Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4521–4528
Marseille, 11–16 May 2020

Tagging Location Phrases in Text

Paul McNamee, James Mayfield, Cash Costello, Caitlyn Bishop, Shelby Anderson
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road, Laurel MD 20723-6099 USA
first.last@jhuapl.edu

Abstract

For over 30 years researchers have studied the problem of automatically detecting named entities in written language. Throughout this time the majority of such work has focused on detection and classification of entities into coarse-grained types like: PERSON, ORGANIZATION, and LOCATION. Less attention has been focused on non-named mentions of entities, including non-named location phrases. In this work we describe the Location Phrase Detection task. Our key accomplishments include: developing a sequential tagging approach; crafting annotation guidelines; building an annotated dataset from news articles; and, conducting experiments in automated detection of location phrases with both statistical and neural taggers.

Keywords: Location Phrase Detection, Named-Entity Recognition, Annotation

1. Introduction

During emerging events such as humanitarian crises, detecting mentions of locations is important. However, named locations such as Haiti, Port-au-Prince, Pétion-Ville, and Gulf of Gonâve represent only a fraction of location information present in text. The research described in this report examines more varied constructions containing finer-grained location information such as "the medical clinic in Telonge," "2 km below the Dolin Maniche bridge," "35 km southeast of the capital," or "behind the Christopher Hotel parking lot." Our work is motivated by the recent DARPA LORELEI program, which tested technologies that support "situational awareness tasks that are important for disaster relief planning and execution" (Christianson et al., 2018).

Representing spatial information is a core part of language. Most of the world's languages use prepositions, postpositions, or circumpositions to convey spatial relations. For example, in English, prepositions such as *on*, *near*, *by*, *to*, *from*, *around*, *under*, *beside*, and *between* express the spatial orientation between two objects. In many languages case markings (*e.g.,* ablative, locative cases) are also used to convey spatial information. For example, in Turkish, the word *village* is köy, and to say *in the village* you would write köyde. This work aims to automatically identify location phrases in text, irrespective of how that information is communicated in a particular language.

In Section 2 we review related work. In Section 3 we present annotation guidelines for a sequential tagging approach to the task. We then describe our experiences annotating data in English and Russian (Section 4). Section 5 gives experiments results for this task. And Section 6 summarizes our findings.

2. Related Work

Hassani and Lee (2017), studied spatial prepositions in English. They built a dataset for classifying when ambiguous prepositions are acting spatially and when they are not. For example, in "Tim is on the committee" **on** is not being used spatially. But in "Tim is on the bus" it is. They built a neural classifier for the task, and reported 94% accuracy in resolving whether ambiguous prepositions in English are being used in a spatial context. Similar work has been reported by Radke *et al.* (2019).

Skoumas *et al.* (2016) extract spatial relations from travel blogs and then georeference tourist sites solely from these textual descriptions. This work is focused on a narrow domain, but conceiveably such data might be used to train geotaggers for locations in other domains.

Moncla *et al.* also look at extracting locations, and in particular, sequences of locations (*i.e.,* itineraries) based on trail descriptions from hiking guides (2016). They use extracted information along with external gazetteers in a graph-based formalism to constrain possible interpretations of narratives. Unique to their dataset are ground-truth GPS positions, which are not available for more general texts.

The closest task to our effort in location phrase identification appears to be spatial role labeling. Kordjamshidi et al. (2011) created the spatial role labeling task and define the following semantic components:

- *Trajector*: the thing being located: "the **painting** on the wall"
- *Landmark*: a reference point relative to which the trajector is being located "the painting on the **wall**"
- *Region*: a region of space defined relative to a landmark (*e.g.,* interior, exterior)
- *Path*: a sequence with beginning, middle, and ending, describing the motion of the trajector
- *Motion*: an indicator of whether motion is perceived
- *Direction*: a path relative to a frame of reference (obviating the need for a landmark), e.g., south, left, above
- *Frame of Reference*: one of {intrinsic, relative or absolute}

In addition to trajectors and landmarks they also annotate *Spatial Indicators*, which are essentially prepositions or occasionally verbs like **surrounded** or **left**; *Motion Indicators*, which are usually prepositional verbs (*e.g.,* **flew**); and *Spatial Relations*, tuples of trajectors, landmarks, and their associated indicators. Shared tasks focused on this problem were held at SemEval 2012 (Kordjamshidi et al., 2012) and SemEval 2013 (Kolomiyets et al., 2013). One of the datasets used was based a set of tourist photographs with captions in English; these data are rich in locations. For the 2013 evaluation using this tourism dataset the top system reported an F_1 score of 93% in detecting spatial indicators, and 79% and 68% on landmarks and trajectors, respectively. Accuracy for extracting complete spatial relations was only 46%.

Early work similar to Kordjamshidi *et al.* has been conducted in Polish using corpora annotated according to the SemEval 2013 guidelines (Marcinczuk et al., 2016). Marcinczuk *et al.* annotated two corpora: 50 articles from Polish Wikipedia with extensive geographic mentions, and about 1,500 documents from the *KPWr Corpus* (Broda et al., 2012). They report an F_1 score of 41% for detecting spatial relations. This is lower than results reported at the English SemEval challenges, possibly suggesting that greater morphological complexity makes the task harder.

Finally, similar work in detecting named and nominal location mentions has been done by Blaylock *et al.* (2012), who identify locations based on a hybrid symbolic-statistical parser, and a manual set of hand-built semantic graph rules. They published experiments using the *PURSUIT* corpus, which was created by automobile drivers wearing headset microphones who were instructed to audibly describe their route.

3. Annotation Schema and Guidelines

The spatial role labeling task, previously described, contains a number of elements useful for fine-grained location analysis. However, for situational awareness in a disaster relief setting, locations should be the focus, and trajectors, the items being spatially located, are not as essential for the task. It is also concerning that the performance of detecting spatial relations (*i.e.,* tuples) is low, in the 40s.

Therefore we decided to adopt some of the framework defined by Kordjamshidi *et al.*, but to reformulate the task based on sequential labeling. We discard trajectors and path annotations, and focus on tagging words in a sentence according to semantic types relevant to locations. We retain Landmarks, Spatial Indicators, and Direction tags, and we add three additional tags: Diffuse, Adjectival, and GAGs (GPE acting as an AGent). Generally speaking we mark locations that are named, nominal, or pronominal. Thus there are six principal types and the implicit "O" type, which indicates unmarked (or **O**ther). In standard PER/ORG/GPE[1] coarse-grained named entity recognition (NER) only named locations are marked; here we remedy this deficit and additionally tag other words that are important for identifying locations and to understand their meaning.

The six types in our Location Phrase task are:

- Spatial Indicator (SI): a word used to indicate a spatial relationship
- Landmark (LAND): a potentially geolocatable place
- Diffuse (DIF): a non-geolocatable place
- Adjectival (ADJ): A GPE or LOC acting in an adjectival or possessive sense
- GPE Acting as an Agent (GAG): a GPE that is not acting in an expressly spatial role
- Directional (DIR): a word or phrase used to indicate cardinal direction, distance, proximity, areal containment, or adjacency.

Only locations are being tagged; other named entities like PERs and ORGs are not annotated for this task. In a real-world setting, separate taggers for traditional named entities and location phrases can be run. In addition to these six tags, we debated adding an additional tag for locational mentions that are part of another entity's name (*e.g.,* "**Baltimore** Department of Social Services." We decided against this, mainly for pragmatic reasons.

Below we describe the six tags in greater detail, and give examples[2] of how each should be marked.

3.1. Spatial Indicators

A Spatial Indicator (SI) is a preposition, or sometimes a verb, that indicates a spatial relationship between an object and a reference location. Prepositions are the most common spatial indicators.

3.1.1. Basic examples

- I stopped $[\text{in}]_{SI}$ Thailand on my trip $[\text{to}]_{SI}$ my aunt's farm $[\text{in}]_{SI}$ Japan.
- The cat is $[\text{in}]_{SI}$ my house.
- The earthquake pushed lava $[\text{into}]_{SI}$ new underground chambers.

Verbs may act as spatial indicators when they indicate a spatial relationships that is not otherwise specified.

- The students $[\text{occupying}]_{SI}$ the dean's office are committed to their cause.
- The troops $[\text{surrounded}]_{SI}$ the compound.
- He $[\text{left}]_{SI}$ Eritrea at age 21.
- Yagana, 18, $[\text{fled}]_{SI}$ her village when Boko Haram attacked.

3.1.2. Exclusions

The prepositions "of" and "from" should usually not be tagged as spatial indicators. To be taggable, those prepositions should have clear locational semantics.

See the discussion about "residents of" below in Section 3.7.5.

[1]A GPE is a geopolitical entity, essentially a populated location with a government. Countries, cities, towns, can all be GPEs. This type of NE was invented to account for the multiple roles that such entities possess. For example, "I went to Berlin" is explicitly spatial. "Berlin will sign the treaty with London" is a governmental/organization usage that is not spatial in nature.

[2]In example sentences in this section we bracket the location phrase words for the *illustrated* type. In some examples there are other words that should be tagged for a different type.

3.2. Landmark

A Landmark is the head of a GPE, LOC, FAC, or other discrete location. A Landmark may be a mention that is named, nominal, or pronominal, though it must act in a spatial sense.

3.2.1. Basic examples

- I stopped in [Thailand]$_{LAND}$ on a trip to my aunt's [farm]$_{LAND}$ in [Japan]$_{LAND}$.
- Yellow cab moving north on [14th Street]$_{LAND}$ from [Harvard Street]$_{LAND}$.
- Muscat's largest [market]$_{LAND}$, the [Muttrah Souq]$_{LAND}$, reopened Saturday, three days after [it]$_{LAND}$ was flooded.

Mentions that are non-specific might require a Diffuse tag (see Section 3.3).

3.2.2. Exclusions

Nouns that refer to a class of object rather than a specific one are not taggable.

- When standing on suspension bridges I get dizzy.
- All elevators are required to have an emergency phone.

Here *bridges* and *elevators* should not be tagged as a Landmark.
Named entities of locations that are part of an organization name or other entity are not tagged.

- I donated money to the Baltimore Museum of Art.

Not tagged. *Baltimore* is part of the official organization name.

- I shot an amazing photograph of an African Elephant while on a safari.

African is part of the species name, and so it is not tagged in this instance.

3.2.3. Permitted Use

Mentions of Landmarks may refer to locations in the past or future tense.

- When it is built the [Acme bridge]$_{LAND}$ will span the [Hudson river]$_{LAND}$.

Though the bridge does not presently exist, it should be marked as a Landmark.

- Yesterday I was buying a ticket at [Disneyland]$_{LAND}$ when I saw Donald Duck.

Mentions may be fictional, conditional, or hypothetical.

- I wish there were a [Wegmans]$_{LAND}$ in [Maple Lawn]$_{LAND}$.

Wegmans should be tagged, even though no such store presently exists. Named vehicles may be Landmarks.

- I just spent two weeks on the [USS Dwight D. Eisenhower]$_{LAND}$ supporting a naval exercise.

3.3. Diffuse Tag

The Diffuse tag is used for non-named locations that are generally not geolocatable. The location may be small, large in area, or imprecisely described.

3.3.1. Basic examples

- I left my keys on the [dresser]$_{DIF}$.

Dresser here is acting locationally; it is where the keys are.

- There was a lot of flooding in the [East]$_{DIF}$.

East is a broad, not crisply defined region, hence not a Landmark.

- I paid $40 for that dresser.

We do not tag *dresser*. It is a common noun and not being used as a location.

3.3.2. Exclusions

In some sense, any physical object has a location. But we avoid tagging objects, especially small objects, when they are not acting as a location. Large immovable objects (*e.g.,* buildings, bridges) should be tagged.

- The storm destroyed seven [homes]$_{DIF}$ in a nearby coastal [town]$_{DIF}$.

We cannot say precisely which homes, thus DIF is chosen over LAND, but a home is in general a markable location.

- He was shot in the [abdomen]$_{DIF}$.

This is not a geolocatable mention, but it is clearly a location, so we mark it as DIF.
Diffuse areas with extents that are unknown to an oracle (see Section 3.7.6), or whose scope is extraordinarily large should not be marked.

- America has many homeless on her urban streets.

Streets is not tagged.

- There are many hungry children in the world.

The scale of *world* is simply too large, so it is not tagged.
Components of a structure are generally not tagged.

- The railings on the [Midtown bridge]$_{LAND}$ are rusty and need to be painted.

Small scale locations and components are not tagged, so *railings* is not marked.

- My keys are in my pocket.

Pocket is not tagged.

3.4. Adjectival Tag

We use the Adjectival (ADJ) tag for named GPEs or LOCs with adjectival or possessive usage. ADJ is also used for demonyms. If a name and its possessive mark (*e.g.,*'s in English) are separate whitespace-delimited tokens, then only the name should be tagged.

3.4.1. Basic examples

- The northern [Ohio]$_{ADJ}$ company laid off 5,000 workers.

- [Muscat's]$_{ADJ}$ largest market, the Muttrah Souq, reopened Saturday, three days after it was flooded.

- [Lebanese]$_{ADJ}$ President Emile Lahoud, meanwhile, was more forthright, strongly denouncing [Washington's]$_{ADJ}$ aggression.

- The dead from the storms included five [Indians]$_{ADJ}$ and four [Pakistanis]$_{ADJ}$.

- The [Springfield]$_{ADJ}$, [Mass]$_{ADJ}$, [plant]$_{LAND}$ employs over 2,500 workers.

Note the compound ADJ use — here both *Springfield* and *Mass* are used as modifiers of plant.

- [Crete]$_{LAND}$ is very beautiful. The [European]$_{ADJ}$ [country]$_{LAND}$ is an island.

3.5. GAG Tag

A GAG (GPE acting as an AGent) tag is used to indicate a GPE that is not acting in a strong locational or territorial sense. However, unlike other tag categories, nominals and pronominals of GAGs are not marked.[3]

3.5.1. Basic examples

- Qatar$_{GAG}$ has economic links with Israel$_{GAG}$ without maintaining diplomatic relations, ...

- Qatar$_{GAG}$ is a very important country on this issue.

As a nominal *country* is not tagged.

3.6. Directional qualifier (DIR)

Directional qualifiers indicate a spatial relationship including: cardinal direction, distance, proximity, adjacency, or areal containment. When the choice seems ambiguous, prefer SI over DIR.

3.6.1. Basic examples

- They need fresh water [in]$_{SI}$ the village [two miles]$_{DIR}$ [down]$_{DIR}$ the road.

- We travelled [north from]$_{DIR}$ Kuwait [to]$_{SI}$ the Iraqi port of Basra.

- The earthquake pushed lava into new [underground]$_{DIR}$ chambers.

- If you do somehow get away, God willing, follow the road [north]$_{DIR}$ for [17 kilometers]$_{DIR}$, then [west]$_{DIR}$ for [ten]$_{DIR}$.

3.7. General Principles

In this section we expand on certain situations that commonly occur, and we attempt to help resolve ambiguous situations that may arise due to an entity being interpretable as multiple types.

[3]To do so, would require marking innumerable mentions of words such as *city*, *country*, *nation*, and *it*.

3.7.1. GPEs

Geo-political Entities (GPEs) are not tagged as GPEs for this task. A location phrase tag must be selected instead. If the GPE is a people group, it should be tagged as an ADJ. GPEs used in a political sense are not tagged as LAND, and should be tagged as a GAG instead. See Section 3.7.2 below.

- The [Soviets]$_{ADJ}$ objected to the [American]$_{ADJ}$ proposal.

"Soviets," meaning "Soviet people," is tagged. If the GPE is used in a physical location sense it should be tagged as a LAND. If the GPE is used as an organizational / governmental sense, tag it as a GAG.

- John McCain (Rep., [Ariz.]$_{GAG}$) is a former POW.

3.7.2. Priority

If a mention can be interpreted by more than one of the LAND, ADJ, or GAG tags, resolve them in the following order. Prefer ADJ over GAG, and GAG over LAND.

- The [German]$_{ADJ}$ government signed a treaty with [France]$_{GAG}$ [in]$_{SI}$ [Cairo]$_{LAND}$, [Egypt]$_{LAND}$.

3.7.3. Pronouns

We will tag pronominal mentions of Landmarks, and the tag should match the entity with which it is coreferent. Relative adverbs (*e.g.,* **where**) may also be tagged. GAGs will only be marked for names; nominals and pronominals that are coreferent with GAGs will not be annotated (see Section 3.5).

- She walked to the [Washington Monument]$_{LAND}$. [It]$_{LAND}$ is near the [Lincoln Memorial]$_{LAND}$.

- [Ellicott City]$_{LAND}$, [Maryland]$_{LAND}$, [where]$_{LAND}$ he lives, experienced several floods.

3.7.4. Nominals

We tag nominal mentions of Landmarks. However, we do not tag nominals of GAGs. Choosing otherwise would require tagging of innumerable mentions of "country", "state", "nation", etc...

3.7.5. Possessive markers

Possessive locations should be marked as ADJ. A "'s" should not be tagged if it is a separate token, because the possessive marker ('s) is not the head.

- [Atlanta]$_{ADJ}$'s [Buckhead]$_{LAND}$ district ...

Consider these three sentences, each of which describes where a person resides:

1. Chicago accountant Scott Taccetta has been hired by Google.

2. Chicago 's Scott Taccetta has been hired by Google.

3. Scott Taccetta of Chicago has been hired by Google.

	setimes	*voa*	*radioliberty*	*voa*
documents	92	16	31	16
sentences	2193	305	2731	248
tokens	56489	7079	46639	5859
tagged	5235	561	3402	476
LAND	1760	224	1415	198
ADJ	1381	139	391	54
DIF	234	18	515	35
GAG	796	52	221	69
SI	912	113	803	108
DIR	152	15	57	12
O	51254	6518	43237	5383

Table 1: Statistics for English/Russian annotations.

All three sentences are describing this person as being a resident or person from Chicago. Example #1 and example #2 are clear cases of ADJ tags. Example #3 looks syntactically like a SI (*of*) and a LAND. However, for such constructions, do not tag the "of" as an SI (see Section 3.1). In examples #1 and #2 **Chicago** should be tagged as ADJ, and in example #3 it should be tagged as a LAND.

3.7.6. Oracles

Sometimes there is a location mentioned, but salient information is unavailable to a third-person reader. For example, a news report might interview a disaster victim who reports that "*her home* was destroyed." In this case, the victim knows where her home is and she could geolocate it, but a reader of the news report cannot. For this sentence *home* should be tagged as DIF because the location is known to someone (*i.e.,* the victim). If the author of a text, or a person described in a text has such information, we call them an oracle.

4. Annotation

We are releasing a dataset with annotations based on the guidelines above in both English (63k tokens) and Russian (52k tokens).[4] Annotations were created using the Dragonfly annotation tool (Lin et al., 2018; Costello et al., 2020), which we have also used for non-native speaker annotation for various tasks in the LORELEI program. A zoomed-in display of the tagging interface is shown in Figure 1.

The released English data are annotated news articles from two sources: the Southeast European Times and Voice of America. The Russian data is also from two sources: RadioLiberty and Voice of America. A statistical summary of the dataset are given in Table 1.

4.1. Russian

A native English speaker who is fluent in Russian annotated the Russian texts; a sample image is shown in Figure 2.

Our experience tagging Russian showed that our English guidelines were generally transferable. We did observed that the genitive case in Russian loses distinction between

[Jamaica's]$_{ADJ}$ [resorts]$_{DIF}$

[resorts]$_{DIF}$ of [Jamaica]$_{LAND}$

[4] https://github.com/iscoe/lrec20-locphr/

Tag	*Precision*	*Recall*	F_1
LAND	89.1	87.8	88.5
DIF	74.8	47.2	57.9
GAG	98.7	85.3	91.4
ADJ	90.7	97.0	93.8
SI	83.0	85.9	84.4
DIR	95.5	71.2	81.6

Table 2: Examining subtype agreement by two annotators.

We also noted that the lack of an indefinite article requires understanding the context to distinguish when a Landmark or a Diffuse tag should be used. For example, in English we can differentiate "I slept in **the** house" from "I slept in **a** house" based on the article. But in Russian both sentences would be glossed as "I slept in house," and contextual information is required to determine the correct tag assignment.

4.2. Inter-Annotator Agreement

It is possible for two people to disagree about the proper tagging of a sentence. Such differences can arise from different interpretations of the task, insufficiently defined guidelines, or a problem that is intrinsically subjective. We identified many problems with our guidelines or with annotation understanding by pairing two annotators and having them adjudicate annotations one-on-one. One issue that we identified is due to the fact that GPEs and Organizations often have inseparable physical presence and institutional characteristics. This tension between multiple roles is one source of ambiguity for annotators.

It is important to give annotators practice on the task before using the annotations they produce. We improved performance on the task by having annotators review several documents that they each tagged and then compare decisions. This helped to identify systematic misunderstandings and mistakes due to task complexity. We show a different view of the *Dragonfly* interface in Figure 3 that allows annotators to compare and correct results.

Across four trained annotators, we found a mean Cohen's Kappa of 0.67 for the six-way tag assignment in English. We interpret this to mean that the task is challenging for humans to perform. If we just look at agreement for taggable extents (and not determining the types), the Kappa score rose to 0.88.

In Table 2 we compare precision, recall, and F_1 scores using doubly-annotated data. We treated one annotator as the gold standard and a different annotator as a system prediction. For most tags there is a high degree of consistency; the Diffuse tag exhibits the most disagreement.

5. Experimental Results

We conducted experiments on English-annotated data using two different sequential taggers. The data used in these experiments is a separate collection of 60k words of news data which we are not able to publicly release. The first tagger is a statistical system named *SVMLattice* (Mayfield et al., 2003) that jointly estimates tag transition and emission probabilities in a Hidden Markov framework. In addition to a baseline using a variety of lexical and subword features,

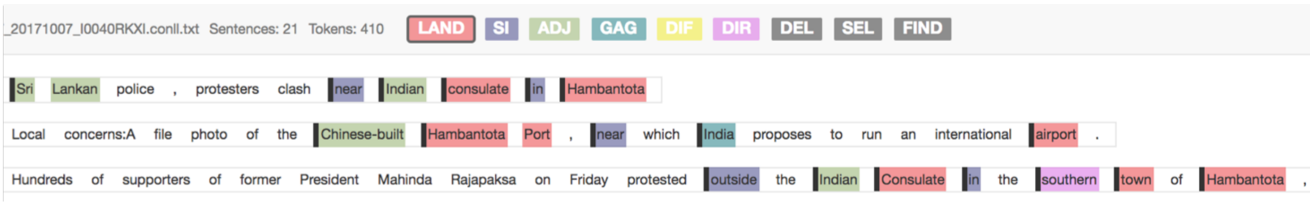

Figure 1: Using the *Dragonfly* annotation interface to annotate a document from IL 10 Set E. The six location tags are distinguished using different colors.

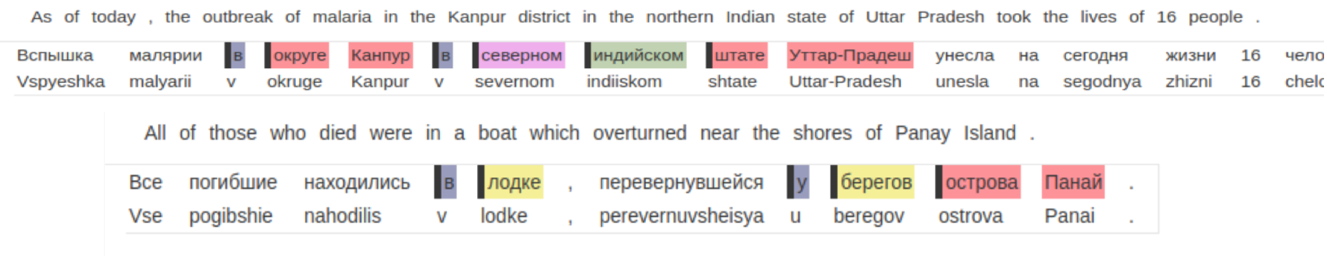

Figure 2: Two Russian sentences annotated using the *Dragonfly* tool.

System	*Precision*	*Recall*	F_1
Statistical baseline	74.0	48.7	58.8
Statistical w/ Brown	74.6	57.1	64.7
Statistical w/ GloVe	74.4	58.6	65.5
Neural (GloVe)	73.2	61.8	67.0

Table 3: Comparing baseline statistical and neural taggers on six-way location phrase tagging.

we also add Brown Cluster (Brown et al., 1992) prefix features or 300-dimensional GloVe embeddings (Pennington et al., 2014) to support generalization. *SVMLattice* works by using a linear support vector machine to compute a margin, which is transformed into a probability using a sigmoid function. Finally Viterbi decoding is used to select the best tag sequence.
The second tagger we utilized is by Liu (2018). It is a neural network, Bi-LSTM model coupled with a conditional random field (CRF) in the style of Lample *et al.* (2016). With the Liu tagger we use 300-dimensional GloVe embeddings (Pennington et al., 2014).

5.1. Baselines

In Table 3 we compare results for the two types of tagger. We find with the addition of either Brown Clusters or GloVe embeddings, our statistical system performs almost on par with a Bi-LSTM+CRF model.

5.2. Conflated Classes

Because we observed that inter-annotator consistency was difficult for some tags (*e.g.,* Diffuse), and we observed that choosing the proper tag is difficult for some entity mentions, we conducted an experiment where we mapped the six tags into a reduced two tag set. We merged LAND, DIF, GAG, and ADJ into one class, chiefly representing the named entities and noun phrases. The SI and DIR

System	*Precision*	*Recall*	F_1	$\Delta\ F_1$
Statistical	82.5	70.5	76.0	10.5
Neural	73.7	69.5	71.5	4.5

Table 4: Experiments using two meta-classes instead of six separate classes.

tags were combined into a second equivalence class, which mainly fits the tokens that are prepositions (*e.g.,* to/from), verbs (*e.g.,* flew/surrounded), and descriptive adjectives (*e.g.,* southwest, nearby). The results are shown in Table 4. Scores are materially elevated for the simpler two-class case. And surprisingly, the statistical model outperforms the neural model under this condition. We believe that the two class model is useful despite bearing less information, for example, entities can still be geotagged and presented to a mission planner in a disaster relief scenario.

5.3. Learning Curve

We also conducted a learning curve study where we trained models on successively larger portions of our training data. Our results are in Figure 4. F_1 scores quickly rise once 10,000 training words are available. Growth continues even at our limit of 50,000 words, so additional gains are likely if more training data is used.

6. Summary

In this work we addressed the problem of detecting named and non-named location phrases in text. Key accomplishments from this effort include: developing a sequential tagging approach to the problem; crafting annotation guidelines; building annotated datasets; and, conducting experiments in automated detection of location phrases using both statistical and neural taggers. While the task appears more difficult than coarse-grained named entity tagging, we believe that the performance obtained from this early work is

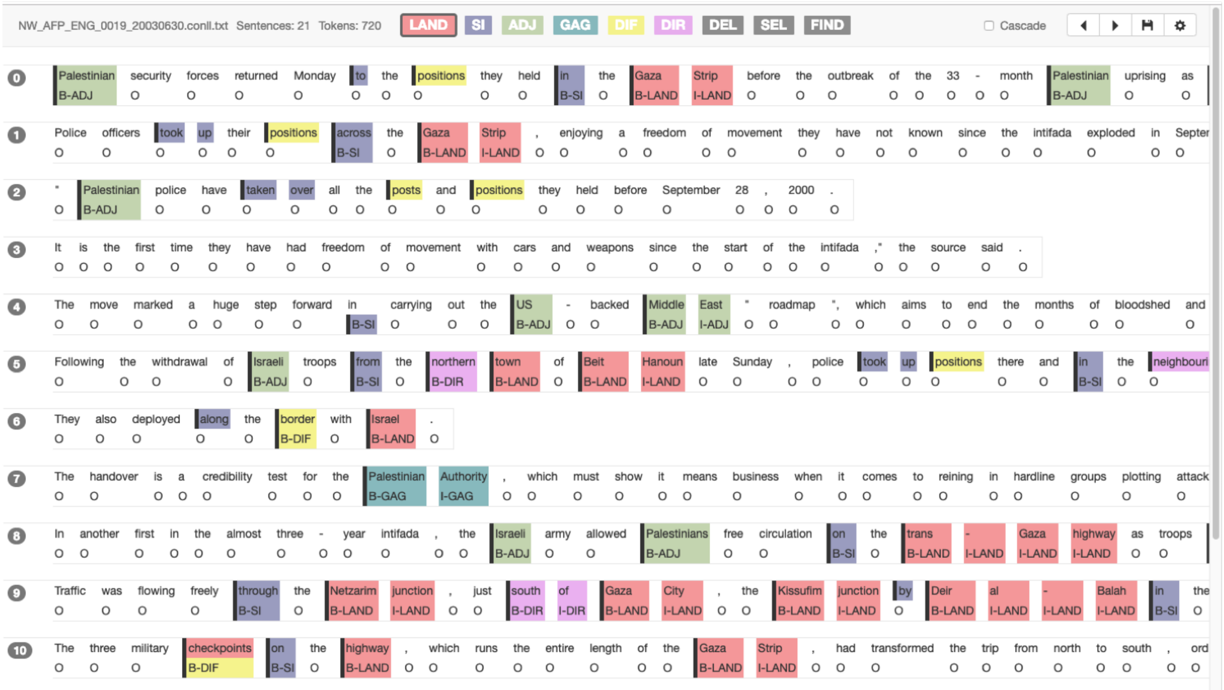

Figure 3: Comparison of annotations by two annotators. Each sentence is displayed on two lines. The first line shows the original text, tagged with the six colored types. The second line corresponds token-by-token with the first, and displays the tags assigned by the second annotator. This view allows the annotators to quickly identify difficult cases and correct mistakes.

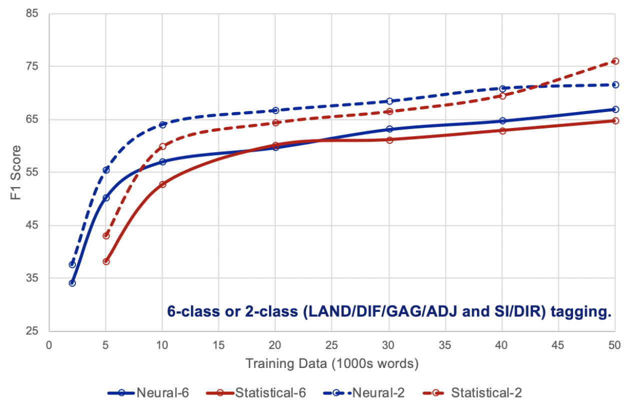

Figure 4: Improvement in performance on location phrase task with additional data.

promising. When trained on a greater quantity of data, we suspect these models can be further improved.
One gap in this research is geotagging non-named locations. The research community has been studying geolocating named locations, but non-named locations have so far received less attention.

Acknowledgment

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contracts No. HR0011-15-C-0115 and No. HR0011-16-C-0102. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

7. Bibliographic References

Blaylock, N., Allen, J. F., de Beaumont, W., Galescu, L., and Jung, H. (2012). Street-level geolocation from natural language descriptions. *Traitement Automatique des Langues*, 53(2):117–205.

Broda, B., Marcińczuk, M., Maziarz, M., Radziszewski, A., and Wardyński, A. (2012). KPWr: Towards a free corpus of polish. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3218–3222, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Brown, P. F., Della Pietra, V. J., deSouza, P. V., Lai, J. C., and Mercer, R. L. (1992). Class-based *n*-gram models of natural language. *Computational Linguistics*, 18(4):467–480.

Christianson, C., Duncan, J., and Onyshkevych, B. (2018). Overview of the DARPA LORELEI program. *Machine Translation*, 32(1-2):3–9.

Costello, C., Anderson, S., Bishop, C., Mayfield, J., and McNamee, P. (2020). Dragonfly: Advances in non-speaker annotation for low resource languages. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC'20)*, Marseille, France, May. European Language Resources Association (ELRA).

Hassani, K. and Lee, W.-S. (2017). Disambiguating spatial prepositions using deep convolutional networks. In

Proceedings of the 31st AAAI Conference on Artificial Intelligence.

Kolomiyets, O., Kordjamshidi, P., Moens, M.-F., and Bethard, S. (2013). SemEval-2013 task 3: Spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 255–262, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Kordjamshidi, P., Van Otterlo, M., and Moens, M.-F. (2011). Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process.*, 8(3):4:1–4:36, December.

Kordjamshidi, P., Bethard, S., and Moens, M.-F. (2012). Semeval-2012 task 3: Spatial role labeling. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, SemEval '12, pages 365–373, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California, June. Association for Computational Linguistics.

Lin, Y., Costello, C., Zhang, B., Lu, D., Ji, H., Mayfield, J., and McNamee, P. (2018). Platforms for non-speakers annotating names in any language. In *Proceedings of ACL 2018, System Demonstrations*, pages 1–6, Melbourne, Australia, July. Association for Computational Linguistics.

Liu, L., Shang, J., Ren, X., Xu, F., Gui, H., Peng, J., and Han, J. (2018). Empower sequence labeling with task-aware neural language model. In *Proceedings of AAAI 2018*.

Marcinczuk, M., Oleksy, M., and Wieczorek, J. (2016). Preliminary study on automatic recognition of spatial expressions in polish texts. In *Text, Speech, and Dialogue - 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings*, pages 154–162.

Mayfield, J., McNamee, P., Piatko, C., and Pearce, C. (2003). Lattice-based tagging using support vector machines. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, CIKM '03, pages 303–308, New York, NY, USA. ACM.

Moncla, L., Gaio, M., Nogueras-Iso, J., and Mustière, S. (2016). Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. *International Journal of Geographical Information Science*, 30(6):1137–1160.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October. Association for Computational Linguistics.

Radke, M., Das, P., Stock, K., and Jones, C. B. (2019). Detecting the Geospatialness of Prepositions from Natural Language Text (Short Paper). In Sabine Timpf, et al., editors, *14th International Conference on Spatial Information Theory (COSIT 2019)*, volume 142 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:8, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Skoumas, G., Pfoser, D., Kyrillidis, A., and Sellis, T. (2016). Location estimation using crowdsourced spatial relations. *ACM Trans. Spatial Algorithms Syst.*, 2(2):5:1–5:23, June.