoas sobrevivem no limiar da fome, estão desprovidos de qualquer serviço básico de saúde, expostos a altíssimos níveis violência. Já abandonaram o país 3,5 milhões de venezuelanos, dos quais um terço se refugiou na Colômbia, havendo regiões onde a situação se antolha dramática. É a maior deslocação de população na história das Américas.



4. Inter-annotator agreement

ADU argumentative roles

Aim: Graph simplification analysis focused on ADU roles (premise or conclusion)



| Labels | P / C | FP / C | FP / IP / C |
|------------------|-------------|-------------|-------------|
| Cohen's κ | .62 (± .34) | .71 (± .35) | .56 (± .33) |
| Obs. Agr. | 82% (± .18) | 85% (± .18) | 74% (± .20) |
| M | 1061 | 1057 | 1061 |

- high IAA scores → annotators have a **shared interpretation** of **ADU roles**
- **higher IAA** when **removing** intermediate premises (IP): confusion with first premises and conclusions

Other Argumentation Corpora

- Persuasive essays [Stab and Gurevych, 2014, 2017]
 - *Major Claim* | *Claims* | *Premises*
- Crowdsourced argumentative texts [Peldszus and Stede, 2016; Skeppstedt et al., 2018]
- TV political debates [Visser et al., 2018]
 - Periodic Table of Arguments - Node types: *Fact*, *Value* or *Policy*

Annotation methodology

Annotation process

(i) ADU detection, (ii) ADU classification

(iii) relation identification, and (iv) relation classification

Annotation tool

- ArgMine (inspired on OVA+)
 - annotate text segments as nodes (ADUs) in a graph with connecting links
 - assign types to nodes (propositions of value, fact, policy)

Annotators

- Recruited four native language annotators with a degree in Language Sciences
- Annotation guidelines (refined with the involvement of the annotators)

ADUs

| | A | B | C | D | mean |
|-----------------|------|------|------|------|--------|
| Fact | 647 | 920 | 386 | 1710 | 915.8 |
| Policy | 56 | 167 | 265 | 179 | 166.8 |
| Value (neutral) | 2059 | 2790 | 2006 | 1247 | 2025.5 |
| Value (+) | 183 | 466 | 481 | 281 | 352.8 |
| Value (-) | 390 | 883 | 973 | 654 | 725.0 |
| Total | 3335 | 5226 | 4111 | 4071 | 4185.8 |
| ADUs per art. | 12.1 | 19.9 | 14.1 | 14 | 15 |

Fact - a piece of information whose veracity can often be verified

Policy - invokes the need to follow a specific directive or course of action

Value - judgments or opinions, which may present stances of an ethical, aesthetic, or political nature

ADU detection

Aim: analyze the extent to which annotators agree on the ADUs included in the text

| Threshold | .50 | .75 | .99 |
|------------------|-----|-----|-----|
| Cohen's κ | .29 | .21 | .15 |
| Obs. Agreement | 76% | 73% | 71% |

- IAA metrics show that this is a **very challenging task**

Relation identification

Aim: IAA for the connection between ADUs pairs

| | Original | P / C | FP / C | FP / IP / C |
|------------------|------------|------------|------------|-------------|
| Cohen's κ | .25 (±.51) | .27 (±.52) | .50 (±.50) | .26 (±.52) |
| Obs. Agr. | 69% (±25) | 66% (±27) | 75% (±27) | 69% (±25) |
| M | 1007 | 1005 | 938 | 1007 |

- "**Original**" setup: "**fair agreement**"
- Graph simplification: removing IPs has considerable impact ("**moderate agreement**")

Article assignment

- Each article is annotated by three annotators

| Annotator | A,B,C | A,B,D | A,C,D | B,C,D |
|--------------------|-------|-------|-------|-------|
| Annotated articles | 83 | 82 | 111 | 97 |

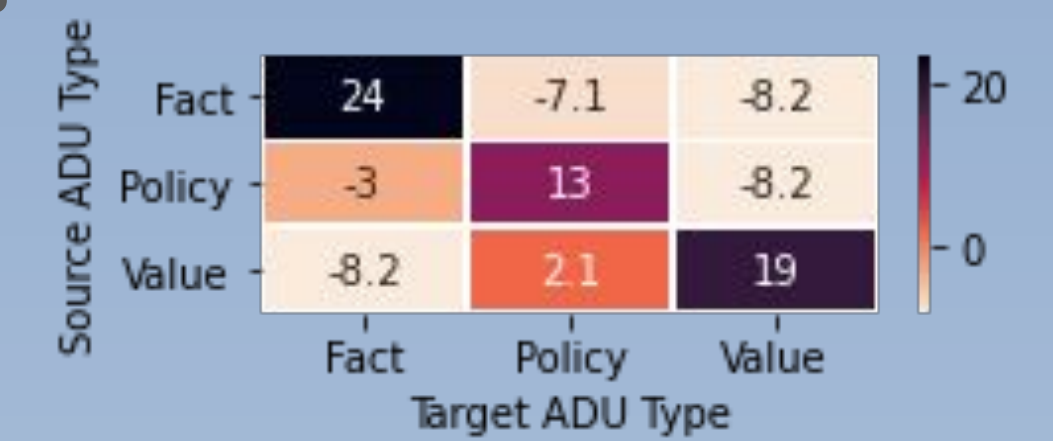
- Article/topic distribution per annotator

| | Culture | Economy | Local | Politics | Sci-Tech | Society | Sports | World | total |
|---|---------|---------|-------|----------|----------|---------|--------|-------|-------|
| A | 31 | 36 | 32 | 34 | 36 | 35 | 40 | 32 | 276 |
| B | 32 | 33 | 29 | 33 | 30 | 39 | 29 | 37 | 262 |
| C | 34 | 34 | 45 | 36 | 36 | 32 | 42 | 32 | 291 |
| D | 38 | 35 | 44 | 32 | 33 | 29 | 45 | 34 | 290 |

Relations

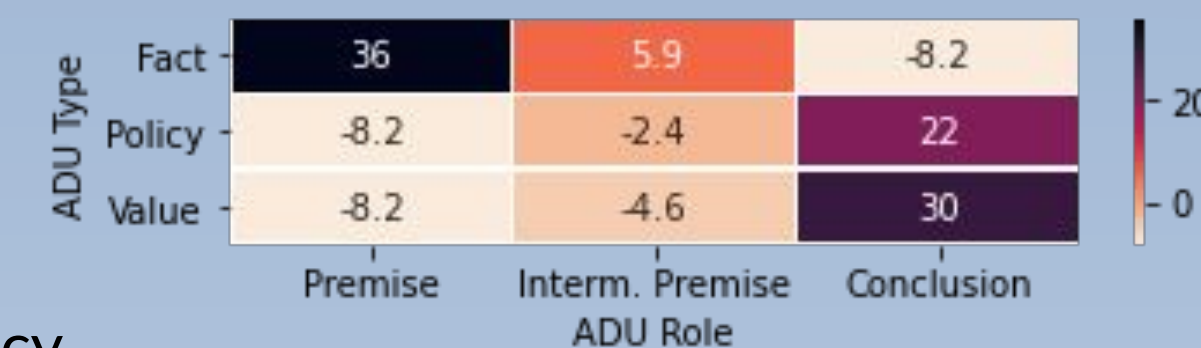
Source and Target types

- Phi-k correlation of 50%
- more relevant in pairs of ADUs with the same type of proposition.



Type and Role

- Phi-k correlation of 55%
- propositions of fact and premise role
- propositions of value/policy and conclusion role



ADU classification

Aim: IAA for ADU classification based on ADUs identified by both annotators

| | Global | Fact | Policy | Value |
|------------------|-------------|-------------|-------------|-------------|
| Cohen's κ | .40 (± .31) | .05 (± .38) | .31 (± .46) | .26 (± .39) |
| Obs. Agr. | 73% (± 17) | 68% (± 18) | 87% (± 13) | 73% (± 17) |
| M | 1061 | 782 | 300 | 1058 |

- "**Moderate agreement**" for the "Global" metric
- **Confusion matrix:**

| | Fact | Policy | Value |
|--------|------|--------|-------|
| Fact | 721 | 6 | 1871 |
| Policy | | 191 | 276 |
| Value | | | 4705 |

Graph-based analysis

Aim: Graph-based agreement measure capturing the semantic similarity between different argument graphs [Kirschner et al., 2015]

| | Original | P / C | FP / C | FP / IP / C |
|----|-------------|-------------|-------------|-------------|
| f1 | .80 (± .21) | .76 (± .23) | .84 (± .22) | .79 (± .21) |
| M | 982 | 963 | 890 | 983 |

- Annotators **tend to agree** with the **overall structure** of the annotation graph, even though some disagreement is observed between IPs and conclusion

Acknowledgments: This research is supported by project DARGMINTS (POCI/01/0145/FEDER/031460), LIACC (FCT/UID/CEC/0027/2020), INESC-ID (UIDB/50021/2020) and CLUP (UIDB/00022/2020), funded by Fundação para a Ciência e a Tecnologia (FCT). Gil Rocha is supported by a PhD grant (SFRH/BD/140125/2018) from FCT.