

## Appendix A

# Types	1,013,024
# Tokens	87,761,626
# Articles	118,532
Average # Articles per month	718

Overview of Izvestia corpus

**Preprocessing** We identified named entities with the [ISPRAS \(texterra\) API](#) and then manually grouped country mentions i.e., collapsing “U.S.A.”, and “Americans” to a single label. We paid particular attention to words referring to the U.S. or Russia and allowed less clean references to other countries. Finally, we tokenized and lowercased all text.

## Appendix B

	<b>Hard</b>	<b>Soft</b>	<b>Avg.</b>
Capacity & Resources	60.00	93.33	76.67
Crime & Punishment	93.33	93.33	93.33
Cultural Identity	73.33	100	86.67
Economic	100	100	100
External Regulation	93.33	100	96.67
Fairness & Equality	20.00	60.00	40.00
Health & Safety	93.33	93.33	93.33
Legality & Constitution.	86.67	93.33	90.00
Morality	46.67	73.33	60.00
Policy Prescription	86.67	100	93.33
Political	73.33	86.67	80.00
Public Sentiment	53.33	86.67	70.00
Quality of Life	66.67	93.33	80.00
Security & Defense	60.00	66.67	63.33

Soft and hard accuracy scores (%) and average precision in Russian framing lexicon intruder detection task