How's Business Going Worldwide ? A Multilingual Annotated Corpus for Business Relation Extraction

Hadjer Khaldi[1,2], Farah Benamara[1], Camille Pradel[2], Grégoire Sigel[2], Nathalie Aussenac-Gilles[1]

[1]IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France, [2] Geotrend, Toulouse, France
{farah.benamara, nathalie.aussenac-gilles}@irit.fr,
{Hadjer, Camille, Grégoire }@geotrend.fr

Abstract

The business world has changed due to the 21st century economy, where borders have melted and trades became free. Nowadays, competition is no longer only at the local market level but also at the global level. In this context, the World Wide Web has become a major source of information for companies and professionals to keep track of their complex, rapidly changing, and competitive business environment. A lot of effort is nonetheless needed to collect and analyze this information due to information overload problem and the huge number of web pages to process and analyze. In this paper, we propose the BIZREL resource, the first multilingual (French, English, Spanish, and Chinese) dataset for automatic extraction of binary business relations involving organizations from the web. This dataset is used to train several monolingual and cross-lingual deep learning models to detect these relations in texts. Our results are encouraging, demonstrating the effectiveness of such a resource for both research and business communities. In particular, we believe multilingual business relation extraction systems are crucial tools for decision makers to identify links between specific market stakeholders and build business networks which enable to anticipate changes and discover new threats or opportunities. Our work is therefore an important direction toward such tools.

Keywords: Business relations, Multilingual linguistic resources, Relation extraction.

1. Introduction

The 21st century economy has shaped the economic landscape and changed the way market stakeholders interact with each other in the global market where national borders have melted and trades became more open and free, leading therefore to a more competitive business world (Hameed et al., 2021). Indeed, competitors of one firm are no longer limited to firms of the same country or region, rivalry has moved from the local market level to the multinational level (Gorodnichenko et al., 2008), inciting companies and industries to reinforce their capacity of innovation in order to deliver competitive products and services, expand their economic growth, and improve economic performances (Passaris, 2006; Hameed et al., 2021).

In this context, the available business and financial textual information on the internet (e.g., companies announcements, industry research reports, online news articles, and policy statements) has become a major source of information for companies and professionals to keep track of their complex, rapidly changing, and competitive business environment (Sewlal, 2004). For years, English has been considered as the international lingua franca to communicate on the web in different formats (Dor, 2004). However, in order to enhance customers experience and expand the regional and global market presence, multilingual contents have also emerged (Ciarlone et al., 2008).

This huge amount of business and financial textual information generated online in different languages makes their exploitation by market stakeholders a laborious task. Therefore, the availability of systems that automatically extract this information (e.g. named entities, relations, events) from multilingual textual contents between market entities becomes crucial.

Roughly, state of the art on information extraction from financial textual data concerns either event extraction or binary business relation extraction (henceforth BRE) involving two organizations (e.g., *startups*, *companies*, *non-profit organizations*, etc.). Event extraction aims at identifying event triggers and their arguments (which can be companies and firms) (Lefever and Hoste, 2016; Jacobs et al., 2018; Qian et al., 2019; Jacobs and Hoste, 2021; Xingyue et al., 2021; Wang et al., 2021), and has been used in multiple applications, such as stock market prediction (Chen et al., 2019; Usmani and Shamsi, 2021), perceiving market trends (Berns et al., 2021; Han et al., 2018), assisting investors decisions, and risk analysis (Liang et al., 2020; Hogenboom et al., 2015). BRE, on the other hand, aims at discovering either Inner-Organizational (Inner-ORG) relations linking a company and its components (e.g. company-employees, company-CEO) or Inter-Organizational (Inter-ORG) relations involving different companies (e.g. company-customer, company-partner) (Zhao et al., 2010; Zuo et al., 2017). For example, from the sentence (1) below, a BRE system can infer that the company *Inmarsat* is a client of the company *Airbus*. BRE has shown to be crucial to valuate companies (Zuo et al., 2017), analyze complex emerging business ecosystems (Braun et al., 2018), understand industries structures (Yamamoto et al., 2017), extract competitive intelligence (Zhao et al., 2010; Xu et al., 2011), and reduce credit risk for financial institutions by identifying links between customers groups (Yan et al., 2019).

(1)   The <u>Airbus</u> group has signed a contract with <u>Inmarsat</u> for the delivery of three reconfigurable geostationary satellites in orbit.

In this paper, we focus on binary Inter-ORG BRE from online web content, and propose the first multilingual dataset annotated for business relations, as well as a set of experiments to detect those relations relying on various deep learning architectures. Our contributions are as follows:

- A unified characterization for Inter-ORG relations focusing on five relations: INVESTMENT, CO-OPERATION, SALE-PURCHASE, COMPETITION, and LEGAL PROCEEDINGS.

- BIZREL, the first manually annotated multilingual dataset annotated according to this characterization and considering four languages: *French*, *Spanish*, *English*, and *Chinese*. [1]

- A set of deep learning experiments to detect these relations in texts. We first experiment with monolingual configurations then cross-lingual relation extraction. We investigate in particular various language transfer settings ranging from zero-shot to joint transfer, relying on pre-trained multilingual language models. Our results are encouraging, beating several monolingual baselines and demonstrating the effectiveness of such a resource for both research and business communities.

The paper is organized as follows. Section 2 presents state of the art. Section 3 describes our data, the characterization of inter-organizational business relations we propose, and the annotation guidelines. Section 4 presents the pilot study we carried out on our data, the results, as well as an error analysis. We conclude by providing some perspectives for future work.

2.   Related Work

2.1.   Business Relations Extraction

Business relations are marginally present in knowledge bases, such as DBpedia (Auer et al., 2007) where relations like *Subsidiary* and *Ownership_of* can be found (Zuo et al., 2017). Some business relations are nevertheless annotated in generic relation datasets with fairly low frequencies, such as *Employment / Membership / Subsidiary* in the ACE 2004 dataset (Mitchell et al., 2005), and *org:subsidiaries*, *org:shareholders* or *org:parents* in TACRED dataset (around 453 , 144, and 444 instances respectively) (Zhang et al., 2017).

Although domain-specific RE has already been explored (see for example, the biomedical (Bossy et al., 2019; Zhou et al., 2019; Thillaisundaram and Togia, 2019) and food (Wiegand et al., 2012) domains), BRE has received much less attention in the literature. Most

existing works rely on semi-supervised approaches using either dependency tree based patterns (Braun et al., 2018), or lexical patterns based on a list of keywords that are specific to each predefined relation type (Lau and Zhang, 2011; Burdick et al., 2015). These patterns are, however, hard to maintain. Supervised approaches were recently proposed where classifiers are trained on monolingual annotated dataset to predict the relation type that holds between two entities (Yamamoto et al., 2017; Yan et al., 2019; Collovini et al., 2020; De Los Reyes et al., 2021).

Overall, BRE studies share three main limitations: **(1)** They rely on datasets that are either small or not freely available to the research community (Zhao et al., 2010; Yan et al., 2019; Collovini et al., 2020), **(2)** Only two relations are considered, namely *Competition* and *Co-operation* (Yamamoto et al., 2017; Lau and Zhang, 2011), **(3)** All the proposed systems are monolingual targeting either English (Burdick et al., 2015; Zhao et al., 2010; Zuo et al., 2017), German (Braun et al., 2018), Chinese (Yan et al., 2019), or Portuguese (De Los Reyes et al., 2021; Collovini et al., 2020).

Our aim here is to go beyond these limitations by proposing for the first time, as far as we know, BRE from economic and financial multilingual contents relying on a unified characterization composed of five business relations (cf. next section) annotated in four languages: *French*, *English*, *Spanish*, and *Chinese*.

2.2.   Multilinguality in Relation Extraction

Distant supervision (Mintz et al., 2009) has been extensively used to build relation extraction datasets. This method uses knowledge bases as a source of supervision to automatically label relations between entities based on the assumption that each sentence with two specific entities is an expression of the same relation. This method was first exclusively used to generate monolingual training data (Riedel et al., 2010; Nam et al., 2018; Mandya et al., 2019; Norman et al., 2019), then recently to generate data in a multilingual setting (Bhartiya et al., 2021; Köksal and Özgür, 2020). Despite the low cost of this method and the abundant training data obtained with it, the availability of knowledge resources related to the domain of relations is required, and significant errors in labels may occur leading to noisy training data which may hurt models precision (Riedel et al., 2010; Xie et al., 2021).

Other works use machine translation or parallel data in order to generate multilingual training data starting from an annotated dataset in one language (Yanan, 2013; Zou et al., 2018; Faruqui and Kumar, 2015). The quality of the generated data depends on the performances and the availability of external resources and machine translation systems, which is not straightforward for many languages and domains.

Finally, manual data annotation has been used to generate monolingual or multilingual data (Zhang et al., 2017; Mitchell et al., 2005; Hendrickx et al., 2010)

---

[1] Link to BizRel dataset

3697

relying on clear and well-defined annotation guidelines about relation and entity types (Han, 2010; LDC, 2004). To reduce human errors and biases that may occur during annotation and produce a high-quality annotated dataset, the annotation is performed in an iteratively assessed process (Grosman et al., 2020), where inter-annotator agreement is evaluated using standard metrics, such as Cohen's coefficient (Cohen, 1960) or Fleiss's kappa (Fleiss, 1971).

Our goal being to build a high quality multilingual resource, we decided to rely on manual annotation. To ensure the quality of the annotated dataset, annotations are made in small batches which helps improve the annotation guidelines, as explained in the next section.

3. Data and Annotation

3.1. Data Collection

We follow the procedure described in (Khaldi et al., 2021) for English business relations and extend it to French, Spanish, and Chinese. Since we are targeting relations that occur at the sentence level, data collection consists in extracting from the web relevant sentences following a three-step procedure:

1. *Document collection*: We select a list of seed keywords as search queries related to activity domains chosen by business intelligence experts, such as: *autonomous cars*, *3D printing*, etc. A set of keywords has been built for each targeted language[2] and used to query Google and Bing search engines. Only the first top 1,000 web pages are selected[3] for a pre-processing stage to filter out headers, footers, and navigation menus. The remaining textual contents are segmented into sentences.

2. *Named entity recognition*: Each sentence is passed to spaCy[4] and StanfordNLP[5] to perform named entity recognition.

3. *Sentence selection*: To increase accuracy, the final sentences are selected according to three main criteria: (i) they must contain entities of type *Organization* (henceforth ORG) that have been recognized by the two taggers; (ii) they must contain at least two named entities of type ORG ; and (iii) sentences whose words are at least 95% of type ORG are discarded (this mainly concerns enumerations of organizations).

This procedure resulted in a total of $25,469$ sentences for French, English, Spanish, and Chinese.

3.2. Characterizing Business Relations

First, we define a characterization of Inter-ORG business relations according to which the dataset will be annotated. We start from a set of four relation types initially proposed by (Zhao et al., 2010): INVESTMENT, COOPERATION, SALE, and SUPPLY. Then, we combine the last two relations into SALE-PURCHASE, since we target non-oriented relations, i.e., $R(EO_1, EO_2) = R(EO_2, EO_1)$, EO_i being named entities of type ORG. Inspired by (Lau and Zhang, 2011; Yamamoto et al., 2017), we add COMPETITION and LEGAL PROCEEDINGS. Finally, the relation OTHERS accounts for the absence of a business relation between two ORG, referring to any other relation type between them.

Our relations are defined below, along with examples taken from our dataset (French, Spanish, and Chinese instances are provided together with their English translations). In each example, the two entities involved in the relation are underlined.

INVESTMENT: an EO is a subsidiary of another EO, or EO holds (all or part) of the shares of another EO.

(2)　据路透中文网23日报道，诺基亚 表示，计划对旗下法国分支阿尔卡特- 朗讯 裁撤1233个岗位，相当于该部门总员工数的三分之一。

(According to a Reuters Chinese website on the 23rd, Nokia stated that it plans to abolish 1,233 positions in its French branch Alcatel-Lucent, which is equivalent to one-third of the department's total employees.)

COMPETITION: a competition/rivalry between two EOs providing the same goods or services, or wanting to access the same relatively small market.

(3)　Boeing et l'avionneur brésilien Embraer, rival de Bombardier sur les avions régionaux, ont annoncé discuter sur un éventuel rapprochement de leurs activités.
(Boeing and the brazilian aircraft manufacturer Embraer, Bombardier's regional aircraft rival, have announced discussions on a possible merger of their activities.)

COOPERATION: a contractual cooperation between two EOs, or when two EOs work together on the same project.

(4)　Xiaomi y Nokia firman acuerdos de cooperación comercial.
(Xiaomi and Nokia sign commercial cooperation agreements.)

LEGAL PROCEEDINGS: one EO launches a legal proceedings against another EO.

(5)　Defendants' fraud was alleged to be contained in affidavits and statements made during the pendency of litigation between Lubrizol and Exxon in New Jersey federal district court.

SALE-PURCHASE: one EO is a client of another, or supplies it with goods or services.

(6)   Even more than Volusion, Squarespace offers a cheaper e-commerce solution to Shopify.

OTHERS: If none of the previously described relations are expressed between the tagged entity pair, or if other types of relations out of this list are expressed, the relation should be OTHERS. In (7), there is no business relation expressed between the two underlined EOs.

(7)   While Airbus partners with Audi , Boeing is cozying to Adient, Mercedes- Benz, and even General Motors.

3.3. Manual Annotation

The collected sentences were manually annotated by non-expert native speakers via the collaborative annotation platform *Isahit*[6]. Given a sentence S, and a set of entity pairs composed of non overlapping entities $\{(EO_1, EO_2), EO_i \in S\}$, annotation consists in assigning one relation R per entity pair among five business relations and one negative relation (cited earlier). It is important to note that many relation types can hold between a given entity pair in real world. In this case, we follow ACE annotation principles (Doddington et al., 2004) and ask annotators to only consider explicit mentions of relations in the current sentence without any additional external knowledge. For example in (8), although the underlined EO can be linked by COMPETITION (it is well known that they share the same automotive market), the final annotation is OTHERS because there are no linguistic signals for COMPETITION.

(8)   Present in the city of Wuhan, PSA, Renault and even Valeo had to close their sites in the containment zone while awaiting the green light from the Chinese authorities to resume their activities.

We first start by annotating *English* and *French* datasets, then *Spanish* and *Chinese*. The annotation was made in batches, each containing $2k$ instances. For each batch from *English* or *French* data, 10% of the annotated data is re-annotated by experts. This helped to assess the quality of the annotations and improve annotation guidelines. Over 1k of re-annotated instances, the average Cohen Kappa between the annotators and the experts is 0.766 for English data and 0.685 for French data, which are strong agreements given the complexity of the task. We, therefore, use the same annotation procedure and guidelines for *Spanish* and *Chinese* data.

3.4. Quantitative Results

Table 1 shows the total number of annotated relations as well as the distribution of instances in the train

---

[6]https://isahit.com/en/

(85%) and test sets (15%). From the table, we can observe that our dataset is imbalanced and that OTHERS is dominant for all languages (66% for English, 68% for French, 56% for Spanish, and 46% for Chinese). The distribution of business relations is similar across languages, the most frequent ones being COMPETITION, followed by COOPERATION, then INVESTMENT. We finally observe that SALE-PURCHASE and LEGAL-PROCEEDINGS are under-represented for all languages.

	EN	FR	ES	ZH
Avg. w_per_s	39	41	34	50
Avg. v_per_s	2	2	2	5
Avg. e_per_s	6	5	6	7
RatioU. e_pairs	77%	53%	42%	52%

Table 2: BIZREL Dataset complexity.

To measure the complexity of business relations and their syntactic richness, we compute the average count of words, verbs, and entities per relation (*Avg. w_per_s*, *Avg. v_per_s*, and *Avg. e_per_s* respectively), and the ratio of unique entity-pairs in the dataset (*RatioU. e_pairs*). Table 2 shows the results. Sentences in our dataset contain on average from 5 to 7 *EOs*, therefore, potentially a maximum of 10 to 21 relations could occur in a single sentence between different *EOs* pairs. In addition, sentences are complex containing in average 2 verbs and the context surrounding a given relation instance is 39 tokens on average for English data (41, 34, 50 for French, Spanish, and Chinese respectively). Moreover, 77% of *EOs* pairs in English data (53%, 42%, 52% in French, Spanish, and Chinese respectively) are unique reflecting entity pairs disparity in the dataset. Overall, these measures confirm the diversity and complexity of business relations expressed in our dataset. This is more salient for the Chinese language where the average number of verbs per sentence is the most important.

4.   Business Relation Extraction

We detail here the experiments we carried out on our multilingual dataset BIZREL. We first start by presenting the monolingual (Section 4.1) and cross-lingual experimental settings (Section 4.2), then give our results. We end this section with an error analysis showing main causes of misclassification.

4.1. Monolingual Experiments

We rely on monolingual pre-trained language models for each language which are pretrained on large non-annotated data using a multi-layer bidirectional Transformer encoder (Vaswani et al., 2017) that uses a multi-head self-attention mechanism to model dependencies between tokens regardless of their distances. We use English and Chinese BERT (Devlin et al., 2019) for English and Chinese data, FlauBERT (Le et al., 2020)

	Dataset	Inv.	Com.	Coo.	Leg.	Sal.	Oth.	#Total_split	#Total
EN	Train	281	1,675	627	50	248	5,647	8,528	10,034
	Test	50	296	111	8	44	997	1,506	
FR	Train	268	1,492	726	50	228	5,764	8,528	10,033
	Test	47	263	128	9	40	1,018	1,505	
ES	Train	53	907	84	69	46	1,463	2,000	3,085
	Test	9	160	15	12	8	259	1,085	
ZH	Train	73	619	333	7	22	914	1,968	2,316
	Test	13	110	59	1	4	161	348	
All.	**#Total**	794	5,522	2,083	206	640	1,5224	25,469	25,469

Table 1: BIZREL dataset distribution per relation type and per dataset type (train, test).

Hyperparmeter	Value
train_batch_size	64
	(16 for monolingual FR)
test_batch_size	64
num_epochs	5
max_seq_length	400
learning_rate	5e-5
adam_epsilon	1e-6
warmup_ratio	1e-1

Table 3: Hyperparameters values in the monolingual experiments.

for French, and Beto (Cañete et al., 2020) for Spanish. All the models use 12 layers of 768 dimensions and 12 heads of attention. Each model is fine-tuned on language specific train/test datasets using the hyperparameters in Table 3. We refer to this setting as (S_0) and these models are considered as strong baselines.

4.2. Cross-lingual Experiments

We conduct a set of experiments using the multilingual pre-trained language model mBERT, [7] a variant of BERT. mBERT is composed of 12 layers of 768 dimensions and 12 heads of attention. It is pre-trained on the concatenation of monolingual Wikipedia corpora from 104 languages. Despite being pre-trained without an explicit objective for multilingual sentence representation, mBERT is able to perform cross-lingual transfer on downstream tasks while fine-tuned on an annotated data of a source language with none or few annotated data in target languages. Here, we fine-tune mBERT on our BIZREL multilingual dataset and consider different settings in order to evaluate the model ability to perform cross-lingual business relation extraction.

Let $L \in \{EN, FR, ES, ZH\}$ be the set the four languages in BIZREL. Let $T = \bigcup_i t_i$ be the dataset composed of training instances t_i from one or several source languages, $i \in L$, and let $E = \bigcup_j e_j$, the dataset composed of test instances e_j from one target language $j \in L$. We propose four experimental settings, each

one involves training and testing mBERT model on different subsets of T and E, as follows:

– (S_1) **Transfer between all-language-pairs.** The model is trained on one language and tested on another, i.e., $T = \{t_i\}$, and $E = \{e_j\}$. Note that when $i = j$, this setting is similar to (S_0) but relies on multilingual contextual embeddings instead of monolingual ones. This setting aims to evaluate the cross-lingual transfer between pairs of languages.

– (S_2) **Zero-shot transfer.** Train on all languages except a given target language and test on that target, i.e., $T = \bigcup_i t_i$ and $E = \{e_j\}$ with $i \neq j$. This allows to evaluate the generalization power across-languages when training data is missing for a specific language. In addition, in order to measure the impact of the unseen target language during training on the overall performances of already seen languages, we further test our models on other source languages. Hence, $zero\_EN$, stands for $T = \{t_{FR}, t_{ES}, t_{ZH}\}$ and $E = \{e_{EN}\}$, in addition we evaluate the performances by testing on $E = \{e_{FR}\}$, $E = \{e_{ES}\}$, and $E = \{e_{ZH}\}$.

– (S_3) **Richly-labeled transfer.** The distribution of relations across languages in BIZREL is imbalanced, with a higher frequency of French and English instances. To evaluate the impact of size on the cross-lingual experiments, we split the dataset into richly labeled (French and English) vs. poorly labeled languages (Chinese and Spanish) and either : (i) Train on $T_{richL} = \{t_{EN}, t_{FR}\}$ then evaluate on $E = \{e_j\}$ with $j \in \{ES, ZH\}$, or (ii) Train on $T_{poorL} = \{t_{ES}, t_{ZH}\}$ then evaluate on $E = \{e_k\}$ with $k \in \{FR, EN\}$.

– (S_4) **All-joint transfer.** In this last stetting, the model is trained on all the languages at the same time, and tested on one target language already seen during training, i.e., $T_{all} = \{t_{EN}, t_{FR}, t_{ES}, t_{ZH}\}$, and $E = \{e_j\}$, $j \in L$.

4.3. Results

Results of the monolingual and cross-lingual experiments are reported in Table 4, in terms of macro precision, recall, and F-score.

Overall, we can observe that models trained on multilingual data outperform their monolingual counterparts for all the languages, except for ZH where the Chinese BERT achieves the best with an F-score of 74.3%.

Settings	Models	EN P	EN R	EN F	FR P	FR R	FR F	ES P	ES R	ES F	ZH P	ZH R	ZH F
S_0‡	Monolg.	67.7	71.9	69.5	72.2	66.8	69.0	74.4	72.5	73.1	**75.8**	**73.2**	**74.3**
S_1	EN	66.8	**72.4**	69.1	67.8	51.9	57.3	72.2	57.3	62.3	41.6	32.4	34.8
	FR	62.6	57.5	59.6	69.0	63.4	65.8	**78.3**	67.0	70.3	39.3	30.9	31.4
	ES	54.1	57.5	54.2	58.8	51.1	53.6	77.1	**76.8**	**76.8**	39.0	43.3	38.4
	ZH	49.6	32.3	35.5	50.7	29.4	32.7	54.5	36.4	40.7	62.9	72.2	66.0
S_2	$zero\_EN$	60.9	63.6	61.3	72.1	65.6	68.0	83.3	**86.3**	**84.6**	72.7	59.2	62.2
	$zero\_FR$	66.3	70.2	67.8	68.3	55.6	60.1	<u>83.6</u>	79.1	81.0	60.0	60.6	60.3
	$zero\_ES$	65.1	69.7	67.1	**73.1**	65.2	68.3	79.6	71.0	73.9	60.4	60.2	60.3
	$zero\_ZH$	66.9	70.8	68.3	<u>74.4</u>	67.0	69.8	80.5	77.7	78.7	60.3	52.0	54.2
S_3	$richL$	66.5	<u>75.0</u>	<u>70.2</u>	71.9	**67.5**	69.4	77.8	66.5	71.1	42.5	36.9	38.4
	$poorL$	58.6	56.7	56.9	61.5	50.7	54.8	75.5	73.8	73.2	53.7	54.3	54.0
S_4	all	<u>67.8</u>	72.9	69.9	<u>74.4</u>	<u>68.8</u>	<u>70.8</u>	79.3	80.4	79.7	73.8	64.1	65.1

Table 4: Monolingual and cross-lingual models results per language. Best performing models in each (S_i) setting are in bold while the best model for each language is underlined. ‡: Baselines models.

Compared to (S_0), transfer between all-language-pairs (i.e., the (S_1) setting) using multilingual embeddings was less productive, except for ES where all the scores increased (e.g., +3.7% F1). As expected, this decrease is however less important when the test concerns the same language. For example, -3.2% F1 when $T = \{t_{FR}\}$ and $E = \{e_{FR}\}$, while -15.4% F1 when $T = \{t_{FR}\}$ and $E = \{e_{ES}\}$. We can also conclude that language transfer from EN, FR, or ES to ZH is very poor (F1 < 50%) while transfer to ES is feasible.

Regarding (S_2), the zero-shot transfer configuration, we note that excluding a target language from the training set was not conclusive, except for ES, where $zero\_ES$ is able to outperform monolingual ES (+0.8% F1). Similarly, excluding ZH, helps to boost performances of the model when evaluated on EN, or FR, while excluding EN or FR yield better results on ES.

Training m-BERT on *richly-labeled* data boosted the results when tested on those data (see for example +0.7% when $E = \{e_{EN}\}$ and +0.4% when $E = \{e_{FR}\}$. However, the results were lower when compared to the baselines (e.g., -2% F1 for ES). On the other hand, training on *poorly-labeled* data has weak transfer power compared to richly labeled data.

Finally, *all-joint transfer* that combines all languages during training was the best, beating all monolingual baselines. This is more salient for FR where we achieve the highest F-score of 70.8%. One reason behind that could be that one relation can be expressed using similar syntactic patterns across languages, which can augment artificially relation instances for one language. Here again, the results when testing on the Chinese test set were not conclusive. This is probably due to the difference in script writing between ZH and the other languages: EN, FR, and ES. Thus, including these languages during training won't improve results on ZH. Also, one possible explanation to the very good results obtained on the other languages when in-

cluding ZH during training, may come from the named entities that are often written in English in our Chinese dataset.

A closer look into the results per class for monolingual models and best performing multilingual models per language (cf. Table 5)[8] shows that, in general, the relation types with the best F-score, for all languages, are the ones with more training data (COMPETITION, COOPERATION). LEGAL PROCEEDINGS has high F-scores, which can be due to the similarity and little variations of relation instance patterns because of the few examples we have. Conversely, under-represented relation types (INVESTMENT, LEGAL PROCEEDINGS, SALE-PURCHASE) gained an improvement over baseline models for many languages when training on more than one language.

	Inv.	Com.	Coo.	Leg.	Sal.	Oth.
$\mathbf{EN}_m$	66.0	78.5	**67.2**	**77.8**	40.5	**86.7**
$\mathbf{EN}_b$	**67.2**	**78.7**	63.9	**77.8**	**47.2**	86.3
$\mathbf{FR}_m$	53.9	**72.4**	68.2	**80.0**	52.5	**87.3**
$\mathbf{FR}_b$	**65.2**	71.7	67.7	76.9	**56.8**	86.6
$\mathbf{ES}_m$	50.0	**86.5**	86.7	95.7	30.8	**88.7**
$\mathbf{ES}_b$	**80.0**	84.0	**93.3**	**100**	**62.5**	88.0
$\mathbf{ZH}_m$	69.6	94.5	89.8	100	0.0	91.8
$\mathbf{ZH}_b$	-	-	-	-	-	-

Table 5: Monolingual ($_m$) and best multilingual models ($_b$) F1-score per relation type and per language. Best results of each language are in bold.

4.4. Error Analysis

We performed a detailed error analysis on the best performing models for each language (cf. Table 4) in order to gain insights into the main shortcomings of the current approach. We can notice the following main sources of errors.

---
[8]In this table, the line ZH$_b$ is empty since the monolingual model was the best.

3701

– One sentence-many relations. This concerns sentences containing more than one relation between different entity pairs, as in (9) and (10). In these examples, only the relation linking the two EO underlined has to be identified. Our best model predicts COOPERATION (EO_2,EO_3) in (9) and (10), whereas the ground-truth annotation is INVESTMENT(EO_2,EO_3) in (9) and OTHERS(EO_2,EO_3) in (10). Note that a COOPERATION relation actually exists between EO_1 and EO_2 in (9), and between EO_1 and EO_3 in (10).

(9)   [Microtronic]$_1$ présentera ses solutions de paiement sans contact, en partenariat avec [Swisscom]$_2$ (groupe [Vodafone]$_3$), une solution de porte-monnaie électronique hébergée sur un téléphone portable et utilisant la technologie NFC pour communiquer.

([Microtronic]$_1$ will present its contactless payment solutions, in partnership with [Swisscom]$_2$ (group [Vodafone]$_3$), a door-to-door solution electronic money hosted on a mobile phone and using NFC technology to communicate.

(10)   2017 年3 月，[Mobileye]$_1$ 被[英特尔]$_2$ 在2017年以153 亿美元收购，此前这家以色列视觉公司也是[特斯拉]$_3$ 的合作伙伴，正是双方联手才有了初代的Autopilot。

(In March 2017, [Mobileye]$_1$ was acquired by [Intel]$_2$ in 2017 for 15.3 billion U.S. dollars. This Israeli vision company was also partner of [Tesla]$_3$, to have the first generation of Autopilot.)

– Use of generic lexical clues. In (11), the lexical clue "*de*" (of) is generally used to express the relation type *Investment* referring to a subsidiary link between two organizations in French language. However, in this example, it does not. Our model misclassifies this sentence as INVESTMENT(EO_1,EO_2) whereas the ground-truth annotation is OTHERS (EO_1,EO_2). Moreover, the clue "*por detrás de*" (behind of) is used in (12) to express a comparison between EO_1 and EO_2 about sponsoring Fifa, whereas it can be used to express the business relation COMPETITION in other contexts. This sentence is, therefore, misclassified as COMPETITION(EO_1,EO_2) whereas the ground-truth annotation is OTHERS (EO_1,EO_2).

(11)   Si [Google]$_1$ est sorti de [Stanford]$_2$, il y a aussi des startups françaises connues qui sont nées au sein d'incubateurs des écoles.

(If [Google]$_1$ came out of [Stanford]$_2$, there are also well-known French startups that were born within school incubators)

(12)   [Hyundai]$_1$ es el tercer patrocinador más antiguo de la Fifa, por detrás de Coca-Cola y [Adidas]$_2$.

([Hyundai]$_1$ is the third oldest sponsor of Fifa, behind Coca-Cola and [Adidas]$_2$.)

– Indirectly expressed relations. In (13), the expression "*has issued Autonomous Vehicle Testing Permits*" triggers a COMPETITION relation between EO_1 and EO_2. However, the model predicts OTHERS.

(13)   Wheego and Valeo now join the likes of Google, Tesla, GM Cruise and Ford on the list of companies the Californian DMV has issued Autonomous Vehicle Testing Permits to, as well as [Volkswagen]$_1$, Mercedes Benz, [Delphi Automotive]$_2$ and Bosch.

5.   Conclusion

In this paper, we presented the first multilingual corpus annotated for business relation extraction. It is composed of about $25,469$ sentences in four languages (*French*, *Spanish*, *English*, and *Chinese*), annotated according to a unified characterization for *Inter-Organizational* relations composed of five important relations: INVESTMENT, COOPERATION, SALE, SUPPLY, COMPETITION, and LEGAL PROCEEDINGS.
We experimented multilingual relation extraction with monolingual models then with various cross-lingual transfer settings ranging from zero-shot to joint transfer. The best results are obtained with m-BERT trained on all-joint datasets. For future work, we plan to add features to our BERT model in order to account for business relations specificities and improve classification based on our error analysis. We believe that this work can advance both research and business communities.

6.   Aknowledgement

7.   Bibliographical References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In The semantic web, pages 722–735. Springer.

Berns, J., Bick, P., Flugum, R., and Houston, R. (2021). Do changes in md&a section tone predict investment behavior? Financial Review.

Bhartiya, A., Badola, K., et al. (2021). Dis-rex: A multilingual dataset for distantly supervised relation extraction. arXiv preprint arXiv:2104.08655.

Bossy, R., Deléger, L., Chaix, E., Ba, M., and Nédellec, C. (2019). Bacteria biotope at bionlp open shared tasks 2019. In Proceedings of The 5th Workshop on BioNLP Open Shared Tasks, pages 121–131.

Braun, D., Faber, A., Hernandez-Mendez, A., and Matthes, F. (2018). Automatic relation extraction

for building smart city ecosystems using dependency parsing. In Pierpaolo Basile, et al., editors, Proceedings of the 2nd Workshop on Natural Language for Artificial Intelligence (NL4AI 2018) co-located with 17th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2018), Trento, Italy, November 22nd to 23rd, 2018, volume 2244 of CEUR Workshop Proceedings, pages 29–39. CEUR-WS.org.

Burdick, D., Hernández, M., Ho, H., Koutrika, G., Krishnamurthy, R., Popa, L. C., Stanoi, I., Vaithyanathan, S., and Das, S. R. (2015). Extracting, linking and integrating data from public sources: A financial case study. Available at SSRN 2666384.

Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In PML4DC at ICLR 2020.

Chen, D., Zou, Y., Harimoto, K., Bao, R., Ren, X., and Sun, X. (2019). Incorporating fine-grained events in stock movement prediction. In Proceedings of the Second Workshop on Economics and Natural Language Processing, pages 31–40.

Ciarlone, L., Kadie, K., and Laplante, M. (2008). Multilingual communications as a business imperative: why organizations need to optimize the global content value chain. The Gilbane Group.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1):37–46.

Collovini, S., Gonçalves, P. N., Cavalheiro, G., Santos, J., and Vieira, R. (2020). Relation extraction for competitive intelligence. In International Conference on Computational Processing of the Portuguese Language, pages 249–258. Springer.

De Los Reyes, D., Barcelos, A., Vieira, R., and Manssour, I. (2021). Related named entities classification in the economic-financial context. In Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation, pages 8–15.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S. M., and Weischedel, R. M. (2004). The automatic content extraction (ace) program-tasks, data, and evaluation. In LREC, volume 2, pages 837–840.

Dor, D. (2004). From englishization to imposed multilingualism: Globalization, the internet, and the polit-

ical economy of the linguistic code. Public Culture, 16(1):97–118.

Faruqui, M. and Kumar, S. (2015). Multilingual open relation extraction using cross-lingual projection. arXiv preprint arXiv:1503.06450.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5):378.

Gorodnichenko, Y., Svejnar, J., and Terrell, K. (2008). Globalization and innovation in emerging markets. Technical report, National Bureau of Economic Research.

Grosman, J. S., Furtado, P. H., Rodrigues, A. M., Schardong, G. G., Barbosa, S. D., and Lopes, H. C. (2020). Eras: Improving the quality control in the annotation process for natural language processing tasks. Information Systems, 93:101553.

Hameed, K., Arshed, N., Yazdani, N., and Munir, M. (2021). On globalization and business competitiveness: A panel data country classification. Estudios De Economia Aplicada, 39(2):1–27.

Han, S., Hao, X., and Huang, H. (2018). An event-extraction approach for business analysis from online chinese news. Electronic Commerce Research and Applications, 28:244–260.

Han, D. (2010). Klue annotation guidelines-version 2.0. Technical report, Technical Report RC25042, IBM Research, August.

Hogenboom, F., de Winter, M., Frasincar, F., and Kaymak, U. (2015). A news event-driven approach for the historical value at risk method. Expert Systems with Applications, 42(10):4667–4675.

Jacobs, G. and Hoste, V. (2021). Sentivent: enabling supervised information extraction of company-specific events in economic and financial news. LANGUAGE RESOURCES AND EVALUATION.

Jacobs, G., Lefever, E., and Hoste, V. (2018). Economic event detection in company-specific news text. In Proceedings of the First Workshop on Economics and Natural Language Processing, pages 1–10.

Khaldi, H., Benamara, F., Abdaoui, A., Aussenac-Gilles, N., and Kang, E. (2021). Multilevel entity-informed business relation extraction. In International Conference on Applications of Natural Language to Information Systems, pages 105–118. Springer.

Köksal, A. and Özgür, A. (2020). The relx dataset and matching the multilingual blanks for cross-lingual relation classification. arXiv preprint arXiv:2010.09381.

Lau, R. and Zhang, W. (2011). Semi-supervised statistical inference for business entities extraction and business relations discovery. In SIGIR 2011 workshop, Beijing, China, pages 41–46.

LDC. (2004). Annotation guidelines for relation detection and characterization (rdc). Technical report, Linguistic Data Consortium.

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbe, B., Besacier, L., and Schwab, D. (2020). Flaubert: Unsupervised language model pre-training for french. In LREC.

Lefever, E. and Hoste, V. (2016). A classification-based approach to economic event detection in Dutch news text. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 330–335, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Liang, X., Cheng, D., Yang, F., Luo, Y., Qian, W., and Zhou, A. (2020). F-hmtc: Detecting financial events for investment decisions based on neural hierarchical multi-label text classification. In IJCAI, pages 4490–4496.

Mandya, A., Bollegala, D., Coenen, F., and Atkinson, K. (2019). A dataset for inter-sentence relation extraction using distant supervision. In LREC 2018-11th International Conference on Language Resources and Evaluation, pages 1559–1565.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pages 1003–1011.

Nam, S., Han, K., Kim, E.-k., and Choi, K.-S. (2018). Distant supervision for relation extraction with multi-sense word embedding. In Proceedings of the 9th Global Wordnet Conference, pages 239–244.

Norman, C., Leeflang, M., Spijker, R., Kanoulas, E., and Névéol, A. (2019). A distantly supervised dataset for automated data extraction from diagnostic studies. In Proceedings of the 18th BioNLP Workshop and Shared Task, pages 105–114.

Passaris, C. (2006). The business of globalization and the globalization of business. Journal of Comparative International Management, 9(1):3–18.

Qian, Y., Deng, X., Ye, Q., Ma, B., and Yuan, H. (2019). On detecting business event from the headlines and leads of massive online news articles. Information Processing & Management, 56(6):102086.

Riedel, S., Yao, L., and McCallum, A. (2010). Modeling relations and their mentions without labeled text. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 148–163. Springer.

Sewlal, R. (2004). Effectiveness of the web as a competitive intelligence tool. South African Journal of Information Management, 6(1).

Thillaisundaram, A. and Togia, T. (2019). Biomedical relation extraction with pre-trained language representations and minimal task-specific architecture. In Proceedings of The 5th Workshop on BioNLP Open Shared Tasks, pages 84–89. ACL.

Usmani, S. and Shamsi, J. A. (2021). News sensitive stock market prediction: literature review and suggestions. PeerJ Computer Science, 7:e490.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems, pages 5998–6008.

Wang, H., Zhu, T., Wang, M., Zhang, G., and Chen, W. (2021). A prior information enhanced extraction framework for document-level financial event extraction. Data Intelligence, pages 1–12.

Wiegand, M., Roth, B., Lasarcyk, E., Köser, S., and Klakow, D. (2012). A gold standard for relation extraction in the food domain. In Proceedings of LREC.

Xie, C., Liang, J., Liu, J., Huang, C., Huang, W., and Xiao, Y. (2021). Revisiting the negative data of distantly supervised relation extraction. arXiv preprint arXiv:2105.10158.

Xingyue, C., Liping, N., and Zhiwei, N. (2021). Financial event extraction based on electra and part-of-speech. Data Analysis and Knowledge Discovery, page 1.

Xu, K., Liao, S. S., Li, J., and Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. Decision Support Systems, 50(4):743–754. Enterprise Risk and Security Management: Data, Text and Web Mining.

Yamamoto, A., Miyamura, Y., Nakata, K., and Okamoto, M. (2017). Company relation extraction from web news articles for analyzing industry structure. In 2017 IEEE 11th International Conference on Semantic Computing (ICSC), pages 89–92. IEEE.

Yan, C., Fu, X., Wu, W., Lu, S., and Wu, J. (2019). Neural network based relation extraction of enterprises in credit risk management. In 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), pages 1–6. IEEE.

Yanan, H. (2013). Cross lingual relation extraction via machine translation. Journal of Chinese Information Processing, 27(5):191–197.

Zhao, J., Jin, P., and Liu, Y. (2010). Business relations in the web: Semantics and a case study. Journal of Software, 5(8):826–833.

Zhou, H., Lang, C., Liu, Z., Ning, S., Lin, Y., and Du, L. (2019). Knowledge-guided convolutional networks for chemical-disease relation extraction. BMC bioinformatics, 20(1):260.

Zou, B., Xu, Z., Hong, Y., and Zhou, G. (2018). Adversarial feature adaptation for cross-lingual relation classification. In Proceedings of the 27th International Conference on Computational Linguistics, pages 437–448.

Zuo, Z., Loster, M., Krestel, R., and Naumann, F. (2017). Uncovering business relationships: Context-sensitive relationship extraction for difficult relationship types. In Michael Leyer, editor, Lernen,

Wissen, Daten, Analysen (LWDA) Conference Proceedings, Rostock, Germany, September 11-13, 2017, volume 1917 of CEUR Workshop Proceedings, page 271. CEUR-WS.org.

8. Language Resource References

Hendrickx, Iris and Kim, Su Nam and Kozareva, Zornitsa and Nakov, Preslav and Séaghdha, Diarmuid Ó. and Padó, Sebastian and Pennacchiotti, Marco and Romano, Lorenza and Szpakowicz, Stan. (2010). SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. Association for Computational Linguistics, SemEval '10.

Mitchell, Alexis and Strassel, Stephanie and Huang, Shudong and Zakhary, Ramez. (2005). Ace 2004 multilingual training corpus.

Zhang, Yuhao and Zhong, Victor and Chen, Danqi and Angeli, Gabor and Manning, Christopher D. (2017). Position-aware Attention and Supervised Data Improve Slot Filling. Association for Computational Linguistics.