Dissecting Biases in Relation Extraction: A Cross-Dataset Analysis on People's Gender and Origin

Marco Antonio Stranisci[*]
Università degli Studi di Torino
aequa-tech, Turin, Italy
marcoantonio.stranisci@unito.it

Pere-Lluís Huguet Cabot[*]
Sapienza University of Rome
huguetcabot@diag.uniroma1.it

Elisa Bassignana[*]
IT University of Copenhagen
Pioneer Center for AI, Denmark
elba@itu.dk

Roberto Navigli
Sapienza University of Rome
navigli@diag.uniroma1.it

Abstract

Relation Extraction (RE) is at the core of many Natural Language Understanding tasks, including knowledge-base population and Question Answering. However, any Natural Language Processing system is exposed to biases, and the analysis of these has not received much attention in RE. We propose a new method for inspecting bias in the RE pipeline, which is completely transparent in terms of interpretability. Specifically, in this work we analyze biases related to gender and place of birth. Our methodology includes (*i*) obtaining semantic triplets (subject, object, semantic relation) involving 'person' entities from RE resources, (*ii*) collecting meta-information ('gender' and 'place of birth') using Entity Linking technologies, and then (*iii*) analyze the distribution of triplets across different groups (e.g., men versus women). We investigate bias at two levels: In the training data of three commonly used RE datasets (SRED$^{\text{FM}}$, CrossRE, NYT), and in the predictions of a state-of-the-art RE approach (ReLiK). To enable cross-dataset analysis, we introduce a taxonomy of relation types mapping the label sets of different RE datasets to a unified label space. Our findings reveal that bias is a compounded issue affecting underrepresented groups within data and predictions for RE.

1 Introduction

Language technologies are widely spreading throughout our everyday life. However, it has been demonstrated that these technologies are often affected by the presence of gender and racial biases (Kurita et al., 2019; Tan and Celis, 2019). "Bias" is a cover term for a number of issues, which according to Hovy and Prabhumoye (2021) may emerge at any stage of the Natural Language Processing (NLP) pipeline. They could come from the data curation process (Sap et al., 2019), be intrinsic into the trained model (Zhao et al., 2017), or they could derive from the cultural background of NLP practitioners (Santy et al., 2023). An orthogonal taxonomy of biases distinguishes between *allocative* and *representational* ones (Suresh and Guttag, 2021). *Allocative* biases regard the unequal distribution of opportunities across different groups, such as disparity in granting loans (Hardt et al., 2016) or the systematic exclusion of certain minorities from public archives (Weathington and Brubaker, 2023). *Representational* biases focus on stereotypical associations between groups and certain features (Caliskan et al., 2017) (e.g., women and lexicon about marriage and parenthood). Blodgett et al. (2020) show that existing works in NLP mainly focus on *representational* biases while the *allocative* ones are often overlooked.

In this context, Relation Extraction (RE) techniques represent a powerful tool to jointly explore the two types of bias described above. RE methods extract fine-grained triples from texts (subject, object, and the semantic relation connecting them), allowing for the discovery of gaps in digital archives. Previous work performed event extraction on Wikipedia biographies to study the presence of systematic gender biases in this archive (Sun and Peng, 2021; Stranisci et al., 2023). Gaut et al. (2020) collected a distantly supervised dataset from Wikipedia for exploring gender bias in RE, but they only include four relation types ('spouse', 'hypernym', 'birthDate', 'birthPlace'). Despite this preliminary work, standards for the adoption and evaluation of RE techniques for bias detection are still missing and are limited to the analysis of gender. Furthermore, before using RE for bias detection there is the pressing need to explore whether these systems portray any themselves.

In this paper, we explore the presence of biases in RE, both at the level of data (by analyzing the training data) and model (by analyzing the model

[*] Equal contribution

Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP), pages 190–202
August 16, 2024 ©2024 Association for Computational Linguistics

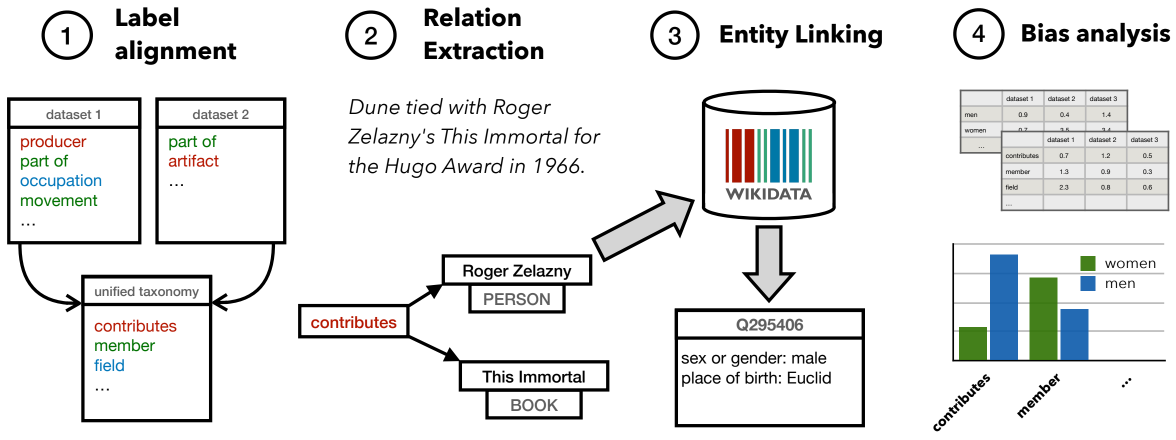

Figure 1: **Overview of our Proposed Methodology.** The first step aligns the label sets of different RE datasets into a unified taxonomy of relation types. In the second step, we extract semantic triplets including 'person' entities. Within the third step, we collect socio-demographic information from Wikidata of the people extracted in the second step. Finally, in the last step we analyze potential *allocative* and *representational* imbalances in the distribution of the extracted information (entities and relations) across different social groups (e.g., men versus women).

predictions). We illustrate our procedure in Figure 1. As a first step, in order to enable cross-dataset analysis, we introduce a taxonomy of relation types mapping the label sets from different RE datasets into a unified label space. Then, as a second and third steps we collect information about people mentioned in a text. This includes semantic relations involving people (from RE), and meta-information related to them (i.e., 'gender' and 'place of birth'; using Entity Linking). As a last step, we explore the *allocative* and *representational* biases by inspecting potential imbalances into the distribution of the extracted triples across different groups (e.g., men versus women). Concretely, we investigate if any relation type (e.g., *member*, *contributes*) is more likely associated with one social group (more details in Section 5). We repeat our procedure both on the training sets on three widely adopted RE datasets: SRED$^{\text{FM}}$ (Huguet Cabot et al., 2023), CrossRE (Bassignana and Plank, 2022a), NYT (Riedel et al., 2010); and on the predictions of a state-of-the-art RE approach, ReLiK (Orlando et al., 2024).

Not only do our findings corroborate existing research regarding the prevalence of gender biases in RE but they also broaden the discourse by uncovering biases along additional dimensions, such as origin. To our knowledge, this is the first investigation that examines bias through the lens of transfer learning and reveals the nuanced effects of simplistic interventions like data balancing. While such strategies may reduce biases for certain target groups, they can inadvertently introduce new biases, underscoring the necessity for a more sophisticated, multi-axial approach for bias mitigation.

The contributions of this paper are:

- We introduce a meticulous bias analysis procedure for RE designed to be applicable across various dimensions, addressing both dataset and model-level biases.
- An in-depth analysis of biases related to 'gender' and 'place of birth' in the train sets of three widely adopted RE datasets and on the predictions of a SotA RE model on those.
- A taxonomy of relation types mapping the label sets of different RE datasets into a unified label space. The taxonomy makes our approach robust and versatile, and opens to cross-dataset analysis.

2 Related Work

Sun et al. (2019) and Blodgett et al. (2020) emphasize current issues in the research about bias detection and mitigation. The first presents a survey aimed at identifying research directions for gender bias detection, while the second criticizes how research in bias detection and mitigation is usually conducted. In order to make explicit potential biases in NLP, Bender and Friedman (2018) and Mitchell et al. (2019) propose to better document datasets and Language Models (LMs) respectively.

Some works released ad-hoc datasets to explore bias detection. Zhao et al. (2018) presented WinoBias, a dataset for coreference resolution aimed at testing stereotypical associations between women and certain types of profession. Nadeem et al. (2021) introduced StereoSet, for testing the presence of stereotypical knowledge in LMs while Gehman et al. (2020) released RealToxicityPrompt, a list of annotated prompts that is intended to measure the toxicity of text generated by LMs. Kiritchenko and Mohammad (2018) presented the Equity Evaluation Corpus, designed to measure gender and racial biases in models trained for sentiment analysis.

Several work on bias analysis focuses on inspecting the internal representation of NLP models. Caliskan et al. (2017) proposed two metrics for bias detection from word embeddings; May et al. (2019) from sentence encoders; and Kurita et al. (2019) from contextualized word embeddings. More recent approaches in this direction use probing strategies (Lauscher et al., 2022; Köksal et al., 2023). However, the outcome of these methods is often hard to interpret because of the black box nature of neural models. In order to prioritize interpretability of the results and obtain a more transparent bias analysis, we propose a new procedure for bias detection in RE technologies, which is applicable both at the level of data and model.

3 Methodology

We introduce a four-step procedure for detecting biases related to 'gender' and 'place of birth' in the Relation Extraction pipeline (see Figure 1). The method can be easily extended to explore other socio-demographic biases.

① First, we align the label spaces of different RE datasets using a unique taxonomy of relations with the aim of performing comparable analysis across corpora (details in Section 3.1).

② As a second step, we employ Relation Extraction in order to gather triplets (subject, object, relation) about people mentioned in a text. This can be done by filtering the triplets in which at least one of the two entities has type 'person'. We leverage the triplets in labeled training sets as well as in the predictions of systems trained using them.

③ We collect socio-demographic data about people that are included in the biographical triplets extracted in step ②. We use Entity Linking (EL) to disambiguate the entity spans with type 'person' and link them to Wikidata (Vrandečić and Krötzsch, 2014) entries. We collect two types of meta-information from Wikidata: 'gender' and 'place of birth'.

④ Last, given the triplets extracted in the second step and the socio-demographic information collected in the third step, we conduct bias analysis by investigating any imbalance in the distribution of relations across different social groups (e.g., men versus women). Since it has been demonstrated that biases may occur at any stage of the NLP pipeline (Hovy and Prabhumoye, 2021), we applied our procedure for assessing the presence of biases both on the corpora used for training RE models and on the entities and relations predicted by them. Specifically, we investigate *allocative* bias in the training data (Section 5.1) and in the predictions made by these models (Section 5.3). Similarly, we examine *representational* bias, adapting metrics from earlier studies to evaluate both the training datasets (Section 5.2) and the predictions (Section 5.4).

3.1 Relation Type Taxonomy

RE datasets often include a label set with relation types which are too fine-grained with respect to our objective of exploring social biases related to 'gender' and 'place of birth' (e.g., *field-of-work* and *occupation* from SRED^{FM}). Aggregating certain types to broader categories enables a higher-level analysis with enough samples per type that would be otherwise unfeasible with infrequent or narrow ones. In addition, we face the issue of lack of standards in dataset annotation for RE (Bassignana and Plank, 2022b), which prevents the comparison of results across corpora (e.g., the relation type */people/person/profession* in NYT versus *occupation* in SRED^{FM}). To overcome these issues we introduce a taxonomy of relation types mapping the original types from the different datasets into a unified label space (e.g., *field-of-work*, *occupation* and */people/person/profession* to *field*). The taxonomy enables cross-dataset comparison and makes our approach versatile. Table 1 reports the ten newly introduced labels, with the co-occurring entity types (one entity type is always a person), and a corresponding example. The taxonomy is organized around the entity types that are part of the triplet. For instance, *contributes* is used to identify all triples with a person and a work, while *relationship* represents triplets where both subject and object are persons.

Relation type	Co-occurring entity	Example
contributes	work	In 2018, *Zhao* directed her third feature film, *Nomadland*, starring Frances McDormand
date	date	*Rosa Luxemburg* born Rozalia Luksenburg, *5 March 1871*
field	occupation, discipline	*Stephen William Hawking* was an English *theoretical physicist*, *cosmologist*
geographical relation	place	Born in *Ogidi*, Colonial Nigeria, *Achebe*'s childhood was influenced by both Igbo traditional culture and postcolonial Christianity
language	language	*Seedorf* speaks six languages fluently: *Dutch*, *English*, *Italian*, *Portuguese*, *Spanish* and *Sranan Tongo*
member	organization	Ahead of the 2009–10 season, *Ronaldo* joined *Real Madrid* for a world record transfer fee at the time of £80 million (€94 million)
participated	event	*Tim Burton* appeared at the *2009 Comic-Con* in San Diego, California, to promote both 9 and Alice in Wonderland
position held	organization	*Meredith Whittaker* is the president of the *Signal Foundation* and serves on their board of directors
relationship	person	*Billy Porter* married *Adam Smith* on January 14, 2017, after meeting him in 2009
topic	work	*Napoleon* appears briefly in the first section of Victor Hugo's *Les Misérables*, and is extensively referenced in later sections

Table 1: **Relation Type Taxonomy.** A list of biographical situations designed for RE. Labels are distinguished on the basis of the co-occurring entities in a triple. All examples are derived from the English Wikipedia.

	Train		**Validation**		**Test**	
	sent.	rel.	sent.	rel.	sent.	rel.
SRED$^{\text{FM}}$	1,199,046	2,480,098	6,333	13,322	3,015	6,474
CrossRE	297	1,220	835	3,483	891	3,604
NYT	19,709	26,267	1,765	2,318	1,773	2,327

Table 2: **Dataset Statistics.** Number of sentences and number of triplets (relations) for each dataset.

4 Experimental Setup

We follow the four-step procedure described in Section 3 to investigate biases in three commonly adopted RE datasets, and the predictions of a popular RE model. Below, we describe our experimental setup in terms of datasets (Section 4.1) and modeling (Section 4.2). Details about their licenses can be found in Appendix B.

4.1 Datasets

SRED$^{\text{FM}}$ (Huguet Cabot et al., 2023). The SRED$^{\text{FM}}$ datasetis a distantly annotated dataset build on top of Wikipedia pages and Wikidata relations, employing a novel triplet critic filtering. The dataset covers 17 languages, but for the scope of this paper we employ only the English portion. Since this is the larger corpus in our study, we use it as a pre-training stage for the experiments on the other two datasets.

CrossRE (Bassignana and Plank, 2022a). CrossREis a multi-domain dataset for RE containing data from the news, politics, natural science, music, literature and artificial intelligence domains. This dataset is the only entirely manually-annotated in our study. Given the small size of the six subsets, in our experiments we join the data across the different domains.

NYT (Riedel et al., 2010). NYT is a RE dataset consisting of news sentences from the New York Times corpus. It contains distantly annotated relations using FreeBase. We use the processed version of Zeng et al. (2018) called NYT-multi.

For each of these datasets, we filter the triplets which include at least one entity 'person'. In Table 2 we report the statistics of the corpora after the filtering phase. In addition, following step ① in Section 3, we map the original relation types of the three datasets, into a unified label space defined by our taxonomy of relation types (Section 3.1). We report our mapping in Table 8 in Appendix A.

4.2 Models

In steps ② and ③ of our proposed procedure (described in Section 3) we employ a Relation Extraction (RE) and an Entity Linking (EL) model respectively. Below we describe them both.

ReLiK (Orlando et al., 2024). For RE, we employ the same setup as ReLiK, a Retriever-Reader model based on DeBERTa-v3 (He et al., 2021). We use the same default parameters as the original paper and train on top of DeBERTa-v3-large.

EntQA (Zhang et al., 2022). To disambiguate the extracted entities 'person' and link them to Wikidata (Vrandečić and Krötzsch, 2014) we use EntQA, a recent state-of-the-art EL system based on the Retriever-Reader paradigm. We employ it to perform entity disambiguation on the entity spans extracted by ReLiK. We only default to these predictions when the original dataset does not have a link to Wikidata, either because a span prediction was not labeled as an entity in the dataset, or because the original dataset did not include disam-

			+ SRED^{FM} pre-train		
	Test	taxonomy	original	taxonomy	balanced
	SRED^{FM}		69.13	71.07	64.84
zero-shot	CrossRE		17.35	20.27	20.07
	NYT		28.58	32.89	33.66
fine-tuned	CrossRE	44.72	51.74	52.04	52.12
	NYT	89.26	88.47	88.52	89.83

Table 3: **Experiments Performance.** Micro-F1 scores of ReLiK trained and evaluated on SRED^{FM}, zero-shot and fine-tuning evaluation on CrossRE and NYT. 'original' refers to a model trained on the original label set; 'taxonomy' indicates that the model was trained on the taxonomy mapping (see Table 8); 'balanced' stands for a gender-balanced version of it (see Section 6). First row indicates performance after pre-training on SRED^{FM} test set.

biguated entities. We use EntQA out-of-the-box (i.e., we do not fine-tune it on our datasets).

4.3 Relation Extraction Experiments

As mentioned in Section 4.1, we use SRED^{FM} for pre-training ReLiK before employing it on the two smaller datasets (CrossRE, NYT). We perform two categories of experiments: 'Zero-shot', where ReLiK is pre-trained on SRED^{FM} and directly evaluated on CrossRE and NYT; and 'fine-tuning', where ReLiK is both pre-trained on SRED^{FM} and fine-tuned on the target dataset.

Zero-shot Experiments. In Table 3 we report the scores of ReLiK trained on SRED^{FM} and evaluated on CrossRE and NYT in a zero-shot fashion. Evaluation is always done in the coarse-grained space of the taxonomy, either on the predictions of a model trained on SRED^{FM} mapped to the taxonomy (column 'taxonomy'), or by mapping the predictions of a model trained on the original labels to the taxonomy (column 'original'). Training on the taxonomy relation types improves the performance for both datasets. These results validate our proposed mapping as a way to unify label sets from different datasets.

Fine-tuning Experiments. Similarly to the previous experiment, in Table 3 we report the scores of ReLiK trained on SRED^{FM} and then fine-tuned on CrossRE or NYT, as well as regular fine-tuning without pre-training (left column). These experiments allow us to explore the use of our shared label space as a means of transfer learning across datasets and later on study how transfer learning affects the bias distribution (see Sections 5.3 and 5.4).

	SRED^{FM}	CrossRE	NYT
Women	20.0%	11.8%	17.3%
Global South	18.9%	10.0%	12.2%

Table 4: **Allocative Bias in Training Data.** The percentage of women and Global South people in SRED^{FM}, CrossRE, and NYT corpora.

Differences in performance are smaller than in the zero-shot counterpart, especially when enough data is available in the target dataset (NYT). Still, this experiment showcases that pre-training on the taxonomy improves performance on low data regimes while it has a small difference on larger ones.

5 Social Bias Analysis

In this section we report our bias analysis conducted on the training sets of the datasets described in Section 4.1 and on the predictions obtained with our trained models. In line with previous work on 'gender' bias analysis, we consider *men* versus *women* (Zhang and Terveen, 2021). For biases related to the 'place of birth', instead, we follow previous work and consider *Global North* versus *Global South* (Dirlik, 2007). Such a distinction has been introduced by the Brandt Commission (Williams, 1980) in the context of an effort of reducing economic issues affecting Third World's countries. Therefore, we design an operational definition of country belonging to the Global South as being a former colony and having a Human Development Index lower than 0.8. We discuss more in details these division in the Limitation Section. We maintain the distinction between *allocative* and *representational* biases and explore both bias types at the level of training sets (Sections 5.1 and 5.2) and in the predictions (Sections 5.3 and 5.4).

5.1 Allocative Bias in Training Data

To assess the *allocative* bias in training data we compare the distributions across two axes between entities that are included in SRED^{FM}, CrossRE, and NYT: The distribution of women against men, and of people born in a Global South countries against ones born in the Global North. As explained in Section 3 we gather this meta-information about people from Wikidata, a collaborative knowledge graph that is part of the Wikimedia ecosystem. Since the analysis relies on metadata extracted from Wikidata, we are only able to compare people whose information about their 'gender' (Wikidata ID P21)

	SRED$^{\text{FM}}$		CrossRE		NYT		SRED$^{\text{FM}}$		CrossRE		NYT	
	M	W	M	W	M	W	N	S	N	S	N	S
contributes	0.28	0.475	**0.407**	0.291	–	–	**0.758**	0.162	**0.447**	0.333	–	–
date	**1.038**	0.926	–	–	–	–	**1.07**	0.993	–	–	–	–
field	**0.388**	0.291	–	–	–	–	0.394	0.451	–	–	0.002	0.0
geographical	**0.469**	0.368	0.218	0.218	**3.251**	2.164	0.501	**0.64**	0.198	**0.644**	0.965	**1.019**
language	0.013	0.006	–	–	–	–	0.025	0.024	–	–	–	–
member	0.21	0.164	0.229	0.218	**0.739**	0.283	0.252	0.201	0.300	0.222	**0.169**	0.121
participated	0.088	0.049	**0.278**	0.145	–	–	0.052	0.08	**0.218**	0.133	–	–
position held	0.091	0.038	0.745	0.727	**0.085**	0.012	0.144	**0.196**	0.742	**1.200**	**0.036**	0.009
relationship	0.124	**0.215**	**0.098**	0.036	0.078	**0.211**	0.132	0.119	0.093	0.111	**0.077**	0.025
topic	0.001	0.001	0.018	0.018	–	–	0.002	0.002	**0.013**	0.0	–	–

Table 5: **Representational Bias in Training Data.** Results of the experiment aimed at identifying statistically-significant differences between social groups for each relation and across corpora. Values represent the proportion of each relation type per person. First six columns report the comparison between men (M) and women (W); last six between Global North (N) and South (S) people. For each relation, we report the group that is significantly more associated with it in bold, if neither is it means that there is not a statistically significant difference ($p \geq 0.05$).

and 'place of birth' (Wikipedia ID P19) are available. This did not have an impact on the analysis of 'gender', while the Wikidata gap with respect to 'place of birth' is 31% of people from SRED$^{\text{FM}}$, 8% from CrossRE and 11% from NYT. Once we obtained this information, in Table 4 we observe the distribution of women and Southern people in order to understand to which extent they are underrepresented in RE corpora. CrossRE is the corpus where both categories are less represented while in SRED$^{\text{FM}}$ they benefit from a higher representation. Overall, the analysis shows a significant underrepresentation of women and people born in the Global South across all corpora, always falling in a range between 10% and 20% of the total. This is even more daring when considering that the Global South accounts for around 80% of the world population. We also want to stress that these allocative biases are compounded from several sources. All our datasets are in English, and from sources that target an English speaking audience. Wikidata and Wikipedia showcase a skewed gender distribution where only 25% and 20% respectively of people's pages are women (Zhang and Terveen, 2021), furthermore Wikipedia collaborators are 83% male.[1] The annotation process for each of the datasets we analyze may also introduce further biases. Our goal here is not to pinpoint where these biases originated but rather how they are reflected in RE resources.

5.2 Representational Bias in Training Data

The analysis of *representational* biases relies on a Monte Carlo experiment that simulates a balanced distribution of people along the axes of 'gender' (men *vs* women) and 'place of birth' (Global North *vs* Global South). For each training set we perform an experiment structured in three parts: (*i*) We randomly pick 100 individuals for each group and average the number of relation in which they are subject or object. (*ii*) We repeat the sampling 10 times for each distribution. (*iii*) For each relation type we calculate the t-test statistics between the 10 mean scores of a majority and a minority group. Results are reported in Table 5. For each relation we report the average per social group and whether there is a significant difference between the two groups. The comparison between genders shows that *member* and *position held* are significantly related to men in the NYT corpus, perhaps due to its nature as a news corpus, along with *geographical* (also in SRED$^{\text{FM}}$). *Relationship* is instead skewed towards women in SRED$^{\text{FM}}$ and NYT, and towards men in CrossRE. From the comparison between Global North and South it emerges that the latter are always more associated to *geographical*. The *position held* property behaves differently across corpora: It is mostly related to South in SRED$^{\text{FM}}$ and CrossRE, and to North people in NYT, which is also skewed towards this group for the *member* relation. *Relationship* is significantly associated to Global South people only in NYT.

In general, some trends emerge when comparing across datasets. The only gender bias that fa-

[1] https://en.wikipedia.org/wiki/Wikipedia:Wikipedians

	SRED$^{\text{FM}}$	CrossRE	NYT
Women	–	- 2.2%	+ 5.6%
+ SRED$^{\text{FM}}$	- 3.5%	- 5.8%	+ 0.6%
+ gen. balanced	- 2.9%	- 4.4%	0.0%
Global South	–	- 8.3%	- 2.1%
+ SRED$^{\text{FM}}$	- 1.7%	- 6.7%	- 1.6%
+ gen. balanced	- 0.3%	- 9.9%	- 5.9%

Table 6: **Allocative Bias in Prediction.** Percentage difference of women and Global South people in false positive and true positive predictions of the model when trained on each dataset (first row), fine-tuned on top of SRED$^{\text{FM}}$ pre-training (second row) or fine-tuned on top of a gender-balanced SRED$^{\text{FM}}$ pre-training (third row).

	gender			place of birth		
	SRED$^{\text{FM}}$	CrossRE	NYT	SRED$^{\text{FM}}$	CrossRE	NYT
contributes	+ 0.03	- 0.01	–	+ 0.04	- 0.30	–
date	+ 0.03	–	–	- 0.05	–	–
field	+ 0.05	–	–	- 0.03	–	–
geographical	- 0.09	+ 0.16	+ 0.04	+ 0.23	+ 0.15	+ 0.05
language	–	–	–	+ 0.29	–	
member	- 0.12	- 0.10	–	–	- 0.10	–
participated	- 0.07	- 0.01	–	+ 0.10	- 0.06	–
position held	- 0.17	0.00	- 0.01	+ 0.10	- 0.04	- 0.02
relationship	+ 0.07	+ 0.14	- 0.17	+ 0.15	+ 0.03	+ 0.16
topic	–	–	–	–	–	–

Table 7: **Representational Bias in Prediction.** The Rec_{Gap} on the evaluation triples with respect to the underrepresented groups (i.e., positive values for women and people from the Global South). '–' means that the relation type appears less than 10 times.

vors women concerns *relationship*, while all the other types (when significant) skew towards men, independently of the dataset. On the other hand, with respect to the North/South analysis, biases are more widespead and of different nature. Of the three datasets, SRED$^{\text{FM}}$ shows less biases on this dimension, and coincidentally it is the one having a higher percentage of people from the Global South (see Table 4). It is worth noticing how the only bias favoring North shared across datasets (with a very high degree in SRED$^{\text{FM}}$) is *contributes*, which may be reflective of an overall cultural bias within the English Wikipedia, from which both SRED$^{\text{FM}}$ and CrossRE are collected.

Summarizing, the analysis shows the presence of recurring *representational* biases against underrepresented groups, specifically for certain relation types: *relationship* for women, *geographical* for Global South. NYT includes the highest number of biases, where men and Northern people mostly appear in relations that emphasize their profession (*member*, *position held*).

5.3 Allocative Bias in Prediction

Our analysis on bias in predictions follows that of Gaut et al. (2020). For *allocative* bias we rely on the False Positive Balance score (FP_{Bal}) inspired by Hardt et al. (2016). This metric is a comparison between the percentage of entities belonging to an underrepresented group in the model's wrong predictions and their distribution in the test and evaluation sets. A positive delta between these two percentages is interpreted as the model tendency to recognize entities from an underrepresented group. The analysis is performed on predictions obtained with and without SRED$^{\text{FM}}$ pre-training, while always fine-tuning on the target dataset (Table 3). This allows to assess the impact of SRED$^{\text{FM}}$ pre-training on the distribution of bias. Table 6 shows that women and Global South people are affected by *allocative* harms in different proportions and that these vary across corpora. The FP_{Bal} score is negative for women in CrossRE, while in NYT it is positive. Using the pre-trained model before fine-tuning amplifies this bias in CrossRE (from -2.2 to -5.8), while it lowers it in the NYT (from +5.6 to +0.06). The opposite happens if Global South people are considered. Given the fact that a negative FP_{Bal} emerges in all distributions, the pre-training step reduces this bias from -8.3 to -6.7 in CrossRE and from -2.1 to -1.6 in NYT.

In summary, while adopting SRED$^{\text{FM}}$ for transfer learning to CrossRE and NYT has a positive effect on the performance (CrossRE goes from 44.72 to 52.04, see Table 3), it has a mixed effect with respect to the biases. On one side, it amplifies the *allocative* biases for women in predictions, on the other it introduces a mitigation in favor of people from Global South. This could be explained by SRED$^{\text{FM}}$ showing a lower starting bias of -1.7 compared to the other datasets, and therefore acting as a mitigator when used as a pre-trained model. The opposite is observed for women, where SRED$^{\text{FM}}$ has a higher starting bias (-3.5).

5.4 Representational Bias in Prediction

We perform the *representational* bias analysis on the predictions by adopting the *Minority Recall Gap* metric (Rec_{Gap}). Inspired by the 'true positive rate gender gap' from De-Arteaga et al. (2019), our metric measures the differences in recall for predictions of two groups. Since the data used for evaluation is unbalanced and some relation types are rare, we only compute the Rec_{Gap} for types appearing at least 10 times in each corpus.

Table 7 shows the Rec_{Gap} for each relation throughout all datasets. A positive value means that the model is more likely to retrieve a relation if it is associated to an underrepresented group (i.e., women and people from the South); on the opposite, a negative value means that the model is more likely to retrieve the relation type if it includes men or people from the Global North respectively. The analysis shows patters that already emerged in the training sets (Section 5.2). *Relationship* and *geographical* triples are more often retrieved when a woman or a Global South person represents its subject or object in five out of six cases. The only exceptions are SRED$^{\text{FM}}$, which achieves a Rec_{Gap} score of -0.09 in favor of men for *geographical*, and NYT, with a score of -0.17 in favor of men for *relationship*. The opposite happens for *position held*, which is mostly retrieved for Global South ($+0.10$) only in SRED$^{\text{FM}}$, while in all the other cases it always leans towards Global North. *Contributes* achieves a positive Rec_{Gap} in SRED$^{\text{FM}}$ and a negative one in CrossRE for both bias analysis, while *member* is always mostly associated with men or people from the North. The same happens for *participated*, except for 'place of birth' in SRED$^{\text{FM}}$. Finally, *field* and *date* are more associated with women and Global North.

These results mostly follow the trends in the training datasets (Section 5.2). Representational biases in predictions regard similar associations between certain categories of people and relation types: Women with *relationship*, Southern people with *geographical*, men and Northern people with *member*. However, the model seems to have its own impact on the propagation of biases. For instance, *field* does not present statistically significant differences between Global North and Global South in the training sets (see Table 5), but it is mostly associated to Northern people in the predictions. This behavior underlines the need of designing approaches for bias detection that encompass all the stages of the RE task.

6 Bias Mitigation

In this section we look at a common approach to tackle skewed distributions of data by balancing the pre-training data (SRED$^{\text{FM}}$) in order to obtain fairer representations of underrepresented groups. This mitigating strategy was the only one shown to be effective in Gaut et al. (2020). Since in Table 6 the 'gender' bias of SRED$^{\text{FM}}$ is more pronounced with respect to the bias related to the 'place of birth' (-3.5% versus -1.7%), we consider the 'gender' axis and re-train ReLiK on a dataset with a balanced distribution across genders. In order to do so, we gather from SRED$^{\text{FM}}$ all triplets involving at least one woman, and then we add triplets involving men until we reach an equal amount. As a results, we obtained a dataset of $836,638$ instances, of which 50.7% involves at least one woman.

As it can be observed in the bottom line of Table 6, the adoption of a gender balanced pre-training dataset has a mitigation effect on the *allocative* biases against both underrepresented groups in SRED$^{\text{FM}}$. The FP_{Bal} decreases from -3.5% to -2.9% against women and from -1.7% to -0.3% against Southern people. The effect on the gender bias of the other datasets is also positive. The balanced distribution improves the FP_{Bal} score from -5.8% to -4.4% in CrossRE, and from $+0.06$ to 0 in the NYT corpus. However, balancing the gender axis has a negative impact on the *allocative* bias against people from the Global South both in CrossRE and NYT. In CrossRE, it amplifies them from -8.3% to -9.9%, while in the NYT corpus from -2.1% to -5.9%. This could be explained by the drop of presence of Southern people in SRED$^{\text{FM}}$ from 18.9% (see Table 4) to 16.9% in the balanced version. An intersectional approach (Crenshaw, 2017) that jointly considers these two sources of underrepresentation could be explored to better understand how to mitigate biases from multiple angles.

7 Conclusion

In this paper we address the critical matter of biases within RE data and systems, and propose a four-step procedure to analyze them. Our approach showcases the widespread nature of biases in the life-cycle of RE systems, encompassing datasets, transfer learning and model predictions. Our findings reveal a concerning underrepresentation of women and individuals from the Global South as well as undesired biases for specific relation types. We demonstrate that tackling bias is a complex and compounded issue which requires careful thought. Simple techniques, such as balancing the data for an underrepresented group, may introduce other unwanted biases. We also provide a carefully designed taxonomy of relation types that enables comparison and effective transfer across RE datasets.

In conclusion our work serves a dual purpose: On one side, it sheds light on the pervasive biases related to gender and origin within RE datasets and systems, on the other it offers a critical perspective on the use of Information Extraction (IE) techniques for bias exploration. This study emphasizes the need for nuanced, multi-faceted approaches to detect and mitigate biases, urging the community to proceed with caution and depth in developing and applying RE technologies.

Bias Statement

In this paper we study the presence of bias in RE models and datasets focusing on two axes: gender (women *versus* men) and origin (Global South *versus* Global North). RE techniques are crucial to extract structured information from unstructured texts and this could lead to a number of downstream tasks, such as the automatic population of knowledge bases or the development of tools for data management and archiving. Biased RE resources can lead to *allocational* harms, since they might exclude people from datasets and models outputs. Additionally, they can represent a *representational* harms for their systematic association between certain categories of people and specific relation types. In this work we present an approach that consider *representational* and *allocational* harms both in datasets and models, since we believe that it is necessary to implement a comprehensive strategy to reduce the harmfulness of RE systems.

Limitations

The first limitation of this work regards the taxonomy adopted for distinguishing people on the basis of their 'place of birth' in the context of a globalized world. We adopt the distinction between Global North and Global South as it has been recently re-proposed as a framework by the United Nations. However, such a conceptualization has been proposed in a Western context and thus might have an impact on the cultural representation of this underrepresented group. Therefore, we design an operational definition of country belonging to the Global South as being a former colony and having a Human Development Index lower than 0.8. In addition, it is worth mentioning that Wikidata comes with many limitations in its taxonomy that hamper a fair collection of data. Squeezing two orthogonal features like 'gender' and 'sexual orientation' in a unique property is not fully respectful of non-binarism. Not only that: the percentage of people who do not identify as men or women in Wikidata is so low that it was non possible to adopt a binary conception of gender in this research. Future work will rely on knowledge bases with a higher representation of non-binary people. .

The second limitation regards the usage of Wikidata for the collection of socio-demographic information about people. The underrepresentation of women and people from the Global South in this knowledge base is a known issue that may impact the analysis. People from the Global South correspond to 85% of the world population, while in Wikidata they represent only the 17.2%. Women are 24.1% in Wikidata, against 49.7% in the real world. This reliance can potentially skew the results, raising questions about whether the identified biases are more reflective of the limitations and biases inherent in Wikidata rather than the RE systems themselves. Unfortunately, at the time of writing there are no alternative open resources with the same coverage of Wikidata.

Another limitation concerns the categorization of relationships. The proposed taxonomy might be too broad in some categories, potentially overlooking more nuanced relation types. For instance, combining 'field' with occupation and sports discipline might obscure specific biases related to distinct professional domains. Additionally, some relation types, like 'relationship,' might be too general. Keeping a more fine-grained taxonomy could help identify specific biases, but as discussed in 3.1 it leads to very infrequent relation types as well as hindering the comparison across datasets.

A final limitation of our work regards gender. Since we rely on Wikidata to augment corpora with socio-demographic information, we must adopt their P21 property that squeezes biological sex, gender identity, and sexual orientation into a single label. Additionally, the representation of people who do not identify as men or women is statistically irrelevant in our RE corpora. Therefore, we were not able to adopt a non-binary perspective on this aspect. While we acknowledge this binary model, it is important to reflect on how it could cause harm by reinforcing gender binaries and excluding non-binary identities. We discuss the term 'gender' and its implications early in the paper, drawing on interdisciplinary perspectives and point to Devinney et al. (2022) for further reading.

Acknowledgements

We thank the NLPnorth group at ITU and the MaiNLP group at LMU for feedback on an earlier version of this paper. Elisa Bassignana is supported by the Independent Research Fund Denmark (Danmarks Frie Forskningsfond; DFF) Sapere Aude grant 9063-00077B and by VILLUM FONDEN grant VIL59826. Pere-Lluís Huguet Cabot is fully funded by the PNRR MUR project PE0000013-FAIR. While working at Babelscape, Pere-Lluís Huguet Cabot was funded by KnowGraphs.

This work was partially supported by the Marie Skłodowska-Curie project *Knowledge Graphs at Scale* (KnowGraphs) No. 860801 under the European Union's Horizon 2020 research and innovation programme.

References

Elisa Bassignana and Barbara Plank. 2022a. CrossRE: A cross-domain dataset for relation extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3592–3604, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Elisa Bassignana and Barbara Plank. 2022b. What do you mean by relation extraction? a survey on datasets and study on scientific relation classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 67–83, Dublin, Ireland. Association for Computational Linguistics.

Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Kimberlé W Crenshaw. 2017. *On intersectionality: Essential writings*. The New Press.

Maria De-Arteaga, Alexey Romanov, Hanna Wallach, Jennifer Chayes, Christian Borgs, Alexandra Chouldechova, Sahin Geyik, Krishnaram Kenthapadi, and Adam Tauman Kalai. 2019. Bias in bios: A case study of semantic representation bias in a high-stakes setting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 120–128, New York, NY, USA. Association for Computing Machinery.

Hannah Devinney, Jenny Björklund, and Henrik Björklund. 2022. Theories of "gender" in NLP bias research. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022*, pages 2083–2102. ACM.

Arif Dirlik. 2007. Global south: Predicament and promise. *The Global South*, 1:12–23.

Andrew Gaut, Tony Sun, Shirlyn Tang, Yuxin Huang, Jing Qian, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2020. Towards understanding gender bias in relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2943–2953, Online. Association for Computational Linguistics.

Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. 2020. RealToxicityPrompts: Evaluating neural toxic degeneration in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3356–3369, Online. Association for Computational Linguistics.

Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pretraining with gradient-disentangled embedding sharing.

Dirk Hovy and Shrimai Prabhumoye. 2021. Five sources of bias in natural language processing. *Language and Linguistics Compass*, 15(8):e12432.

Pere-Lluís Huguet Cabot, Simone Tedeschi, Axel-Cyrille Ngonga Ngomo, and Roberto Navigli. 2023. RED$^{\text{fm}}$: a filtered and multilingual relation extraction dataset. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4326–4343, Toronto, Canada. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR (Poster)*.

Svetlana Kiritchenko and Saif Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 43–53, New Orleans, Louisiana. Association for Computational Linguistics.

Abdullatif Köksal, Omer Yalcin, Ahmet Akbiyik, M. Kilavuz, Anna Korhonen, and Hinrich Schuetze. 2023. Language-agnostic bias detection in language models with bias probing. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12735–12747, Singapore. Association for Computational Linguistics.

Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black, and Yulia Tsvetkov. 2019. Measuring bias in contextualized word representations. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 166–172, Florence, Italy. Association for Computational Linguistics.

Anne Lauscher, Federico Bianchi, Samuel R. Bowman, and Dirk Hovy. 2022. SocioProbe: What, when, and where language models learn about sociodemographics. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7901–7918, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Chandler May, Alex Wang, Shikha Bordia, Samuel R. Bowman, and Rachel Rudinger. 2019. On measuring social biases in sentence encoders. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 622–628, Minneapolis, Minnesota. Association for Computational Linguistics.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 220–229, New York, NY, USA. Association for Computing Machinery.

Moin Nadeem, Anna Bethke, and Siva Reddy. 2021. StereoSet: Measuring stereotypical bias in pretrained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5356–5371, Online. Association for Computational Linguistics.

Riccardo Orlando, Pere-Lluís Huguet Cabot, Edoardo Barba, and Roberto Navigli. 2024. Retrieve, read and link: Fast and accurate entity linking and relation extraction on an academic budget. In *Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand. Association for Computational Linguistics.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sebastin Santy, Jenny Liang, Ronan Le Bras, Katharina Reinecke, and Maarten Sap. 2023. NLPositionality: Characterizing design biases of datasets and models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9080–9102, Toronto, Canada. Association for Computational Linguistics.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy. Association for Computational Linguistics.

Marco Antonio Stranisci, Rossana Damiano, Enrico Mensa, Viviana Patti, Daniele Radicioni, and Tommaso Caselli. 2023. WikiBio: a semantic resource for the intersectional analysis of biographical events. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12370–12384, Toronto, Canada. Association for Computational Linguistics.

Jiao Sun and Nanyun Peng. 2021. Men are elected, women are married: Events gender bias on Wikipedia. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 350–360, Online. Association for Computational Linguistics.

Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1630–1640, Florence, Italy. Association for Computational Linguistics.

Harini Suresh and John Guttag. 2021. A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '21, New York, NY, USA. Association for Computing Machinery.

Yi Chern Tan and L. Elisa Celis. 2019. Assessing social and intersectional biases in contextualized word representations. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledge base. *Communications of the ACM*, 57:78–85.

Katy Weathington and Jed R Brubaker. 2023. Queer identities, normative databases: Challenges to capturing queerness on wikidata. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–26.

Gavin Williams. 1980. The brandt report: A critical introduction. *Review of African Political Economy*, 7(19):77–86.

Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 506–514, Melbourne, Australia. Association for Computational Linguistics.

Charles Chuankai Zhang and Loren Terveen. 2021. Quantifying the gap: A case study of wikidata gender disparities. In *Proceedings of the 17th International Symposium on Open Collaboration*, OpenSym '21, New York, NY, USA. Association for Computing Machinery.

Wenzheng Zhang, Wenyue Hua, and Karl Stratos. 2022. EntQA: Entity linking as question answering. In *International Conference on Learning Representations*.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2017. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2979–2989, Copenhagen, Denmark. Association for Computational Linguistics.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 15–20, New Orleans, Louisiana. Association for Computational Linguistics.

A Relation Type Mapping

In Table 8 we report the mapping that we apply from the original labels of SRED$^{\text{FM}}$, CrossRE, NYT to our proposed unified taxonomy of relation types.

B Resources

The datasets and models utilized in this paper are governed by the following licenses:

- SRED^{FM} Dataset: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.
- CrossRE Dataset: GNU General Public License v3.0.
- NYT Dataset: Linguistic Data Consortium (LDC) Data Use Agreement.
- ReLiK Model: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.
- EntQA Model: MIT License.

C Hardware

We train every model on a single NVIDIA® RTX 3090 graphic card with 24GB of VRAM. We use the default hyperparameters used in the original paper for ReLiK with Adam (Kingma and Ba, 2015) as optimizer.

	SRED^{FM}			CrossRE	NYT
contributes	cast member author producer creator librettist architect	notable work screenwriter composer lyrics by designed by film editor	director performer discoverer or inventor after a work by executive producer voice actor	artifact origin	
date	date of birth work period (end)	date of death time period	work period (start)		
field	occupation field of work	sport instrument	field of this occupation sports discipline competed in		/people/person/profession
geographical relation	place of death country league allegiance	place of birth work location educated at place of burial	country of citizenship country for sport residence indigenous to	physical	/people/person/nationality /people/deceased_person/place_of_death /people/person/place_of_birth /people/ethnicity/geographic_distribution /people/person/place_lived
language	native language	writing language	languages spoken, written or signed		
member	part of member of movement record label	genre crew member(s) ethnic group religious order	member of sports team religion or worldview military branch	part-of general-affiliation	/people/person/religion /people/person/ethnicity /people/ethnicity/people /sports/sports_team_location/teams
participated	participant winner significant event	award received candidate conflict	successful candidate nominated for		
position held	position held chairperson head of state owned by employer	founded by military rank director / manager commanded by	position played on team / speciality office held by head of the organization member of political party head of government	role	/business/company_shareholder/major_shareholder_of /business/person/company /business/company/advisors /business/company/major_shareholders /business/company/founders
relationship	spouse parent relative unmarried partner	sibling family influenced by	child partner in business or sport student	social	/people/person/children
topic	characters	depicts	main subject	topic	

Table 8: **Taxonomy Mapping.** Mapping of the original relation types from SRED^{FM}, CrossRE, NYT into the taxonomy of relation types of Table 1.