

# nLORE: A Linguistically Rich Deep-Learning System for Locative-Reference Extraction in Tweets

Nicolás José FERNÁNDEZ-MARTÍNEZ<sup>a,1</sup> and Carlos PERIÑÁN-PASCUAL<sup>b</sup>

<sup>a</sup> *Universidad Católica San Antonio de Murcia (Spain)*

<sup>b</sup> *Universitat Politècnica de València (Spain)*

**Abstract.** Location-based systems require rich geospatial data in emergency and crisis-related situations (e.g. earthquakes, floods, terrorist attacks, car accidents or pandemics) for the geolocation of not only a given incident but also the affected places and people in need of immediate help, which could potentially save lives and prevent further damage to urban or environmental areas. Given the sparsity of geotagged tweets, geospatial data must be obtained from the locative references mentioned in textual data such as tweets. In this context, we introduce nLORE (neural LOcative Reference Extractor), a deep-learning system that serves to detect locative references in English tweets by making use of the linguistic knowledge provided by LORE. nLORE, which captures fine-grained complex locative references of any type, outperforms not only LORE, but also well-known general-purpose or domain-specific off-the-shelf entity-recognizer systems, both qualitatively and quantitatively. However, LORE shows much better runtime efficiency, which is especially important in emergency-based and crisis-related scenarios that demand quick intervention to send first responders to affected areas and people. This highlights the often undervalued yet very important role of rule-based models in natural language processing for real-life and real-time scenarios.

**Keywords.** location extraction, geolocation, named entity recognition, deep learning

## 1. Introduction

The large volume of user-generated content on Twitter can be exploited in social sensing settings where crisis management and tracking become of utmost importance for disaster relief operations [1,2]. Having a rapid understanding about crisis-related and emergency events can help handle human and economic resources effectively through immediate and timely decisions and actions taken by aid organizations and competent authorities. As a result, emergency responders can coordinate effective aid and help allocate resources in the affected areas and/or to the affected people, potentially saving lives and preventing further damage to environmental or urban areas. Because of the lack of geotagged data, e.g. geotagged tweets represent only around 1% of tweets [3], and its recent sharing restrictions, it becomes necessary to turn to other geospatial evidence, such as that found in tweets in the form of locative references, which are usually much more frequent than geotagged data [4]. Twitter has in fact been

---

<sup>1</sup> Corresponding author. E-mail: njfernandez@ucam.edu

exploited in many geolocation systems that handle real-life scenarios, e.g. natural or human-made disaster detection and tracking in floods, earthquakes, storms, civil unrest, war, crime, etc. [1,2,5], health surveillance and disease tracking, including the current COVID-19 pandemic [6], marketing and advertising [7], or traffic incident detection, road traffic control and/or traffic congestion [8]. In this regard, the location dimension proves to be critical for raising situation awareness of crisis-related events and understanding their impact, i.e. gaining insight into where the incident happened, where people require assistance, and/or which areas were affected.

Departing from previous work on location detection from tweets [9,10], we introduce a deep-learning implementation of LORE (i.e. neural LORE, or nLORE), which, by means of a bi-directional Recurrent Neural Network (bi-RNN) with linguistic feature engineering, automatically learns the linguistic features and rules provided by LORE to extract locative references of any type from English tweets with even greater accuracy. Indeed, nLORE slightly outperforms LORE, which already outperformed other entity recognizers such as Stanford NER, spaCy, Stanza, OpenNLP, and Google Entity Recognizer, offering a great qualitative and quantitative advantage over these general-domain entity recognizers and qualitatively over tweet-specific location-detection models. However, LORE shows much better speed and efficiency, being the preferred model in real-life crisis-related and emergency-based contexts, where quick deployment and allocation of resources by competent authorities is essential to come to the rescue of affected people or areas.

The article is organized as follows. Section 2 presents the major works in locative reference detection from tweets. Section 3 explains the methodology followed in the training phase of nLORE. Section 4 describes the evaluation phase. Section 5 presents the results obtained by nLORE in comparison with LORE and discusses the results and implications derived from this study, including future lines of research. Finally, Section 5 presents some conclusive remarks.

## 2. Related Work

Many works propose rule-based systems that rely on a combination of linguistic rules (e.g. locative preposition + proper nouns) and/or lexical resources (e.g. gazetteers of place names or lists of locative cues) together with supportive natural language processing (NLP) tasks, such as the part-of-speech (POS) tagging of n-grams [3,11,12]. Other systems are based on probabilistic frameworks, using Conditional Random Fields (CRF) or neural networks such as Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN) with or without additional linguistic feature engineering [7,13]. Others present hybrid systems, which combine rule-based methods with machine- or deep-learning techniques and/or other named entity recognition (NER) tools [8,14]. What all these works share is the limited semantic coverage of fine-grained location types, as well as the lack of a sound linguistic-based approach for their identification. In the remainder of this section, we describe the most prominent works in locative reference extraction from tweets.

A microtext location-detection model was proposed by [14] using regex-based rules, the Open Calais NER software, machine-learning techniques for abbreviation disambiguation, and the National Geospatial Intelligence Agency gazetteer for the identification of places in New Zealand and Australia at and within city level, such as geopolitical entities, buildings, and streets. It was tested on two evaluation datasets: one

about the 2011 Christchurch dataset in New Zealand and another about the 2011 Texas wildfire in the US. In the first evaluation dataset, the model achieved an F1 score of 0.85 for streets, 0.86 for buildings, 0.96 for geopolitical entities, and 0.88 for place abbreviations, giving an average of 0.9. In the second dataset, it obtained an F1 score of 0.71.

A linguistic-based unsupervised location-detection model was presented by [11] based on linguistic techniques and rules such as noun-phrase extraction and n-gram matching techniques using regex-based rules and the GeoNames database. It targeted geopolitical entities, points of interest (POIs), addresses, and surrounding distance and direction markers, giving an F1 score of 0.79.

A rule-based location-detection model for English tweets was developed by [3] using the OpenStreetMaps database. They used NLP techniques such as the NLTK sentence tokenizer, entity-based matching with OpenStreetMaps using an n-gram-based module, their own corpus of building and street types, and the NLTK stopword list enriched with a list of names. They focused on geopolitical entities, buildings, and streets. The evaluation stage was carried out with separate corpora of tweets about different incidents (i.e. blackout, earthquake, and hurricane) in different geographic areas (i.e. Christchurch, Milan, New York, and Turkey). Precision numbers were impressive, ranging from 0.93 to 0.99, and F1 scores ranged from 0.90 to 0.97, except for the Turkey earthquake dataset, where an F1 score of 0.28 was achieved.

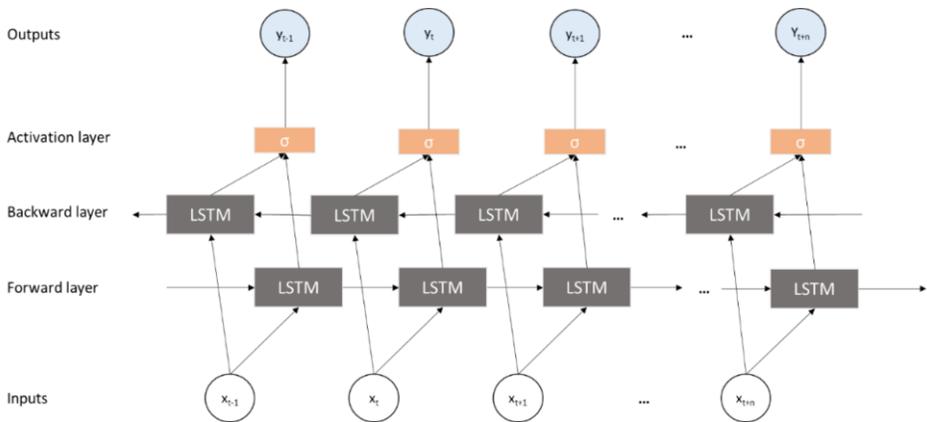
A location extractor called NeuroTPR was built by [13], who used a bi-directional RNN with Long Short Term Memory (LSTM) enriched with linguistic features. Apart from building a tweet corpus from the 2017 Hurricane Harvey dataset, they also used GeoCorpora [4]. The targeted location types were geopolitical entities, natural landforms, POIs, and a few traffic ways. In the evaluation stage, they compared their model against standard off-the-shelf NER tools such as Stanford NER, spaCy NER, a basic bi-directional LSTM-CRF model and another deep-learning model from the 2019 SemEval geoparsing competition. The evaluation phase was carried out using the exact matching of location references. The best NeuroTPR model, trained on 3000 Wikipedia articles and 599 tweets, achieved a precision score of 0.787, a recall score of 0.678, and an F1 score of 0.728 on the *Harvey2017* corpus.

### 3. Method

We implemented a hybrid approach to NER by means of a bi-directional RNN with LSTM as a hidden layer structure and a CRF layer on top exploiting linguistic feature engineering and semantic information contained in word embeddings. Our primary goal was to check the power of linguistic feature engineering in nLORE. Derived from this goal, our secondary goals were (i) to elucidate whether linguistic feature engineering could overcome the sparsity of data in probabilistic models and (ii) to compare the performance of nLORE against its rule-based counterpart LORE, which outperformed well-known general-domain NER systems [9], with the aim of resolving whether our rule-based model LORE is as powerful and accurate as nLORE without the computational cost, time, and resources required in probabilistic approaches.

### 3.1. Architecture of the Model

The underlying idea in the use of RNNs in NLP scenarios is that language must be treated as a temporal phenomenon, that is, a linguistic realization is perceived as a sequence of tokens that are combined one after another, where the prediction of a given word is dependent on earlier words. A bi-directional RNN consists of two simple RNNs stacked on top of each other, where not only the previous values are taken into account but also the following ones by means of backpropagation and forward-propagation. For the hidden layer structure, LSTM served to keep in memory distant contextual information, which is appropriate given the nature of language, since nearby contextual information is not sufficient to predict sequences of words. For the output layer structure, we used a CRF layer on top of the bi-directional RNN with LSTM for its great predictive power when contextual information is provided (Figure 1).



**Figure 1.** Neural architecture of the model.

### 3.2. Training Stage

Table 1 summarizes the optimal hyperparameters used for training nLORE. These values were chosen taking into account not only performance but also computational time and cost, striking a balance between both. Dropout values typically range from 0.2 to 0.5, representing a tiny fraction of output values which are set to zero during training; 0.5 provided the best performance. As for the hidden LSTM layer(s), we set only one hidden LSTM layer with a size of 200 neurons, to find a compromise between time and performance. The learning rate, whose value typically ranges from 0.001 to 1, affects the learning speed by determining how quick the weights shift in each training iteration. The lower the learning rate, the more accurate the estimation because of smaller increments in weights, but the more time it takes to train the network; 0.1 gave the best results. Mini-batch size selects a random subset of training data (i.e. a mini-batch) to be used in each iteration. Optimal performance in terms of speed and accuracy was observed with mini-batch sizes of 2 up to 32; 16 was the optimal number in terms of the cost-benefit ratio.

**Table 1.** Hyperparameters in the training phase of nLORE

Parameters	Values
Network type	Bidirectional RNN
Dropout	0.5
Hidden layer type, number and size	LSTM, 1, 200
Output layer type	CRF
Learning rate	0.1
Mini-batch size	16

We trained nLORE through eight models resulting from the combination of different corpus sizes (i.e. 1,000, 3,000, 5,000, and 7,000 tweets) and linguistic features (i.e. basic or extended features). The whole collection of tweets was compiled using a set of keywords related to a variety of crisis and emergency-related events: *earthquake, flood, car accident, bombing attack, shooting attack, terrorist attack, and incident*. After pre-processing the collection of tweets, which involved removing duplicates and nearly identical tweets, we obtained 8,063 microtexts, of which 7,000 were used for training and 1,063 were used for validation. The corpora were annotated taking into consideration the location types shown in Table 2.

**Table 2.** Location types in LORE and nLORE

Location type	Subtypes	Examples
Geopolitical entities	country, state, region, province, city, town, kingdom, villa...	<i>China, New York, Buenos Aires</i>
Natural landforms	mountain, mount, ridge, volcano, valley, lake, river, shore, beach, park, canyon...	<i>Mount Everest, Grand Canyon, Lake Michigan</i>
POIs	building, museum, school, station, stadium, garden, café, tavern, hospital, court, theater, residence, zoo, casino, square...	<i>Victoria Coach Station, Fox Valley Animal Referral Center, Hotel Park Villa</i>
Traffic ways	Highways, roads, addresses...	<i>I-90 SW, 11 Croydon Road, northbound J19</i>

The locative markers surrounding these location types were also annotated on the basis of the following taxonomy:

- Distance markers: kilometre(s), kilometer(s), km(s), metre(s), meter(s), m(s), mile(s), mi(s), yard(s), yd(s)...
- Directional markers: North, N, Northeast, NE, Northwest, NW, South, S...
- Movement markers: Northbound, NB, Southbound, SW...
- Temporal markers: hour(s), hr(s), h(s)...

In this regard, the complexity of the phrasal structure of a locative reference is illustrated in Figure 2, where the asterisk is used to mark optionality, and the double asterisk refers to the optional presence of locative markers either at the beginning or at the end of the locative reference.

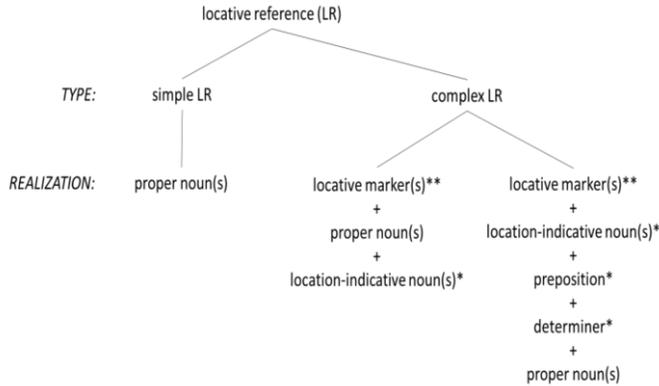


Figure 2. Phrasal structure of locative references

We adjusted the size of the training corpus (i.e. 1,000, 3,000, 5,000, and 7,000 tweets) in the training phase to determine whether linguistic feature engineering can overcome the problem of sparse data in a probabilistic framework. Moreover, considering that one of our purposes was to examine the usefulness of linguistic features, we differentiated between a basic model of features, focusing on those typically used in NER (i.e. token form and POS tag), and an extended model, including features such as token form, POS tag, presence of locative marker, and record in our place-name dataset obtained from GeoNames or in our location-indicative noun dataset obtained from WordNet (including words such as *street*, *restaurant*, *highway*, etc.)<sup>2</sup>.

Other features employed were template features, context template features and word embeddings trained on a corpus of over 3.8m English tweets (Cheng et al., 2010)<sup>3</sup>. The template features for the basic model consisted of the preceding, current and following token form (i.e.  $t_{i-1}$ ,  $t_i$ , and  $t_{i+1}$ ) as well as their POS tags (i.e.  $p_{i-1}$ ,  $p_i$ , and  $p_{i+1}$ ) and the combination of both. For the extended model, template features also included:

- the preceding, current and following tokens belonging or not to the place-name dataset (i.e.  $pn_{i-1}$ ,  $pn_i$ , and  $pn_{i+1}$ ) in combination with their POS tag,
- the current and following token that may belong in the location-indicative noun dataset (i.e.  $li_i$ ,  $li_{i+1}$ ) in combination with their token form, and
- the preceding, current and following token that may be a locative marker (i.e.  $lm_{i-1}$ ,  $lm_i$ ,  $lm_{i+1}$ ).

With respect to context template features, the context window considered the preceding token, the current token, and the following token, so that template features previously discussed could also be combined with the preceding and following tokens. Regarding

<sup>2</sup> We refer the reader to [9,10] for a much more detailed explanation on the linguistic-based criteria for the development of LORE, i.e. the semantic coverage of locative references and the compilation of lexical resources.

<sup>3</sup> Available on the following link: [https://archive.org/details/twitter\\_cikm\\_2010](https://archive.org/details/twitter_cikm_2010)

the trained word embeddings, we used the skip-gram model with Word2Vec vectors [16]. The intuition behind the role of word embeddings in our model was that it could help detect tokens that are similar to those which are place names in GeoNames, location-indicative words or locative markers, and also neighboring tokens that are part of or that predict a given locative reference, such as prepositions, proper nouns, etc. We specified a context window for word embeddings that included the current token, the preceding two tokens, and the following two tokens. All these features provided the best results in terms of computational resources and time in the training and evaluation phases. The resulting eight models were saved when they achieved the best result in the validation corpus. Iterations ranged from 20 to 25: the larger the corpus and the higher the number of feature types that were taken into account, the greater time it took to train each model.

#### 4. Evaluation

The evaluation corpus was compiled on a different date to obtain different types of incidents with different locative references, thus ensuring that there was no overfitting. The compilation, pre-processing and annotation steps were the same, resulting in a corpus of 1,372 tweets.

We followed the evaluation measures most widely used in NER (i.e. precision, recall, and F1), as shown in Eqs. (1)-(3), where TP, FP and FN denote the number of true positives, false positives and false negatives, respectively.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

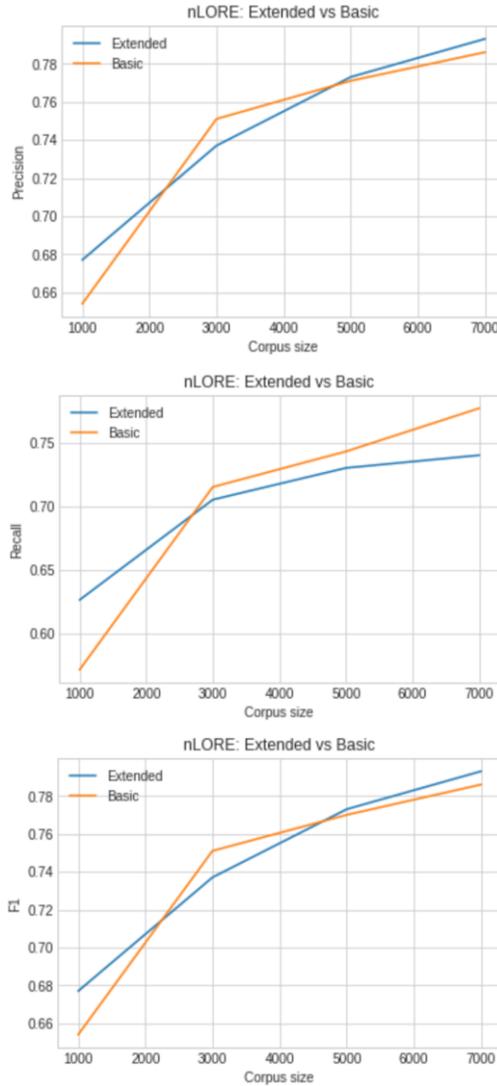
$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

As can be noted, F1 is the harmonic mean of precision and recall. All the measures return a value ranging from 0 to 1. The evaluation method was strict, thus considering full matches only.

#### 5. Results and Discussion

Figure 3 shows the precision, recall, and F1 scores of nLORE in the eight models resulting from the combination of different corpus sizes and linguistic-based features.



**Figure 3.** Precision, Recall, and F1 scores of nLORE, respectively.

It should be noted that precision was higher in the basic 1k and 3k models but increased exponentially in the extended model when the training dataset was larger, outperforming the basic 5k and 7k models. In the case of recall, the opposite is true. Whereas recall was higher for the extended 1k model, it was eventually outperformed by the basic 3k, 5k and 7k models. Therefore, the contribution of the extended linguistic features seems to be, at best, poorly significant and, at worst, only incidental. However, a trend can be noted with the 1k extended model. In fact, if we were to sketch the main reason in support of the use of extended linguistic features, this would be the better performance of the extended model over the basic model when corpus size is fairly small. In contrast, the difference in performance between basic and extended models tends to diminish as corpus size becomes increasingly larger and, even in this scenario, providing extra linguistic features might seem a bit counterproductive, since

they worsen performance very slightly, as can be concluded from the recall scores, despite improving precision.

Our intuition behind the higher precision of the extended 5k and 7k models is that the extended linguistic features may help avoiding the extraction of wrong instances that cannot be properly discarded by the token form and POS tag features alone. However, adding extended linguistic features might have a negative impact –though very slight– on the identification of right locative references, as spotted in the recall scores. Incorporating these features may have incidentally made the process of locative-reference extraction stricter, which explains the fewer number of FPs and the greater number of FNs.

In the light of these results, we selected the 7k extended model because it offered the best F1 score. Therefore, we compared the 7k extended model against our rule-based model (Figure 4), where we can observe that nLORE slightly outperformed LORE in terms of precision (0.85 vs 0.73) and F1 (0.79 vs 0.76) but not recall (0.74 vs 0.79).

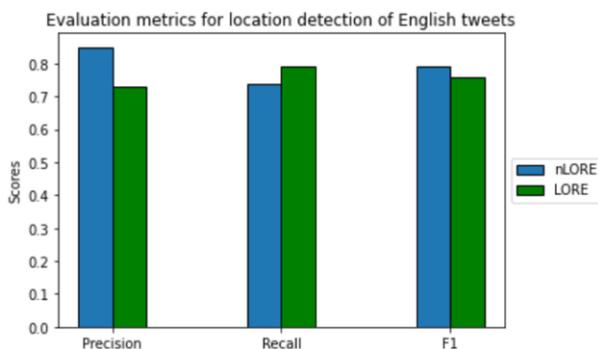


Figure 4. nLORE vs LORE.

The 7k extended nLORE model was able to extract all the locative references in Example 1 and Example 2, whereas LORE missed part of the locative references *NY 22*, *NY 138*, and *Northbound Exit Ramp*.

- (1) Incident on I684 NB at Exit 6A – NY 22 to NY 138 – Goldens Bridge (Northbound Exit Ramp).
- (2) earthquake 995 km north-east of Whangarei

At times, nLORE failed in the extraction of locative references such as *Port Harcourt airport* (Example 3), whereas LORE was able to detect it.

- (3) FAAN shuts down Port Harcourt airport temporarily over bush fire incident

Sometimes, LORE misidentified person names as locative references. In Example 4, *Clancy* was extracted as a locative reference because the rule-based module could not filter it out. However, nLORE was able to prevent its extraction.

- (4) I've read all of Clancy novels

In Example 5, LORE relies on rule-based patterns to detect *I-39B* as a traffic way, but nLORE missed this one, despite traffic ways being annotated as locative references.

#### (5) Police activity on I-39B

In short, nLORE incorporates in the system symbolic-based knowledge through linguistic feature engineering. This serves to intelligently infer and predict locative patterns in tweets by avoiding falling into errors caused by POS tags mislabeled by the Stanford POS tagger or mismatched locative items from the place-name dataset, which LORE could not prevent. The rule-based system tended to overmatch, negatively affecting precision but correlated with better recall, whereas nLORE was stricter, avoiding the extraction of wrong items but missing a few instances, hence negatively affecting recall but improving precision.

However, in terms of speed and efficiency, LORE greatly outperformed nLORE. On the one hand, LORE did not need a training phase, hence saving a great deal of time in its development. On the other hand, whereas LORE was able to process and extract all the locative references from the evaluation corpus in 12 seconds, nLORE lasted for almost 2 minutes. In other words, the rule-based model was almost 10x times quicker than its probabilistic counterpart. This supports the claim that rule-based models show better runtime efficiency than probabilistic models [17]. This is especially important in real-life, real-time emergency-based or crisis-related scenarios where time management becomes of utmost importance for emergency responders to locate the incident and act as quickly as possible. Moreover, LORE was already tested with a different evaluation corpus of English tweets [9], achieving similar evaluation numbers, which confirms its generalizability to other unseen collections of tweets. In addition to that, LORE already supports Spanish and French, having better performance than other NER systems with evaluation corpora in those languages [10]. Therefore, nLORE would be preferred over LORE in terms of performance, whereas LORE would be the preferred option when speed and efficiency are the most important concerns, which is indeed the case in emergency- or crisis-related settings. Even if rule-based approaches are seen as a dead-end technology in academia, they are nevertheless employed in companies and industries for real-life scenarios that demand efficiency, quick performance and scalability [17].

For future research, we could use larger training corpora with nLORE to check if performance can be enhanced and how speed is affected. We also leave for future research the use of different novel approaches in NLP, such as contextualized word embeddings and pre-trained language models, e.g. BERT, or the adaptation of nLORE to other languages such as Spanish or French, so that nLORE could become a multilingual NER system.

## 6. Conclusion

We have presented nLORE, a deep-learning system for fine-grained locative-reference extraction in tweets, feeding off the linguistic knowledge provided by LORE and trained on a relatively small corpus of English tweets, so that locative references can be inferred and extracted with slightly greater accuracy than LORE. In some respects, nLORE overcame some of the limitations presented by LORE, as evidenced by evaluation results and the examples above discussed. This research also demonstrates

that linguistic feature engineering in probabilistic approaches may still provide a much-valued added benefit, especially when confronted with data sparsity, which could pave the way for more linguistic-oriented computational work in the field of NER. This approach goes in line with recent calls in linguistic and computational communities, requesting a greater interaction between linguistics and artificial intelligence [18]. At the same time, LORE still shows promising speed and efficiency and a performance almost as good as that of nLORE, which means that rule-based models are still a valuable and viable solution, despite the often underappreciated value of such models in the current NLP landscape. Given the importance of an efficient and quick response of emergency-based services and authorities in crisis-related scenarios, rule-based models such as LORE can help provide great value to systems oriented to give aid to affected people as quickly as possible, potentially saving lives or preventing further damage in affected areas.

### Acknowledgements

Financial support for this research has been provided by the Spanish Ministry of Science, Innovation and Universities [grant number RTC 2017-6389-5: project WATERoT] and by the European Union's Horizon 2020 research and innovation program [grant number 101017861: project SMARTLAGOON].

### References

- [1] Zhang C, Fan C, Yao W, Hu X, Mostafavi A. Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*. 2019; 49: 190-207.
- [2] Martínez-Rojas M, Pardo-Ferreira M del C, Rubio-Romero JC. Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. *International Journal of Information Management*. 2018; 43: 196-208.
- [3] Middleton SE, Kordopatis-Zilos G, Papadopoulos S, Kompatsiaris Y. Location Extraction from Social Media. *ACM Transactions on Information Systems*. 2018; 36(4): 1-27.
- [4] Wallgrün JO, Karimzadeh M, MacEachren AM, Pezanowski S. GeoCorpora: building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science*. 2018; 32(1): 1-29.
- [5] Imran M, Castillo C, Diaz F, Vieweg S. Processing Social Media Messages in Mass Emergency: A Survey. *ACM Computing Surveys*. 2014; 47: 1-38.
- [6] Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. 2020; Available from: <http://arxiv.org/abs/2003.13907>
- [7] Inkpen D, Liu J, Farzindar A, Kazemi F, Ghazi D. Location detection and disambiguation from twitter messages. *Journal of Intelligent Information Systems*. 2017; 49 (2): 237-253.
- [8] Das RD, Purves RS. Exploring the Potential of Twitter to Understand Traffic Events and Their Locations in Greater Mumbai, India. *IEEE Transactions on Intelligent Transportation Systems*. 2019; 1-10.

- [9] Fernández-Martínez NJ, Perrián-Pascual C. LORE: a model for the detection of fine-grained locative references in tweets. *Onomázein*. 2021.
- [10] Fernández-Martínez NJ, Perrián-Pascual C. Knowledge-based rules for the extraction of complex, fine-grained locative references from tweets. *Revista Electronica de Linguistica Aplicada*. 2020; 19(1): 136-63.
- [11] Malmasi S, Dras M. Location mention detection in tweets and microblogs. In: Hasida K, Purwarianti A, editors. *Communications in Computer and Information Science*. Singapore: Springer; 2016; 123-134.
- [12] Al-Olimat HS, Thirunarayan K, Shalin V, Sheth A. Location Name Extraction from Targeted Text Streams using Gazetteer-based Statistical Language Models. In: *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics; 2018; 1986-1997.
- [13] Wang J, Hu Y, Joseph K. NeuroTPR: A Neuro-net ToPonym Recognition Model for Extracting Locations from Social Media Messages. *Transactions in GIS*. 2020; 1-22.
- [14] Gelernter J, Balaji S. An algorithm for local geoparsing of microtext. *GeoInformatica*. 2013; 17(4): 635-67.
- [15] Cheng Z, Caverlee J, Lee K. You are where you tweet: A content-based approach to geo-locating Twitter users. In: *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. 2010; 759-768.
- [16] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. In: *Proceedings of the International Conference on Learning Representations (ICLR 2013)*. 2013; 1-12.
- [17] Chiticariu L, Li Y, Reiss FR. Rule-based information extraction is dead! Long live rule-based information extraction systems! In: *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. 2013; 827-832.
- [18] Linzen TAL. What can linguistics and deep learning contribute to each other? Response to Pater. *Language*. 2019; 95(1): 99-108.