SCIENCE
PASSION
TECHNOLOGY

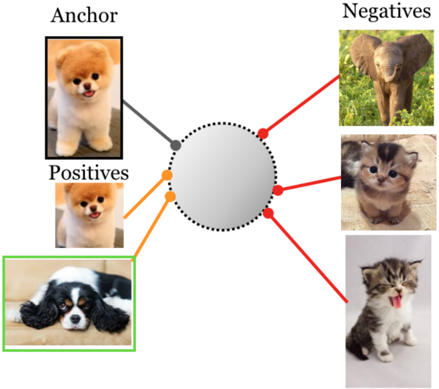
Contrastive Pre-Training of Transformer Models for Computational Framing Analysis

Alexander Ertl

Graz, October 2023

Contrastive Pre-Training of Transformer Models

- Contrastive refers to the training method
- Pre-training is a training stage
- Transformer — a deep learning architecture suited for natural language processing



P. Khosla et al. Supervised contrastive learning. Advances in Neural Information Processing Systems, volume 33, pages 18661–18673. (2020).

Computational Framing Analysis

To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described [Ent93].

Framing Example

Quality of Life

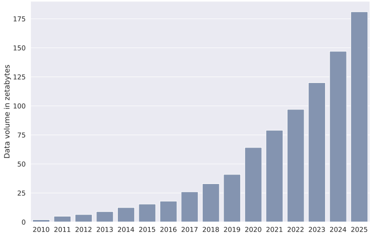
. . . lawyers. . . are aiding the caravan migrants before they attempt to seek asylum in the U.S. . .

Security and Defence

. . . President Donald Trump said he'll mobilize the U.S. military to close the border with Mexico to stop an "assault" on the nation by a caravan of migrants. . .

Motivation

- Frames are omnipresent
- Which frames are prevalent in society?
- Manual framing analysis is slow and expensive
- Exponential growth of data generation



IDC, und Statista. "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes)." Chart. June 7, 2021. Statista. Accessed October 12, 2023. https://www.statista.com/statistics/871513/worldwide-data-created/

Research Questions I

Q1 *How can the training procedure of transformers be optimised to improve computational framing analysis methods?*

Research Questions II

Q2 *How does contrastive learning transform the embedding space in multi-label settings?*

Research Questions III

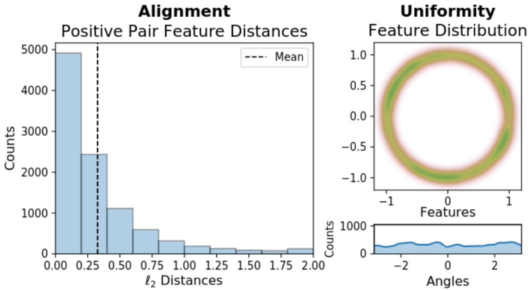
Q3 *How do moral frames differ from topics and how should they be interpreted?*

Research Questions IV

Q4 *Do unsupervised framing detection methods agree with the frames produced by the supervised model and/or do they introduce any new frames?*

Contrastive Learning

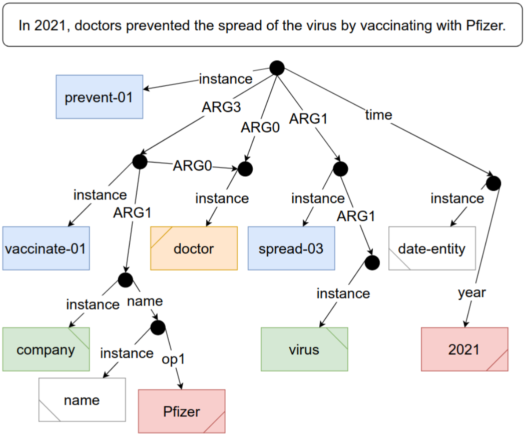
- Optimises for
  - uniformity (representations distributed evenly over entire available space)
  - alignment (positive pairs are close to one another)



Tongzhou Wang and Phillip Isola. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. International Conference on Machine Learning. 2020, pp. 9929–9939.

Narrative Analysis

- Important for human cognition
- Have a *setting or context*, *temporal aspect*, *agents*, and a *moral*
- Extract narratives via AMRs

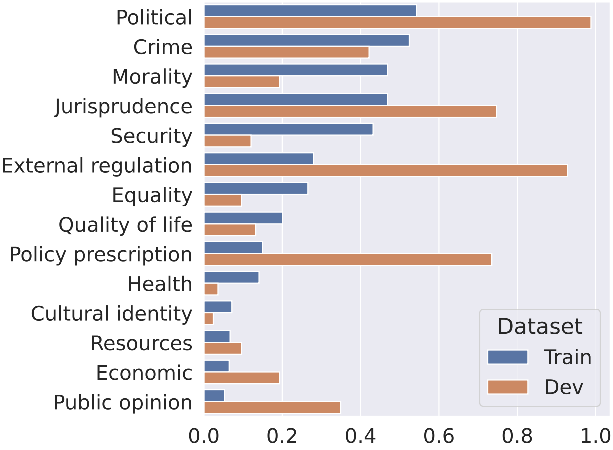


Ongoing work from our research group.

Q1 *How can the training procedure of transformers be optimised to improve computational framing analysis methods?*

SemEval 2023

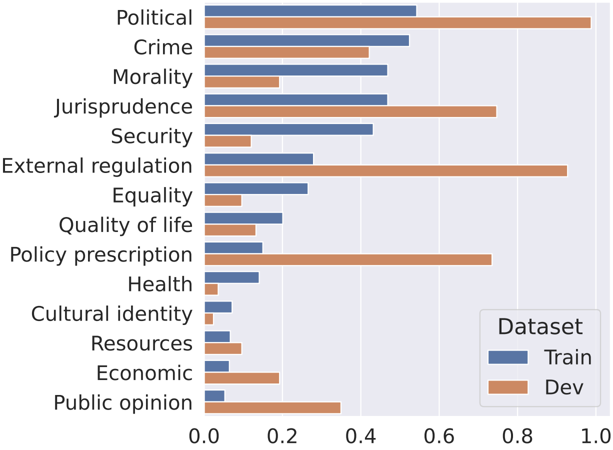
- Advance multilingual frame detection
  - 6 few-shot languages
  - 3 zero-shot languages
- Supervised multi-label classification



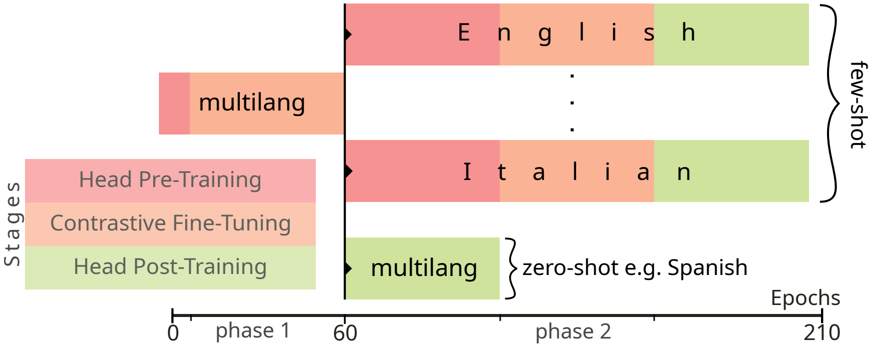
Desiderata

Approach that handles

- Multi-label few-shot setting
- Imbalanced datasets
- Shifts in distribution



The Multilingual Contrastive Pre-Training of Transformers Pipeline



Alexander Ertl, Markus Reiter-Haas, Kevin Innerhofer, Elisabeth Lex. MiniCPT at SemEval-2023 Task 3: Multi-Label-Aware Contrastive Pretraining for Framing Prediction with Limited Multilingual Data. Proceedings of the 17th International Workshop on Semantic Evaluation. SemEval 2023. Toronto, Canada, July 2023.

HeroCon Loss [Zhe+22]

- Multi-Component contrastive loss function
- Positive pairs should be similar
- Negative pairs should be dissimilar
- How close should representations be?

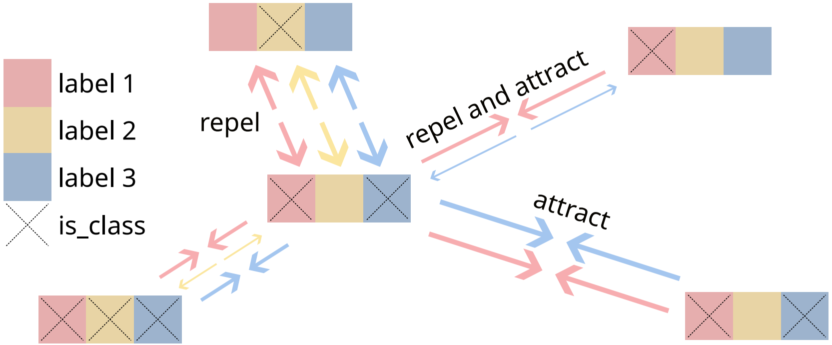
HeroCon Loss

$$\mathcal{L} = \mathcal{L}_{BCE} + \alpha \mathcal{L}_{CON}$$

$\mathcal{L}_{BCE} =$ binary cross-entropy

$\mathcal{L}_{CON} =$ contrastive

Label-Aware Contrastive Learning [Ert+23]



Weighting based on Hamming distances between label vectors represented by arrow thickness

Ablation Study [Ert+23]

Model	en	it	ru	fr	ge	po	$\overline{\Delta}$
mCPT+CS[1]	**.688**	**.590**	.519	**.575**	**.591**	.638	

[1] contrast-sampling

Ablation Study [Ert+23]

Model	en	it	ru	fr	ge	po	$\overline{\Delta}$
mCPT+CS[1]	**.688**	**.590**	.519	**.575**	**.591**	.638	
- CS	.682	.585	**.520**	.570	.561	.636	-.008
- PT[2]	.681	.545	.475	.563	.583	.616	-.015
- $\mathcal{L}_{CON}$	.657	.521	.436	.524	.570	**.645**	-.018

[1] contrast-sampling [2] multilingual pre-training

Ablation Study [Ert+23]

Model	en	it	ru	fr	ge	po	$\overline{\Delta}$
mCPT+CS[1]	**.688**	**.590**	.519	**.575**	**.591**	.638	
- CS	.682	.585	**.520**	.570	.561	.636	-.008
- PT[2]	.681	.545	.475	.563	.583	.616	-.015
- $\mathcal{L}_{CON}$	.657	.521	.436	.524	.570	**.645**	-.018
- E2E[3]	.629	.519	.500	.535	.586	.633	.008

[1] contrast-sampling [2] multilingual pre-training [3] end-to-end training

The Competition

- Most teams employed transformers and trained on all languages
- Best team used ensemble methods and trained on additional data
- Second-best team also used contrastive learning

Spanish - Subtask 2

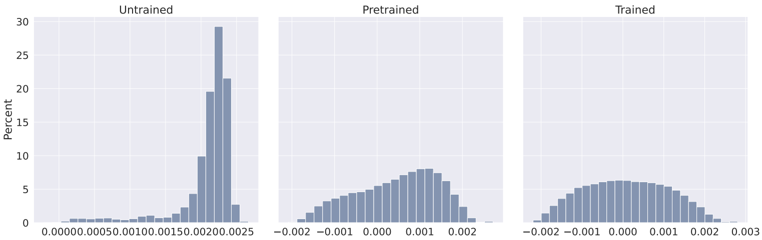
Rank	Team	F1 micro	F1 macro
1	PolarIce (mCPT)	**0.57143**	0.45491
2	UMUTeam	**0.55844**	0.46541
3	vera	**0.55758**	0.52425
4	SinaaAI	**0.55484**	0.44105
5	TeamAmpa	**0.50575**	0.38650
6	MarsEclipse	**0.49032**	0.43161
7	Riga	**0.48889**	0.42605
8	QCRITeam	**0.48780**	0.38955
15	SharoffAndLepekhin	**0.26506**	0.20117
16	FramingFreaks	**0.21538**	0.21132
17	Baseline	**0.12000**	0.09524

Q2 *How does contrastive learning transform the embedding space in multi-label settings?*

Embeddings

$\mathcal{L}_{CON}$ [WI20] in multi-label settings also optimises for

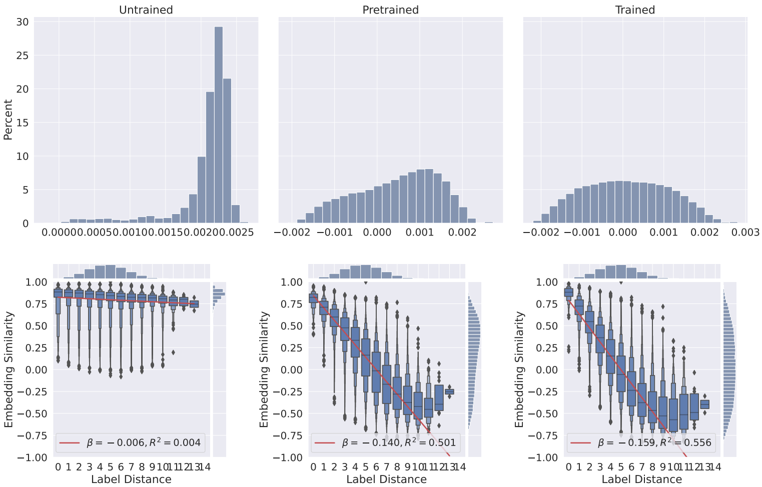
- Uniformity



Embeddings

$\mathcal{L}_{CON}$ [WI20] in multi-label settings also optimises for

- Uniformity
- Alignment



Q3 *How do moral frames differ from topics and how should they be interpreted?*

Frame Keywords

- Goal: find keywords most predictive of frames

  $\Rightarrow$ Logistic regression + coefficient analysis

- Contrast with BERTopic topics by examining
  - conditional probabilities
  - coherence measures

Frame Keywords

- Goal: find keywords most predictive of frames

  $\Rightarrow$ Logistic regression + coefficient analysis

- Contrast with BERTopic topics by examining
  - conditional probabilities
  - coherence measures

Frame Keywords - Health and Safety

levin, mass genocide, wage, infect human, osteoporosis, converging, somatic, watch tv, **guitarist**, 890, kidney cells, debra, cares act, people hiv aids, citizens catastrophic

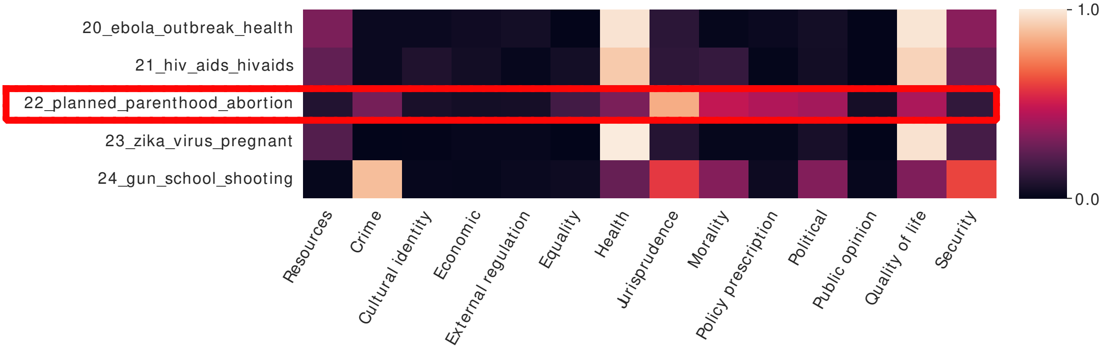
. . . pioneering rock guitarist whose sharp, graceful style helped Elvis Presley shape his revolutionary sound and inspired a generation of musicians . . . died Tuesday.

Frame Keywords - Health and Safety

levin, mass genocide, wage, infect human, osteoporosis, converging, somatic, watch tv, **guitarist**, 890, kidney cells, debra, cares act, people hiv aids, citizens catastrophic

. . . pioneering rock guitarist whose sharp, graceful style helped Elvis Presley shape his revolutionary sound and inspired a generation of musicians . . . died Tuesday.

Frames versus Topics - $\mathcal{P}(\text{frame}|\text{topic})$



Frames versus Topics

Frames and topics are distinct

- Frames are not well defined by their keywords
- Perceived coherence is low
- Topics do not imply frames

Q4 *Do unsupervised framing detection methods agree with the frames produced by the supervised model and/or do they introduce any new frames?*

Unsupervised Frame Detection

- Extraction of subject-predicate-object tuples
- Filtering operations
  - i Frames are employed to hammer home a message
  - ii The dataset is not homogeneous
  - iii Frames employ high-valence keywords

Narrative Frames

Frame	Sentiment	%	Frame	Sentiment	%
disease cause die	-0.60	3.72	terrorist attack	-0.83	1.75
person criticize	-0.38	2.64	person protest	-0.25	1.66
person play	0.34	2.52	person rebel	-0.15	1.33
person pay tax	-0.10	2.11	you risk	-0.27	1.05
person fight	-0.38	2.02	person shoot	-0.34	1.02
person prosecute	-0.40	1.86	i thank you	0.36	1.02
vaccine effective	0.48	1.82	person fight fire	-0.61	0.70
person pay	-0.10	1.81	tooth decay	-0.40	0.52

Top-Level Narratives

With goats, sheep and fruit also testing positive, nothing is safe. Assuming you trust the test.

- nothing safe, test cause safe, you trust test

President Magufuli says tests were found to be faulty after goat, sheep and pawpaw[fruit] samples test positive for COVID-19

- person [Magufuli] say find fault test
- thing cause recommend dismiss kit

Top-Level Narratives

With goats, sheep and fruit also testing positive, nothing is safe. Assuming you trust the test.

- nothing safe, test cause safe, you trust test

President Magufuli says tests were found to be faulty after goat, sheep and pawpaw[fruit] samples test positive for COVID-19

- person [Magufuli] say find fault test
- thing cause recommend dismiss kit

Limitations and Future Work

- Empirical construction of training pipeline
- Frames remain fairly ambiguous
  - Additional datasets
  - Expert validation of frames
- Filtering operations likely discard relevant narratives (e.g. synonymous narratives)
- Construction of top-level narratives

Q1 - Improvements to the Training Procedure

Multilingual Pre-Training

effectively increases the available amount of training data

Contrastive Learning

acts as a regularizer and is essential in low-data situations

Q2 - Transformation of the Embedding Space

Uniformity and Alignment

The embedding space is transformed such that the entire available space is utilised (uniformity) and samples with similar label vectors are close to one another (alignment).

Q3 - Differences between Frames and Topics

Frames are not Topics

The frames extracted by *mCPT* are not well captured by keywords lists or implied by topics.

Q4 - Unsupervised Frame Detection

Narratives are more concrete

Narratives are more granular than frames.

Contextualisation

Topics and narratives provide useful contextual information for the interpretation of conventional frames.

Bibliography I

[Boy+13] Amber E Boydstun et al. **Identifying Media Frames and Frame Dynamics Within and Across Policy Issues.** (2013).

[Ent93] Robert M Entman. **Framing: Toward Clarification of a Fractured Paradigm.** *Journal of communication* 43.4 (1993), pp. 51–58.

[Ert+23] Alexander Ertl et al. **MiniCPT at SemEval-2023 Task 3: Multi-Label-Aware Contrastive Pretraining for Framing Prediction with Limited Multilingual Data.** Proceedings of the 17th International Workshop on Semantic Evaluation. SemEval 2023. Toronto, Canada, July 2023.

[WI20] Tongzhou Wang and Phillip Isola. **Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere.** International Conference on Machine Learning. 2020, pp. 9929–9939.

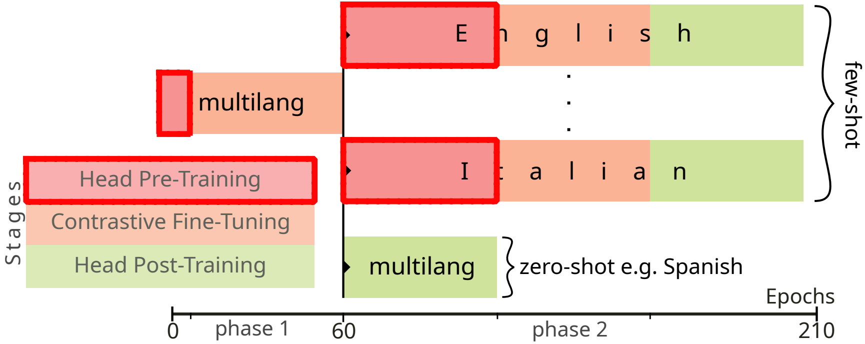
Bibliography II

[Zhe+22] Lecheng Zheng et al. **Contrastive learning with complex heterogeneity.** Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022, pp. 2594–2604.

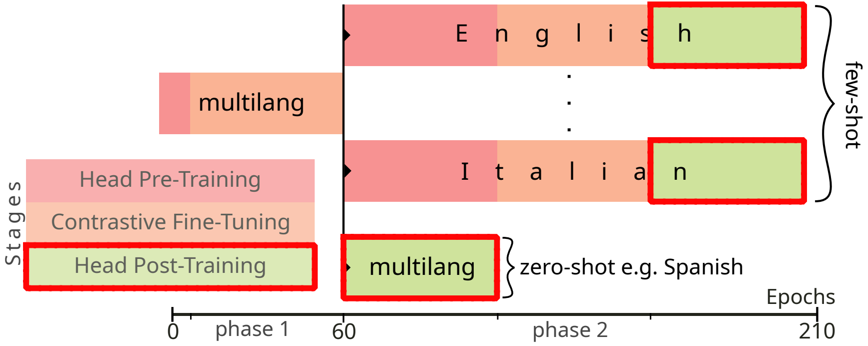
Thank You for Your Attention

https://github.com/lambdasonly/mCPT

The mCPT Pipeline



The mCPT Pipeline



Toy Example [Ert+23]

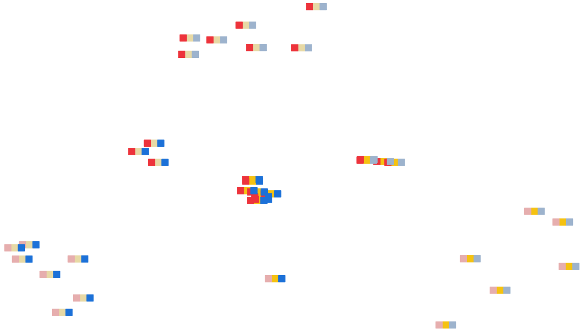
- Randomly sampled embeddings
- PCA projection from 3 dimensions

Toy Example

- HeroCon Loss
- Aligned along axes drawn from the origin

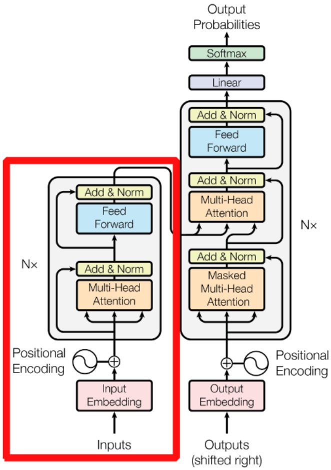
Toy Example

- Euclidean loss
- Curse of dimensionality



Transformers

- Encoder-Decoder model
- Attention layers
- MiniLM distilled BERT



Vaswani et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.

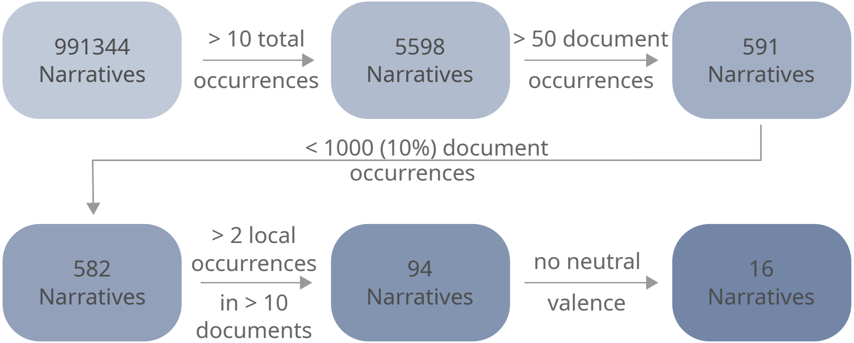
Contrast Sampling [Ert+23]

- Loss function computed on batches
  - at least one sample per class for negative pairs
  - at least two samples per class for positive pairs
- Epoch ends when all samples have been seen at least once
- Samples of less frequently occurring classes have to be reused
  $\Rightarrow$ implicit oversampling

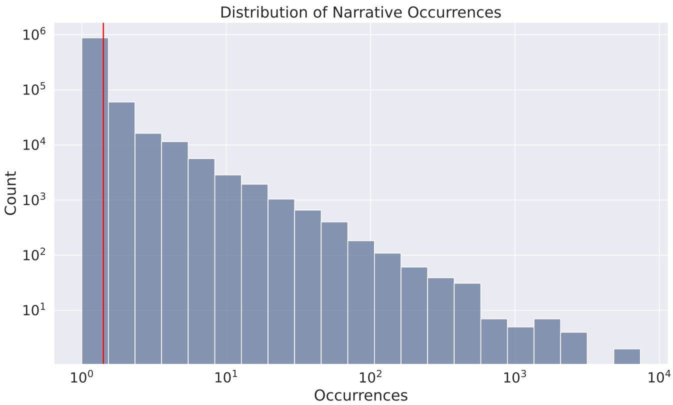
The Policy Frames Codebook [Boy+13]

- “System for categorizing frames across policy issues”
- 14 frames such as *policy*, *health and safety*, or *quality of life*
- E.g. *quality of life frames: The effects of a policy on individuals’ wealth, mobility, access to resources, happiness, social structures, ease of day-to-day routines, quality of community life, etc.*

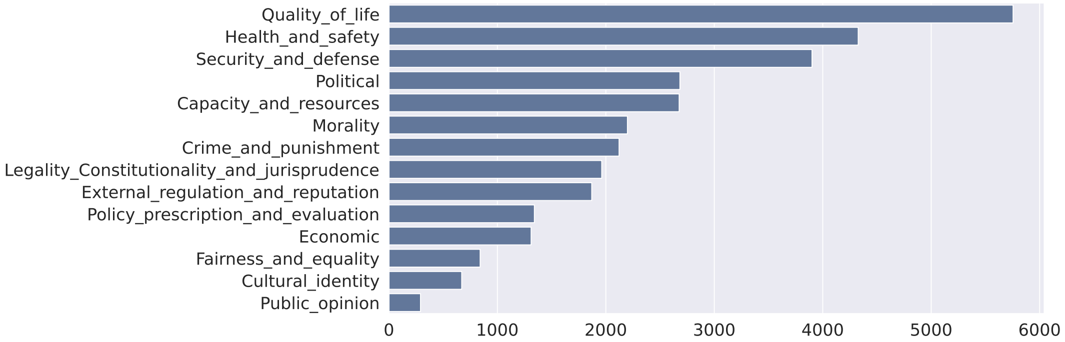
Filtering


991344
Narratives
> 10 total
occurrences
5598
Narratives
> 50 document
occurrences
591
Narratives
< 1000 (10%) document
occurrences
582
Narratives
> 2 local
occurrences
in > 10
documents
94
Narratives
no neutral
valence
16
Narratives

Narrative Distribution



Frame Counts



Topic Coherence Measures

	Frames
Capacity_and_resources	0.68
Policy_prescription_and_evaluation	0.67
Legality_Constitutionality_and_jurisprudence	0.66
Fairness_and_equality	0.66
Economic	0.65
External_regulation_and_reputation	0.64
Quality_of_life	0.64
Morality	0.62
Health_and_safety	0.6
Political	0.59
Security_and_defense	0.58
Cultural_identity	0.56
Crime_and_punishment	0.51
Public_opinion	0.17

	Topics
2_moon_space_apollo_lunar	0.94
4_fluoride_water_teeth_fluoridation	0.92
3_plane_search_mh370_flight	0.87
10_digital_cash_bitcoin_currency	0.79
7_league_arsenal_football_wenger	0.78
12_marijuana_cannabis_cbd_medical	0.78
8_market_analysis_anhydrous_ahf	0.71
6_jonestown_jones_temple_shamo	0.64
11_alien_game_ripley_isolation	0.6
5_brain_neurons_control_mind	0.48
1_bigfoot_species_dogs_reptiles	0.44
0_said_people_us_one	0.35
-1_us_one_people_said	0.34
9_vector_field_hydrogen_derivative	0.29