

Identifying Motion Entities in Natural Language and A Case Study for Named Entity Recognition

Ngoc Phuoc An Vo, Irene Manotas, Vadim Sheinin, Octavian Popescu

IBM Research

Yorktown Heights, NY, US

{ngoc.phuoc.an.vo, irene.manotas}@ibm.com
{vadims, o.popescu}@us.ibm.com

Abstract

Motion recognition is one of the basic cognitive capabilities of many life forms, however, detecting and understanding motion in text is not a trivial task. In addition, identifying motion entities in natural language is not only challenging but also beneficial for a better natural language understanding. In this paper, we present a Motion Entity Tagging (MET) model to identify entities in motion in a text using the Literal-Motion-in-Text (LiMiT) dataset for training and evaluating the model. Then we propose a new method to split clauses and phrases from complex and long motion sentences to improve the performance of our MET model. We also present results showing that motion features, in particular, entity in motion benefits the Named-Entity Recognition (NER) task. Finally, we present an analysis for the special co-occurrence relation between the person category in NER and animate entities in motion, which significantly improves the classification performance for the person category in NER.

1 Introduction

In semantic approaches such as Semantic Role Labeling (SRL) and Abstract Meaning Representation (AMR), semantic features can have peculiarities of a particular language (Zhu et al., 2019). In contrast, the irreducible semantic core considered by the Natural Semantic Metalanguage (NSM) approach, builds a kind of universal mini-language where its core components, the *semantic primes*¹, are universal to all languages (Goddard, 1997; Wierzbicka, 1980). *Semantic primes*, are defined as universal core meanings. The hypothesis in NSM is that meaning could be reconstructed into basic elements or semantic primes. Then, any complex meaning can be decomposed without circularity and without residue into a combination of discrete other meanings, the *semantic primes*, among which is the meaning of motion represented by the MOVE semantic prime.

In natural language, motion can describe different movement types (e.g. rotational, transactional, internal, etc), it can relate to changes in a concept or abstract object when it is *figurative motion* (e.g. "*Her voice twisted from incredulity to astonishment*"), or it can describe the movement of physical object(s) when it is *literal motion* (e.g. "*The player twisted his leg before kicking the ball*"). Thus, motion analysis and detection can be challenging because of the different ways in which motion can be used in natural language (Beavers et al., 2010). Overall, motion is a linguistic primitive that allows us to express complicated events more concisely, because of its features motion has been considered extensively in theoretical linguistic analysis. Many linguists have agreed that motion is a semantic fundamental (Talmy, 1985; Goddard, 1997; Beavers et al., 2010), thus identifying motion in text and its features is important, and so it is to investigate how motion features could help improve Natural Language Processing (NLP) tasks.

Although analysis of spatial features in natural language has considered and investigated semantics related with motion events, their elements, including encoding schemes (Pustejovsky and Moszkowicz, 2008; Pustejovsky and Yocum, 2013; Pustejovsky et al., 2015; Lee et al., 2018), to the best of our

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

¹<https://intranet.secure.griffith.edu.au/schools-departments/natural-semantic-metalanguage/what-is-nsm/semantic-primes>

knowledge, there has not been an investigation of how motion detection in natural language, and motion features, might impact NLP tasks. To start this investigation, we selected Named-Entity Recognition (NER), a tagging task that identifies and classifies named entities in a text into predefined semantic types such as person and organization, among others. NER is a core task of many NLP systems, especially information extraction and question answering. This paper explores and shows results of a motion entity tagging model and how leveraging motion entities positively impact models for the NER task.

This paper is organized as follows, Section 2 presents characteristics of motion in natural language and related work; Section 3 presents the LiMiT dataset, created as a resource to build motion in text classification and motion entity tagger models, which is publicly released to the community. A model for motion entity tagging is presented in Section 4; In Section 5, we present the results of our investigation about the impact of the entity in motion feature in the NER task; Conclusions and future work are presented in Section 6.

2 Background

In natural language, motion instances can be present as *literal* or *figurative* (Talmy, 2000; Matlock, 2004; Pustejovsky and Yocum, 2013). Motion describing the movement of a physical object is considered *literal* motion (e.g. "*William lifted the weight*"), while *figurative* motion is when there is no direct occurrence of motion (e.g. "*A gentle slope brought us in one hour to the main summit*"), or not occurrence of physical motion at all (e.g. "*The range runs east and west*"). Although figurative motion can sometimes indirectly indicate motion of physical objects, this is not always the case. Therefore, we decided to focus our analysis on literal motion which directly indicates motion events related with physical entities in motion.

Four elements of a motion event are identified by contemporary linguistics: (a) an object, also called the 'figure', 'theme', or 'participant', (b) moving along (c) a 'path' (with 'source' and 'goal') (d) with respect to another reference-object, also called 'ground' (Talmy, 1985). Nevertheless, not all of motion elements are present in all instances of motion in natural language. For instance, verbs such as '*wiggle*' or '*wave*' describe the manner of motion of an object but do not require traversal of a path. Similarly, other verbs such as '*dance*' may sometimes not specify a motion path. For this reason, we decided to identify, as one core element of the motion event, the entities in motion (i.e. figure or participant), and start our analysis of motion in natural language by focusing on the occurrence of physical motion in text.

To the best of our knowledge, we are the first to investigate how motion features impact an NLP task. However, there have been related work around motion data sets and motion classification tasks in natural language. In the context of capturing spatial information in text, MotionBank (Pustejovsky and Yocum, 2013) was a corpus created with fictive and literal motion sentences, including information about different features related with location and non-location, one of which is motion. Then the entity-in-motion annotation was proposed as part of ISO-Space, as the MOVER entity associated with the MOVELINK relation annotation (Pustejovsky, 2017). Recently, and with the aim to better understand the semantics of spatial information in text, (Egorova et al., 2018) presented a fiction motion data set and classification model for fictive motion sentences. Although previous motion data sets could have been considered for our study, we were not able to find any of the mentioned data sets publicly available online yet. Therefore, we decided to use the recently released Literal-Motion-in-Text (LiMiT) dataset (Manotas et al., 2020) of literal motion sentences.

3 The Literal Motion in Text Dataset

The Literal-Motion-in-Text (LiMiT) dataset (Manotas et al., 2020) was created as a resource to identify both literal motion in text and the entities in motion from a natural language expression describing the motion of physical entities. With models built with LiMiT dataset, it is also possible to investigate the impact of motion features in natural language tasks. LiMiT was publicly released² to promote further research in this area. The size of the LiMiT dataset is 24,559 sentences: 15,346 sentences describing literal motion, and 9,213 sentences describing figurative or no motion events. In this section we briefly

²https://github.com/ilmgut/limit_dataset

MET Model	Test	Prec.(%)	Recall(%)	F1(%)
BERT	All	74.00	68.04	70.90
	NAC	71.18	73.64	72.39
	e-books	77.51	62.61	69.27

Table 1: Performance of BERT-based MET model.

describe the characteristics of the LiMiT dataset, which was used for training the model to identify physical entities in motion in a sentence.

3.1 Data Sources and Sentence Extraction

The LiMiT data set is composed of sentences drawn from two main sources: fiction e-books and novels, collected from online websites, and sentences from video descriptions extracted from the Net Activity Captions (NAC) dataset (Krishna et al., 2017). Text documents were first processed with a parser to identify the main text, then pre-processing was applied on the resulting sentences to remove special characters and other unnecessary formatting. After identifying all sentences from a document with a NLP parser, sentences were filtered by using information from their linguistic features, obtained from an NLP parser, and considering a list of verbs of motion. The final output for each document was a set of potential motion sentences i.e., sentences containing a *motion* verb and therefore possibly describing one or multiple motion events.

3.2 Annotation of Motion Sentences

After potential motion sentences were identified, crowdsourced annotator workers annotated sentences in the Appen Platform³. In the annotation task, workers were asked to both annotate whether a sentence described the movement of a physical entity (e.g., animate or inanimate). For literal motion sentences, workers identified the animate and inanimate entities in motion in the given sentence. To guarantee the quality of the labels and work provided by annotation workers, participants had to both complete a quiz before each annotation job, and in every work page workers were given a random test row. Test rows in the quiz and in work pages look the same as rows in the annotation job, but they had predefined golden answers hidden from the worker that allow to measure the quality of the responses given by annotators. After all annotation judgements were completed, the Fleiss Kappa measure of agreement was computed. The computed Fleiss kappa for the combined sentences was 0.71, which indicates a substantial agreement among the workers participating in the annotation job.

Sentences in the LiMiT dataset cover a vocabulary of 29,021 unique words, with an average sentence length of 98.81 characters. For literal motion sentences in LiMiT, there is at least one annotated entity in motion and about 40% of sentences have two or more entities in motion.

4 Motion in Text Baseline Model

To investigate how motion features might possibly impact NLP tasks we need a model to identify motion entities in natural language text. In this section we first describe the baseline model we trained to identify entities in motion in a sentence. Then we describe an algorithm designed to extract short sentences from complex long sentences to improve the performance of the model output.

Data. For each of the sentences in the LiMiT dataset, we transformed the text and its labeled entities in motion to the Inside-outside-beginning (IOB) tagging format. For example, for the tokens sequence [Thomas, Williams, jumped, over, the, rocks], the corresponding IOB tag sequence with MET labels was [B-MOT, I-MOT, O, O, O, O].

³<https://appen.com/>

MET Model	Prec.(%)	Recall(%)	F1(%)	Support
BERT (Sent)	85.83	62.80	72.54	164
BERT (CnP)	80.613	73.78	76.83	164

Table 2: Motion Entity Tagging on sentences (Sent) and on clauses and phrases (CnP).

Model. We built a Motion Entity Tagging (MET) model to identify the physical entities in motion in a sentence. For an input sentence, the MET model predicts and tags the tokens that belong to entities in motion in the text. For example, for the sentence "John took the book from the shelf", the MET model will predict *John* and *book* as the entities in motion. We used Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019; Vaswani et al., 2017) to build a MET classifier. Specifically, we used the BERT_{BASE} (whole word masking) language model, which takes a maximum 512 input word piece token sequence $X = [x_1; x_2; \dots; x_T]$ and uses a $L = 12$ layer transformer network (with 12 attention heads and 768 embedding dimensions) to output a sequence of contextualized token representations $H_L = [h_1^L; h_2^L; \dots; h_T^L]$. For the BERT-NER model, we used the representation of the first sub-token as the input to the token-level classifier over the MET label set. We fine-tuned the model using five epochs, with a learning-rate of 0.1, and using a 256 maximum sequence length.

Results. Table 1 shows the results of the evaluation for the MET classifier built using the BERT-NER model. We used sequence labeling evaluation considering exact matches, calculating macro recall, precision and F1-scores. On literal motion sentences from NAC, the recall is better than on sentences extracted from fiction e-books; but precision is better on fiction e-books literal motion sentences than on NAC sentences. We think this is because sentences from NAC are often shorter and contain simpler language constructs for describing motion than sentences from fiction e-books. Also, sentences from fiction e-books have a more complex use of natural language (i.e. more descriptive and verbose) than NAC sentences, which sometimes make it more challenging to tag the correct entities in motion. In some cases, although the correct motion entity is partially predicted, a good prediction is not considered as such due to a mismatch to the surface string e.g., when 'bowling ball' is predicted as the entity in motion but 'ball' was the labeled golden entity in motion.

A Method for Improving Predictions on Complex Motion Sentences. The following sentence shows an example of a motion sentence where there are multiple entities in motion highlighted in bold:

*"For five, ten minutes afterwards **Frieda** continued to hold the floor, and then in the midst of an account of a party given at the Johnson home **she** had suddenly stopped talking and thrown herself down on the floor, tucking a **sofa pillow** under her blonde head."*

Analyzing the MET classifier results we noticed that the model was not always able to identify all entities in motion in complex sentences. For instance, for the sentence shown above, the MET model identifies as entities in motion only *Frieda*. We observed that recall could be improved if a long sentence would be split into separate short sentences, clauses or phrases. Then, to tackle this problem and to improve the performance of our entity in motion prediction, we designed an algorithm to identify and split clauses and phrases (CnP) from a sentence.

In order to split a sentence we used the following steps: First, analyze the sentence punctuation. Most of the times, the period (e.g. ".") is a clear delimiter of independent segments in text, but we also considered text between matching commas, parentheses and semicolons as separated phrases; Second, consider the sentence dependencies. A phrase introduced by a sentence complementizer is independent. Examples include relative clauses, and other phrases such as "in the case of", temporal conjunctions like "before", "after", etc; Third, consider the Verb Phrase (VP) domain. Since most of the times semantic roles are given by the verb, determining the borders of the VP domains is instrumental into separating the sentence into short phrases. Determining the VP domain is of course difficult, but in our case even an approximate delimitation is beneficial. We used the verb valencies and prepositions affinity as learned

from GIGAWORD⁴ to determine the boundaries between tensed verbs. The steps above produce a series of short sentences, including clauses and phrases.

After short sentences were identified, we ran the MET classifier on the extracted clauses from a sentence, and consolidated all predicted motion entities for clauses in a sentence. We evaluated the results of our motion prediction over clauses and phrases (CnP) of sentences and compared them with the motion entity predictions over the original sentence for 100 examples. Table 2 shows the results of our evaluation comparing the model performance when predicting motion entities for the sentence as a whole BERT (Sent), and when predicting the entities in motion over the clauses/phrases in a sentence BERT (CnP). We can see that by loosing a little on precision, we can boost recall and F1 scores of the motion entity classifier and predict more motion entities over the identified clauses and phrases of long/complex motion sentences.

5 Case Study: Impact of Motion Entity Features on Named Entity Recognition Task

In this section we present a preliminary study about the proportion of motion sentences in well-known NLP datasets. Then, we present the case study showing the impact of entity in motion features on NER models.

5.1 Motion Sentences in NLP Datasets

Before investigating the impact of entity in motion features in NLP tasks, we conducted a study to analyze the proportion of potential motion sentences in datasets of some established NLP tasks. For this study, we selected some well-known NLP datasets for our motion in natural language analysis. For each selected dataset, we identified unique sentences and run our motion text classification model to identify the proportion of motion sentences in the dataset. The left side of Table 3 shows the proportion of motion sentences for the SQuAD 2.0 (Rajpurkar et al., 2018), SNLI (Bowman et al., 2015), SICK (Marelli et al., 2014), MSR Paraphrase (Dolan and Brockett, 2005), and CoNLL-2003 NER (Sang and De Meulder, 2003) datasets. From these results, we can see that overall datasets have between 7% – 53% of motion sentences, which make them candidates for analyzing how motion-related features could impact the performance of their related NLP tasks. NLP datasets, such as SNLI and SICK, having more than half of motion sentences, could be used to further study and show impact on motion features analysis.

5.2 Impact of Entity in Motion Features on NER models

In this section, we present the case study in which motion features benefit a particular task in NLP: Named Entity Recognition (NER). We investigated and present preliminary results of potential impact of motion features on the NER task. We selected NER for our case study because: 1) both NER and Motion Entity Tagging are sequence tagging tasks, and 2) the target of these tasks is identifying entities and labeling them. Thus, we hypothesized that motion features will benefit the NER task. Specifically, we also hypothesized a co-occurrence relation between the [PER] category in NER and identified motion entities in a sentence. For example, the sentence “*John kicked the ball*”, would have *John* annotated as [B-PER] as well as animate motion entity [B-MOT]. To understand how motion features can benefit the NER task, we purposely built basic models in which no sophisticated or advanced features (such as word embedding, semantic or syntactic features) are used or significantly contribute to the model performance. Section 5.2.1 describes the NER models and in section 5.3, we discuss co-occurrence between [PER] category in NER and identified motion entities.

5.2.1 Datasets and Experiments

Datasets We run experiments on the CoNLL-2003 NER dataset (Sang and De Meulder, 2003). This dataset has four categories [PER, LOC, ORG, MISC] in which [PER] is label for person names, [LOC] for locations, [ORG] for organizations, and [MISC] for other categories of entity. This dataset consists of train (14,987 sentences), validation (3,466 sentences), and test (3,684 sentence) sets.

⁴<https://catalog.ldc.upenn.edu/LDC2003T05>

Dataset	#Unique Sentences	#Motion Sentences	%Motion	%PER	%LOC	%ORG	%MISC
SQuAD 2.0 (dev)	6,411	910	14.2	NA	NA	NA	NA
SQuAD 2.0 (train)	93,768	6617	7.1	NA	NA	NA	NA
SNLI 1.0 (train)	629,518	331,897	52.7	NA	NA	NA	NA
SNLI 1.0 (dev)	13,138	6,964	53	NA	NA	NA	NA
SNLI 1.0 (test)	13,137	6,857	52.2	NA	NA	NA	NA
SICK (train)	5,034	2,581	51.3	NA	NA	NA	NA
SICK (test)	5,002	2,551	51	NA	NA	NA	NA
MSRPar (train)	7,923	1,419	17.9	NA	NA	NA	NA
MSRPar (test)	3,440	649	18.9	NA	NA	NA	NA
CoNLL-2003 Train	14,041	1,858	13.3	8.53	0.29	0.61	0.62
CoNLL-2003 Dev	3,250	525	16.2	10.68	0.89	1.02	0.89
CoNLL-2003 Test	3,453	346	10	6.23	0.81	0.29	0.49

Table 3: Motion Texts in NLP Datasets - Co-occurrence between NER Entities vs. Motion Entities.

Experiment 1. We built two simple NER models using bag-of-words representation and a typical multi-class classifier. We used this experiment to examine how motion features alone can benefit the performance of NER task.

Bag-of-Words and Classifier (BoWC). This is our baseline model. It takes in sequences of words and NER tags (labels). We used bag of words representation from scikit-learn⁵ to produce word vector \vec{w} for each word in the NER datasets. The word vectors and their associated NER tags are fed into the Gradient Boosting Classifier⁶ for training the classification model.

Bag-of-Words and Classifier with Motion Features (BoWC-M). This is the enhanced model from the baseline BoWC by integrating motion features. First we produced one-hot vectors \vec{m} for our three motion labels [B-MOT, I-MOT, O]. We run our motion tagger (presented in Section 4) to annotate motion label for each word in the NER dataset, so each word will have two labels, the motion entity label and the NER label (e.g., (*WORD MOTION_tag NER_tag*)). For example, for the sentence “*Peter kicked the ball*”, the labels for each word would be: (*Peter B-MOT B-PER*) (*kicked O O*) (*the O O*) (*ball B-MOT O*), where *Peter* and *ball* are entities in motion. We enhanced the BoWC model that at the output of every word vector, we concatenated each word vector \vec{w} with the one-hot vector derived from its associated motion label \vec{m} to obtain the fusion vector \vec{v} (see 1). The resulting fusion vector \vec{v} is used for the classification model training.

$$\vec{v} = \text{concatenate}(\vec{w}, \vec{m}) \quad (1)$$

Experiment 2. Since NER is a sequence labelling task, we built 2 other models using Conditional Random Fields (CRFs)⁷. We used this experiment to examine how motion features interact with other features and eventually how it can benefit the performance of NER task.

Conditional Random Fields (CRF). This model takes in basic features such as word, word spanning, is_uppercase, is_title, is_digit, and Part-of-Speech (POS) tag of the current word w , the previous word $w - 1$, and the next word $w + 1$.

Conditional Random Field with Motion Features (CRF-M). This is an enhanced model of CRF in which the motion labels are integrated as a new feature together with other basic features.

5.2.2 Evaluations

Table 4 presents and contrasts the performance between model BoWC vs. BoWC-M for CoNLL-2003 validation and test sets. The model BoWC-M with integrated motion features outperforms the baseline

⁵https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html

⁶<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>

⁷<https://sklearn-crfsuite.readthedocs.io/en/latest/>

Model BoWC			Model BoWC-M			
VALIDATION						
Precision	Recall	F1		Precision	Recall	F1
0.92	0.41	0.57	B-LOC	0.92	0.42	0.57
0.91	0.62	0.74	B-MISC	0.93	0.62	0.75
0.77	0.08	0.14	B-ORG	0.78	0.09	0.17
0.96	0.12	0.21	B-PER	0.78	0.26	0.39
0.85	0.60	0.70	I-LOC	0.83	0.56	0.67
0.83	0.40	0.54	I-MISC	0.83	0.39	0.53
0.73	0.20	0.31	I-ORG	0.75	0.19	0.31
0.89	0.07	0.12	I-PER	0.70	0.18	0.28
0.87	1.00	0.93	O	0.88	0.99	0.93
	0.87		Accuracy		0.88	
0.86	0.39	0.47	Macro Avg	0.82	0.41	0.51
0.87	0.87	0.83	Weighted Avg	0.87	0.88	0.85
TEST						
0.96	0.28	0.43	B-LOC	0.95	0.27	0.43
0.87	0.48	0.62	B-MISC	0.86	0.51	0.64
0.74	0.06	0.11	B-ORG	0.73	0.06	0.11
0.96	0.08	0.15	B-PER	0.74	0.17	0.28
0.73	0.55	0.63	I-LOC	0.79	0.44	0.57
0.71	0.48	0.57	I-MISC	0.74	0.48	0.58
0.75	0.18	0.29	I-ORG	0.74	0.19	0.30
0.77	0.02	0.05	I-PER	0.72	0.11	0.20
0.86	1.00	0.92	O	0.86	1.00	0.92
	0.86		Accuracy	0.86		
0.82	0.35	0.42	Macro Avg	0.79	0.36	0.45
0.85	0.86	0.81	Weighted Avg	0.85	0.86	0.82

Table 4: Model Performance BoWC vs. BoWC-M on CoNLL-2003 Datasets.

Model CRF			Model CRF-M			
VALIDATION						
Precision	Recall	F1		Precision	Recall	F1
0.91	0.88	0.90	B-LOC	0.92	0.88	0.90
0.93	0.84	0.88	B-MISC	0.92	0.84	0.88
0.85	0.81	0.83	B-ORG	0.84	0.81	0.83
0.90	0.91	0.90	B-PER	0.91	0.92	0.92
0.96	0.77	0.85	I-LOC	0.90	0.77	0.83
0.88	0.72	0.79	I-MISC	0.86	0.73	0.79
0.81	0.84	0.83	I-ORG	0.80	0.83	0.82
0.94	0.95	0.95	I-PER	0.95	0.95	0.95
0.99	1.00	0.99	O	0.99	1.00	0.99
	0.98		Accuracy		0.98	
0.91	0.86	0.88	Macro Avg	0.90	0.86	0.88
0.98	0.98	0.98	Weighted Avg	0.98	0.98	0.98
TEST						
0.86	0.81	0.83	B-LOC	0.85	0.81	0.83
0.81	0.76	0.78	B-MISC	0.82	0.76	0.79
0.76	0.73	0.75	B-ORG	0.76	0.73	0.75
0.83	0.86	0.84	B-PER	0.84	0.86	0.85
0.75	0.60	0.67	I-LOC	0.77	0.64	0.70
0.69	0.67	0.68	I-MISC	0.69	0.67	0.68
0.67	0.74	0.70	I-ORG	0.66	0.75	0.70
0.87	0.95	0.91	I-PER	0.88	0.95	0.91
0.99	0.99	0.99	O	0.99	0.99	0.99
	0.96		Accuracy		0.96	
0.80	0.79	0.79	Macro Avg	0.81	0.79	0.80
0.96	0.96	0.96	Weighted Avg	0.96	0.96	0.96

Table 5: Model Performance CRF vs. CRF-M on CoNLL-2003 Datasets.

model BoWC on all measures for most of the label classes. The results prove that using a basic level of language representation like bag-of-words, integrated with motion features, benefits the NER task.

In contrast, Table 5 shows the same overall performance between CRF and CRF-M models but with a slight improvement for category [PER] when motion features are integrated.

5.3 Co-occurrence Relation between Category [PER] and Animate Motion Entity

In this section, we discuss the co-occurrence relation between category [PER] in NER and animate motion entity in texts describing the motion of physical entities. The category [PER] represents person names in natural language, and motion entity represents a physical entity in motion. In literal motion, a physical motion entity can be either animate or inanimate. And an animate motion entity may be a person(s). Therefore, there should be a co-occurrence relation between category [PER] and animate motion entities. For example in the sentence “*John kicked the ball*”, *John* is an animate entity and *ball* is an inanimate entity. Table 4 shows the performance improvements of model BoWC-M (enhanced motion features) for most of categories in NER, however, the improvements for category [PER] (B-PER and I-PER) on validation and test sets are most significant. As shown in Table 5, although the model CRF-M does not return better overall result, it slightly improves the performance for category [PER]. The following examples illustrate the co-occurrence relation between category [PER] and animate motion entities. We present the data in format (*WORD MOTION-label NER-label*) for better reading.

- (**Zhirinovsky B-MOT B-PER**) (visited O O) (Iraq O B-LOC) (twice O O) (in O O) (1995 O O)
- (**Kenny B-MOT B-PER**) (Dalglish I-MOT I-PER) (spoke O O) (on O O) (Thursday O O) (of O O) (his O O) (sadness O O) (at O O) (leaving O O) (Blackburn O B-ORG)
- (Thousands O O) (of O O) (**demonstrators B-MOT O**) (have O O) (marched O O) (through O O) (London O B-GEO)
- (**They B-MOT O**) (marched O O) (from O O) (the O O) (Houses O O) (of O O) (Parliament O O)

The first 2 examples show the co-occurrence of [PER] entity and animate motion entity in text which helps our model to identify category [PER] better. However, the animate motion entities also can co-occur with general entities, such as *demonstrators* or *they* in the 3rd and 4th examples. In this case, the animate motion entity can help to identify general entities in other applications. The right side of Table 3 shows the co-occurrence between each NER category and our animate motion entity in the CoNLL-2003 datasets. We can see a strong co-occurrence relation between [PER] entities and animate motion entities which results in more significant classification improvement for [PER] category by the models BoWC-M and CRF-M.

6 Conclusions and Future Research

We presented a motion entity tagger model to identify entities in motion in text, and presented a method to improve entity in motion predictions for complex, long motion sentences. Furthermore, we present our initial results towards analyzing how features of literal motion, specifically motion entities, can impact NLP tasks. Specifically, we investigated the impact of motion features on the NER task and proved our hypothesis that motion features can improve the performance of the NER model. Finally, we show the study of co-occurrence between category [PER] in NER and animate motion entity, which indicates how the motion feature benefits the classification results for the [PER] category. For future research, we plan to continue investigating how motion features can benefit other more complex NLP tasks such as textual inference. We also study approaches for integrating motion features to obtain better performance in different tasks which helps with better understanding of natural language overall.

References

- John Beavers, Beth Levin, and Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics*, 46(2):331–377.
- Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. The SNLI corpus.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Association for Computational Linguistics.

- Bill Dolan and Chris Brockett. 2005. Automatically constructing a corpus of sentential paraphrases. In *Third International Workshop on Paraphrasing (IWP2005)*. Asia Federation of Natural Language Processing.
- Ekaterina Egorova, Ludovic Moncla, Mauro Gaio, Christophe Claramunt, and Ross S. Purves. 2018. Fictive motion extraction and classification. *International Journal of Geographical Information Science*, 32(11):2247–2271.
- Cliff Goddard. 1997. The semantics of coming and going. *Pragmatics; Vol 7, No 2 (1997)*, 7, 01.
- Ranjay Krishna, Kenji Hata, Frederic Ren, Li Fei-Fei, and Juan Carlos Niebles. 2017. Dense-captioning events in videos. In *International Conference on Computer Vision (ICCV)*.
- Kiyong Lee, James Pustejovsky, and Harry Bunt. 2018. The revision of ISO-space, focused on the movement link. In *Proceedings 14th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 35–44. Association for Computational Linguistics.
- Irene Manotas, Ngoc Phuoc An Vo, and Vadim Sheinin. 2020. LiMiT: The Literal Motion in Text Dataset. In *Findings of EMNLP*. Association for Computational Linguistics.
- Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, Roberto Zamparelli, et al. 2014. A SICK cure for the evaluation of compositional distributional semantic models. In *LREC*, pages 216–223.
- Teenie Matlock. 2004. The conceptual motivation of fictive motion.
- James Pustejovsky and Jessica L. Moszkowicz. 2008. Integrating motion predicate classes with spatial and temporal annotations. In *Coling: Companion volume: Posters*, pages 95–98.
- James Pustejovsky and Zachary Yocum. 2013. Capturing motion in ISO-SpaceBank. In *Proceedings of the Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*. Association for Computational Linguistics.
- James Pustejovsky, Parisa Kordjamshidi, Marie-Francine Moens, Aaron Levine, Seth Dworman, and Zachary Yocum. 2015. SemEval-2015 task 8: SpaceEval. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval)*. Association for Computational Linguistics.
- James Pustejovsky. 2017. ISO-Space: Annotating static and dynamic spatial information. In *Handbook of linguistic annotation*, pages 989–1024. Springer.
- Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don't know: Unanswerable questions for squad. *arXiv preprint arXiv:1806.03822*.
- Erik F Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.
- Leonard Talmy. 1985. *Language Typology and Syntactic Description*, volume 3. Cambridge University Press, 2 edition.
- Leonard Talmy. 2000. Toward a cognitive semantics, vol. 1: Concept structuring systems. language, speech, and communication. *MIT Press, Language Arts and Disciplines*, 10.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS*, page 6000–6010. Curran Associates Inc.
- A. Wierzbicka. 1980. *Lingua Mentalis: The Semantics of Natural Language*. Academic Press.
- Huaiyu Zhu, Yunyao Li, and Laura Chiticariu. 2019. Towards universal semantic representation. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 177–181. Association for Computational Linguistics.