Applying event detection to reveal the *Estado da Índia*

Gonçalo C. Albuquerque
goncalo.albuquerque@uevora.pt
Renata Vieira
renatav@uevora.pt
Ana Sofia Ribeiro
asvribeiro@uevora.pt
CIDEHUS, Universidade de Évora, Portugal

Marlo Souza
msouza1@ufba.br
Institute of Computing, UFBA, Brazil

Abstract

This paper presents a study based on the application of a Portuguese event detection (extraction and classification) tool (TEFE) to historical texts. It shows how historical analysis and interpretation can use this tool in historical research, on the basis of a historiography analysis of the Portuguese Empire in East. TEFE has been applied to one volume of the Portuguese *Livros das Monções* (Monsoon Books), concerning the time gap between 1614 and 1616, highlighting conflict-related categories of events. A historical analysis of this special category of events is performed, revealing aspects of a generalized juncture of war and conflict in Asia in the early seventeenth century.

1 Introduction

This paper intends to investigate a digital methodology employing Natural Language Processing to study, in a fully-integrated way, the period in which Portugal was part of the Hispanic Monarchy (1580-1640), that is, analysing the Portuguese presence in all the Indian ocean in relation with local populations and political units and with other European powers which were directly competing with the Portuguese for the control of Indian ocean and the European access to luxury Asian commodities, such as spices, cloths, silk or porcelain, particularly.

The *Estado da Índia* constituted the most complex overseas Portuguese set of territories, encompassing a geography as wide as the Indian Ocean borders, from the Eastern coast of Africa to Macao or Japan (Pearson, 1987; Subrahmanyan, 1993; Thomaz, 1998). These territories included different types of jurisdictional realities: conquered territories as Goa or the Northern province of Daman and Diu or factories and fortresses under Portuguese administration within local political units, such as Kochi, implemented after negotiations with local governments. This administrative and jurisdictional diversity increased the difficulties of the Portuguese administration, centralized in Goa under the authority of a vice-roy, to rule this set of geographical discontinuous territories. At the same time, the literature underlines that this political unit should also be perceived within the geographical scope of Portuguese informal presence in Asia, that is, the places where free-riding Portuguese individuals constituted significant communities (mostly *mestizos*), albeit there were nor Portuguese formal structures of any sort, nor under Portuguese jurisdictional power (Antunes, 2012).This was particularly the case of the Portuguese communities settled around the Bay of Bengal.

We explore the application of an event extraction method aiming to support the process of data identification of historical junctures, which is a huge time-consuming task for historians dealing with massive document corpora. A process like this, applied to the study of a colonial macro-region as the eastern Portuguese Empire, identifies and categorize human actions and episodes which determine not only patterns of historical junctures in time and place, but also disruptive events that underline changing processes. In this work, we focus on selecting particular categories of events, those related to war and conflict, and try to evaluate how they reflect a particular historical juncture and how they allow a more holistic comprehension of the Eastern sphere of the Portuguese overseas empire.

We start by discussing related work on event detection (ED) in Section 2, and the novelty of using this approach to historical research. Then Section 3 presents and describes the functioning of the computational tool used for event detection and classification in this work. Section 4 describes our studied source, the Monsoon Books. Section 5, then describes the methodology of applying this computational tool to our historical 17th-century text. An historical analysis and interpretation of the results is presented on Section 6. Finally, the paper is concluded in Section 7.

2 Related Work

The literature on Event Detection and its subtasks, namely event identification and event classification, has been mainly focused on English and Chinese languages (Ahn, 2006; Nguyen and Grishman, 2015; Liu et al., 2018; Nguyen and Nguyen, 2019), for which there are standard corpora for the task (Xiang and Wang, 2019). For the Portuguese language, however, few works on such task have been developed.

Event identification (though not classification) was addressed in some extent in the HAREM (Carvalho et al., 2008) evaluation. The first work, to our knowledge, that addresses Event Detection is that of Costa and Branco (2012a), employing decision trees and feature engineering on the TimeBank-PT corpus. Quaresma et al. (2019), on the other hand, investigates the task of event extraction, i.e. identification of events and their arguments, based on Semantic Role Labelling. Finally, the work of Sacramento and Souza (2021) studies event extraction for the Portuguese language with a rich semantic typology of events based on the FrameNet, trained on the TimeBack-PT corpus.

While some work on natural language processing for the digital humanities have been proposed (Piotrowski, 2012), as argued by McGillivray et al. (2020), there is still a great unmet potential for the application of NLP tools for processing large textual datasets for humanities researchers.

Interest on the application of computational methods for historiographical research has recently increased in the literature, particularly the application of geographical information systems and network analysis. For example, Dahmen et al. (2017) applies network analysis to study coalitions and conflicts in the the crisis of 1225 – 1235 within the Holy Roman Empire, the conflict between the Emperor Frederick II and his son, Henry VII, similar to that performed by Gramsch (2014). Prado et al. (2020) also employ computer-based network analysis to study the presentation of women and their political and social roles in sources on the early history of Britain. Also, social network analysis has been particularly used for the study of trade and finance in early modern and contemporary Europe (Ribeiro, 2016).

Event detection to historical sources has been applied to narrative sequential novels in Italian language (Sprugnoli and Tonelli, 2019), or to geographical events in colonial narratives in 16th-century Spanish of New Spain (Jimenez-Badillo et al., 2020) (todays Mexico). Although previous work for Portuguese has considered the detection of entities (Vieira et al., 2021; Cameron et al., 2022), we are not aware of event detection works that consider Portuguese historical sources.

3 Event Detection

The goal of event detection is to identify and classify event mentions in plain text. Given an input text, an ED system should be able to identify whether the sentences contain events of interest by identifying event trigger terms (event identification) and classify them into specific event types (event classification).

Similar to Sacramento and Souza (2021), following the ACE 2005 annotation guidelines (Consortium, 2005), we understand events as things that happen in time, i.e. "a specific occurrence involving participants; [...] something that happens and can frequently be described as a change of state.". In this context, an event is denoted by an event trigger, which may be expressed primarily through verbs and nominalizations but also by other word classes such as adjectives and prepositions.

For instance, in the following sentence from the TimeBankPT corpus (Costa and Branco, 2012b):

> "Meridian National Corp. **said** it **sold** 750,000 shares of its common stock to the McAlpine family interests, for $1 million, or $1.35 a share."

The words "**said**" and "**sold**" describe event occurrences (triggers) for two distinct event mentions, one of type *Statement* and the other of type *Commerce Selling*, respectively, if we consider the FrameNet (Baker et al., 1998) lexicon as a source of target event types.

Sacramento and Souza (2021) developed a method on Portuguese sentences (named TEFE), which employs a rich semantic typology of events based on the FrameNet (Baker et al., 1998). TEFE encodes ED as a sequence labelling problem, in assuming that event triggers are single words/tokens in the sentences. It employs bidirectional recurrent neural networks to simultaneously predict event triggers, their types and arguments.

In this work, we have only considered the simplified model EDFF for Event Detection, described by Sacramento (2021). The method assumes that the token representations, obtained using a pre-trained

BERT model for the Portuguese language (Souza et al., 2020), encode enough information to identify event mentions and their types. The word embeddings, represented by the vector $\vec{x}$, are processed through a time-distributed dense layer, followed by a softmax layer, as described in Equations 1 and 2 below. The model was trained on an enriched TimeBankPT corpus, annotated with events types from the FrameNet project.

$$c = RELU(W_1\vec{x} + b_1) \quad (1)$$

$$O = softmax(W_2c + b_2) \quad (2)$$

In this work, we directly apply TEFE to historical data, composed of previously transcribed texts from the Monsoon Books. We were interested in evaluating its usefulness to the identification of historical junctures in historical corpora, considering that 17th-century Portuguese is lexically, grammarly and semantically different from the current language. Note that no adaptation of the model trained by Sacramento (2021) was made to deal with the differences between contemporary and 17th century Portuguese.

4 The Monsoon Books

The *Documentos Remetidos da Índia* or *Livros das Monções* (Monsoon books) collect letters exchanged between the monarchs and Portuguese government councils and India viceroys, where all types of affairs concerning the so-called Portuguese *Estado da Índia* were discussed. They comprise a geographical scope from Eastern Africa to Japan. The use of this collection is paramount to understand the internal dynamics of the Portuguese Estado da Índia until the 19th century. In fact, they are considered the core documents produced by Portuguese authorities in Asia. The fact of being a type of documental corpora concerning all types of issues makes the Monsoon Books unique and a privileged lab for building a new analytic model and approach to understand internal dynamics of colonial empires macro-regions. As internal dynamics we consider a full scan of political, economic affairs, social organization, cultural and religious interactions both in a diachronic and sincronic perspectives in a cosmopolitan world in continuous change. The Monsoon Books are composed by the sets of documents located in both in the Portuguese National Archives, in Lisbon, and in the Historical Archives of Goa, in Panjin, India. Since this paper intends to assess an automatic event extraction model in order to conceive an interpretative framework of European colonial presence in overseas macro-regions, we employ some of the already transcribed and published books referring to the years of 1614-1616 (Patto, 1893). Presently, both the handwritten and the printed Monsoon Books are in no means indexed, compromising historical research.

5 Applying Event Identification to the Monsoon Letters

Table 1: Events identified by TEFE - Examples

Trigger	Event Type
"saber"	Awareness
"entendido"	Awareness
"partiram"	Departing
"comaçaram"	Activity Start
"fazer"	Intentionally act
"causa"	Causation
"receber"	Receiving
"tire"	Removing
"ver"	Perception experience
"cumprirá"	Activity ongoing

Each volume of Monsoon books encompasses more than 300 printed pages of narrative text. In this sense, applying a computational tool of event extraction enhances historical research by rapidly extracting textual information from a large collection. It is not feasible for an historian, studying a specific theme in a certain time period and specific location, to rapidly locate in this documental collection the letters concerning an specific theme under research. The automatic identification and classification of events in the Monsoon books allows the historian to more efficiently locate the particular passages that are relevant.

Also, the statistics of identified events and its semantic classification is itself an analytical tool for historians, since it makes clear which were the main administrative concerns of Portuguese authorities in Asia, in a certain chronology. Language technology may help the reader with hints, extraction and quantification of these events. The system described by Sacramento and Souza (2021), and discussed in Section 3, was developed with the purpose of finding and classifying mentions to events, as well as identifying the participants of

these events. The system receives an input sentence such as "*Eu el-rey faço saber aos que este alvará virem que tenho entendido que pelo mau comcerto que tiveram as naus que, o anno passado de seiscentos e quinze, partiram do porto de Goa para este reino*[...]" and identifies that "partiram" (departed) is an instance of a Departure event (described by the Departing Frame) and that "*as naus*" (the ships), "*porto de Goa*" (Goa's harbor) and "*este reino*" (this kingdom) are entities participating in such event.

In the extracting process, the text source is first fragmented into sentences, using NLTK (Bird, 2006) Portuguese sentence segmenter. Later the event detection model is applied to each sentence independently. The resulting events identified by the system were then manually analyzed to understand the potential of this method to identify information about conflicts in the region, in a period of intense political change.

We would like to note that an analysis of accuracy of the tool in the studied corpus is out of the scope of this paper. The accuracy of the tool is presented in previous work (Sacramento, 2021), where it was evaluated in a different corpus. Here we make instead an analysis of its usefulness to historical research. While understanding the accuracy of the model when applied to historical texts can be valuable to indicate strategies for adapting theses models to new corpora, to our knowledge, there is still no dataset of historical Portuguese texts annotated with events that could be used for such an evaluation.

6 Analysis of conflict related events

The extraction of events from Volume III of the *Livros das Monções*, between 1614 and 1616, allowed us to identify around 101 different event categories and 4,688 occurrences of events. A total of 18 types of conflict-related events were identified (see table 2). The team instantly realised that such statistics indicate a period of severe stress and open conflict in Asia. However, these events are not related to conflict occurrences in the same way, due to the meaning of the category and its term (trigger), i.e. the word that identifies the event in the context in which it arises. That said, we decided to divide the types of events related to conflict occurrences into two groups: specific categories and generic categories. We understand by specific categories the types of events that directly indicate the occurrence of conflicts, through their meaning.

Example: "e no particular da *guerra (Hostile Encounter)* que o dito rey de cochim tem com o samorim, tereis"

The Hostile Encounter category associated to the term "guerra" has a direct meaning with the event conflict. We know that the presence of this term indicates that a conflict is present in the context of competition between the king of Kochi and the Calicut Samorin. However, the relationship between the events identified and the occurrence of conflicts does not appear in the document in the same way, since the term that triggers the event is not always directly associated with the event in question. As the example below shows:

Example: "**fortificacao** e provimento da cidade , que convem muito que se remedeie , de maneira que movendo o mogor **guerra** , ou pondo - lhe **cerco** (como se deve recear) lhe nao possa *fazer (Intentionally Act)* **damno**"

The category *Intentionally Act*, associated with the trigger "fazer", which identifies the event in this sentence, is not directly related to the occurrence of a conflict. However, the presence of the terms "fortificaçao", "guerra", "cerco" and "damno" indirectly refer to the presence of a conflict. This is due to their relationship with this type of occurrence, since they belong to the conflict lexicon. The case presented here is not unique; throughout the documents we see the presence of terms that are indirectly related to the word "conflito", such as "defesa", "defensão", "imigos", "rebeldes", "hollandezes", "ataques", in addition to those presented above. Because of this indirect relationship, we classify these types of events as generic categories.

Although our main focus is on the specific categories, because of their direct relationship with the occurrence of conflicts, we can not leave out the generic categories because of their importance. If we count the total categories where events associated with conflicts were found, we see that 66% are represented by generic categories (12 out of 18) and of the 233 events linked to conflict occurrences, 56% (131 events) refer to these categories, reinforcing the importance of introducing them into this analysis. Tables 3 and 4 show the percentage of verified conflict occurrences in each of the categories, specific and generic, respectivelly.

As we can see from Tables 5 and 6 most of the triggers, i.e. the terms that trigger the events, are verbs. Verbs do not appear in a single tense, and

Specific categories	Generic categories
Hostile Encounter	Attempt
Cause Harm	Cause change of position a scale
Destroying	Preventing or letting
Death	Cause change
Conquering	Seeking to achieve
Killing	Removing
	Purpose
	Event
	Assistance
	Causation
	Intentionally act
	Success or failure

Table 2: Categories of conflict-related events

Events	Occurrences	Conflict occurrences	Percentages
Hostile encounter	76	76	100%
Killing	19	7	37%
Conquering	12	9	75%
Death	10	6	60%
Destroying	5	5	100%
Cause harm	1	1	100%

Table 3: Specific categories

Events	Occurrences	Conflict occurrences	Percentages
Intentionally act	431	15	3%
Attempt	107	35	33%
Causation	100	16	16%
Assistance	68	17	25%
Purpose	45	17	38%
Removing	44	2	5%
Seeking to achieve	33	12	36%
Preventing or letting	17	6	35%
Cause change	14	2	14%
Cause change of position on a scale	7	4	57%
Event	4	2	50%
Success or failure	1	1	100%

Table 4: Generic categories

Events	Triggers	Definition
Hostile encounter	"guerras"	To report a conflict
Killing	“morrer”, “mortos”, “matar”,	To report an assasination or cause of the death
Conquering	"conquista"	To search to conquer a fortress/ city/ territory
Death	“morte”	To report a death
Destroying	“destruir”	To report destruction of something or someone
Cause harm	“feriram”	To cause or to report an injury

Table 5: Specific categories and respective trigger terms and definitions

Events	Triggers	Definition
Intentionally act	“fazer”, “proceder”, “efectuar”	To intentionally take a concrete action
Attempt	“procurar”, “intentar”, “pretensão”	To aim to take a certain action
Causation	“causar”, “resultar”, “causa”	To provoke a reaction
Assistance	“ajuda”, “ajudar”, “servir”	To help achieving something
Purpose	“intentem”, “mandar”, “pretender”	To express intention to achieve a goal
Removing	"tirar"	To take something off
Seeking to achieve	"procurar", "buscar"	To take an action to achieve one goal
Preventing or letting	“deixar”, “impedir”, “permitir”	To allow or impede something
Cause change	“mudar”, “converter”, “alterar”	To take an action to promote change
Change pos on a scale	"diminuir", "reduzir"	To intent the defeat of something
Event	“acontecer”	To report something that had occurred
Success or failure	“conseguir”	To accomplish or fail in a certain goal

Table 6: Generic categories and respective trigger terms and definitions

there can be more than one tense per event. Still triggers may not necessarily be verbs, such as the triggers for the event category hostile encounter, whose main term is the word “guerra (s)”.

We identified events that are directly related to conflicts, such as the hostile encounter event, whose trigger is the word "guerra". As the results in table 3 reflect, the hostile encounter event category is the one with the highest number of conflict occurrences, totalling 76 events. The 1614-16 period, under the rule of the Portuguese Viceroy Dom Jerónimo de Azevedo, was part of a specific conjuncture of internal political (re)equilibrium in most of Asian regions (Subrahmanyan, 1993; Thomaz, 1998). Since then, the Portuguese Estado da Índia has never been a continuous set of continental territories, nor also complied to territories under formal administration of the Portuguese crown, as were the cities where a significant Portuguese community was settled as those in the Malabar Coast. As F. Bethencourt describes it encompassed also "(...) all the Christian communities, sedentary or in transit, who were in some way involved in the various forms of jurisdiction delegated by the Portuguese king." (Bethencourt, 1998). With the Habsburg dinasty in Portugal, and despite the crisis in the Cape Route navigation system, there was a set of territorial conquests, but the uprising of great Asian empires as the Mughals or the Marathas have determined a political reconfiguration of a vast territory from Persian and the Arabic Peninsula until the Eastern bank of the Bengal bay (Flores, 2015). Apart from that, the political powers of Asia foreseen the benefits of allying with other European powers in order to diminish the naval power of the Portuguese. That was the case of the alliance between the Persia Xa and the English East India Company resulting in the loss of the Portuguese fortress of Goombron (1615), the establishment of the English fectory in Jask (1616) both events related in these set of events (Chaudhuri, 1985). Therefore, it is not striking that 5 per cent of the events report directly to conflicts in this time period.

Table 3 shows how the 6 specific categories of events are almost entirely dedicated to conflicts and war, as the semantic domain of them reports to killing, destruction, conquer and war. Nevertheless, event generic categories reporting the achievement of a success or reducing or weak an enemy are totally or significantly dedicated to military events. As table 4 reveals, the frequency of more generic event categories such as *Intentionally Act* or *Attempt* related to conflict mostly describe orders issued by the Portuguese Crown to take action to build or strengthen defensive structures or to attack certain Asian powers, and to plan certain actions mostly against the Dutch, who were allied with certain authorities from Eastern Indian coast and Ceilan as well (Abeyasinghe, 1966; Boxer, 1969).

Curious is the presence of the categories like *Assistance* or *Purpose* in conflict terms. The first, as a counter effect of the troubled historical juncture seeks to ask for solution to help solving problems related with the competition of both Europeans and Asian powers.The second, also profoundly related with the category *Seeking to Achieve*, relates to the intention to achieve a certain goal. As table 4 shows, 38 and 36 per cent of such categories are linked to conflict related events which demonstrate

the worrying of the Portuguese authorities with this climate of general confrontation against the Portuguese power in the Indian ocean.

Based on this analysis, we consider that an automatic event extraction with a semantic categorization allows the historian to rapidly identify characteristics of a certain time period, as this one in the Portuguese Asia, by identifying semantics of the text. Although we have tried to apply such analysis to conflict related events, the events classification and the proper statistics of such extraction only allow a rapid consideration of the hot topics discussed among the authorities of *Estado da Índia.*

7 Conclusion

This paper presented the application of TEFE, an event detection tool, to the study of 17th century historical events in Portuguese Asia, as registered in the Monsoon Books. The collection of extracted events and its classification have enhanced the historian to realise the concern of 17th century Portuguese authorities with the overall conjuncture of conflict in Asia between 1641 and 1616, in different areas of Estado da India (Persian Gulf, the Mughal Empire in northern Industan Peninsula, Ceilan) and with different Asian and European political units.

We have identified specific categories related within the semantic field of conflicts, but did not limit the analysis to these categories, since we have found conflict-related events also distributed in other more generic categories. That analysis has enabled us to analyse war as an historical phenomenon in the entire region in detail. We could observe how certain events reveal open direct conflict and others report a tense political relation that could have derived or not in a certain form of conflict.

Although our analysis could indicate points where the tool may be improved, in this work, we primarily focused on the output provided by the tool for an analysis of the semantic field of conflict. In fact we found that the set of events detected and classified were helpful to corroborate aspects investigated by historians regarding that period of time. Also, the events' extraction and classification allows the historian to rapidly detect trends on the topics mostly discussed in such a vast documental corpora and to raise his/her awareness that the semantics employed relates to specific historical junctures.

Acknowledgements

This work has received financial support from the Portuguese Science Foundation FCT, in the context of the projects 2022.07730.PTDC, UIDB/00057/2020, and CEECIND/01997/2017, and from the Brazilian funding agency CAPES Finance Code 001.

References

Tikiri Abeyasinghe. 1966. *Portuguese rule in Ceylon 1594-1612*. Colombo.

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8.

Cátia Antunes. 2012. Free agents and formal institutions in the portuguese empire: Towards a framework of analysis. *Portuguese Studies*, 28(2):173–185.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Francisco Bethencourt. 1998. O estado da índia. In *F. Bethencourt and K. Chaudhuri (eds.), História da Expansão Portuguesa*, pages 284–314. Círculo de Leitores.

Steven Bird. 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 69–72.

C. R. Boxer. 1969. Portuguese and spanish projects for the conquest of southeast asia, 1580—1600. *Journal of Asian History*, 3(2):118–136.

Helena Freire Cameron, Fernanda Olival, Renata Vieira, and Joaquim Francisco Santos Neto. 2022. Named entity annotation of an 18th century transcribed corpus: problems, challenges. In *Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing (2nd DHandNLP 2022) co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2022), Virtual Event, Fortaleza, Brazil, 21st March, 2022*, volume 3128 of *CEUR Workshop Proceedings*, pages 18–25. CEUR-WS.org.

Paula Carvalho, Hugo Gonçalo Oliveira, Diana Santos, Cláudia Freitas, and Cristina Mota. 2008. Segundo HAREM: Modelo geral, novidades e avaliaçao. *quot; In Cristina Mota; Diana Santos (ed) Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM Linguateca 2008.*

Kirti N. Chaudhuri. 1985. *Trade and civilisation in the Indian Ocean: an economic history from the rise of Islam to 1750*. Cambridge University Press.

Linguistic Data Consortium. 2005. ACE (automatic content extraction) english annotation guidelines for events. *Version*, (5.4.3).

Francisco Costa and António Branco. 2012a. LX-TimeAnalyzer: A temporal information processing system for portuguese.

Francisco Costa and António Branco. 2012b. TimeBankPT: A TimeML annotated corpus of Portuguese. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3727–3734, Istanbul, Turkey. European Language Resources Association (ELRA).

Silvio R. Dahmen, Ana L. C. Bazzan, and Robert Gramsch. 2017. Community detection in the network of german princes in 1225: A case study. In *Complex Networks VIII: Proceedings of the 8th Conference on Complex Networks CompleNet 2017 8*, pages 193–200. Springer.

Jorge Flores. 2015. *Nas Margens do Hindustão: o estado da Índia e a expansão mongol ca. 1570-1640*. Imprensa da Universidade de Coimbra/Coimbra University Press.

R Gramsch. 2014. Conflicts as a structure-forming force: the reign of henry (vii)(1225–1235) in network-analytic perspective. *Multiplying Middle Ages. New Methods and Approaches for the Study of the Multiplicity of the Middle Ages in a Global Perspective (3rd 16th CE)*.

Diego Jimenez-Badillo, Patricia Murrieta-Flores, Bruno Martins, Ian Gregory, Mariana Favila-Vásquez, and Raquel Liceras-Garrido. 2020. Developing geographically-oriented nlp approaches to sixteenth-century historical documents: digging into early colonial mexico. *Digital Humanities Quarterly*, 14(4).

Jian Liu, Yubo Chen, Kang Liu, and Jun Zhao. 2018. Event detection via gated multilingual attention mechanism. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Barbara McGillivray, Thierry Poibeau, and Pablo Ruiz. 2020. Digital humanities and natural language processing:"je t'aime... moi non plus". *Digital Humanities Quarterly*, 14(2).

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 365–371. Association for Computational Linguistics.

Trung Minh Nguyen and Thien Huu Nguyen. 2019. One for all: Neural joint modeling of entities and events. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6851–6858.

R. A. B. Patto. 1893. *Documentos Remetidos da India ou Livros das Monções. Tomo IV*. Academia Real das Ciências de Lisboa, Lisboa.

M. N. Pearson. 1987. *The Portuguese in India*. Cambridge University Press, Cambridge.

Michael Piotrowski. 2012. *Natural language processing for historical texts*. Morgan & Claypool Publishers.

Sandra D. Prado, Sílvio R. Dahmen, Ana LC Bazzan, Máirín MacCarron, and Julia Hillner. 2020. Gendered networks and communicability in medieval historical narratives. *Advances in Complex Systems*, 23(03):2050006.

Paulo Quaresma, Vítor Beires Nogueira, Kashyap Raiyani, and Roy Bayot. 2019. Event extraction and representation: A case study for the portuguese language. *Information*, 10(6):205.

Ana Sofia Ribeiro. 2016. *Early Modern Trading Networks in Europe. Cooperation and the case of Simon Ruiz*. Routledge.

Anderson da Silva Brito Sacramento. 2021. Um método computacional de extração automática de eventos em domínio fechado na língua portuguesa. Master's thesis, Unviersidade Federal da Bahia.

Anderson da Silva Brito Sacramento and Marlo Souza. 2021. Joint event extraction with contextualized word embeddings for the portuguese language. In *Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10*, pages 496–510. Springer.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS*, pages 403–417, Rio Grande do Sul, Brazil. Springer.

Rachele Sprugnoli and Sara Tonelli. 2019. Novel event detection and classification for historical texts. *Computational Linguistics*, 45(2):229–265.

S. Subrahmanyan. 1993. *The Portuguese Empire in Asia, 1500-1700: a political and economic history*. Longman, London/New York.

L. F. R. Thomaz. 1998. *De Ceuta a Timor*. Difel, Lisboa.

Renata Vieira, Fernanda Olival, Helena Cameron, Joaquim Santos, Ofélia Sequeira, and Ivo Santos. 2021. Enriching the 1758 portuguese parish memories (alentejo) with named entities. *Journal of Open Humanities Data*, 7:20.

Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.