LiMiT: The Literal Motion in Text Dataset

Irene Manotas   **Ngoc Phuoc An Vo**   **Vadim Sheinin**
Thomas J. Watson Research Center
IBM Research AI, Yorktown Heights, NY, US
{irene.manotas, ngoc.phuoc.an.vo}@ibm.com, vadims@us.ibm.com

Abstract

Motion recognition is one of the basic cognitive capabilities of many life forms, yet identifying motion of physical entities in natural language have not been explored extensively and empirically. We present the Literal-Motion-in-Text (LiMiT) dataset, a large human-annotated collection of English text sentences describing physical occurrence of motion, with annotated physical entities in motion. We describe the annotation process for the dataset, analyze its scale and diversity, and report results of several baseline models. We also present future research directions and applications of the LiMiT dataset and share it publicly as a new resource for the research community.

1 Introduction

Natural Language Understanding (NLU) tasks involving semantic and pragmatic levels in Natural Language Processing (NLP) are challenging mainly due to the pervasive ambiguity of language and the subtly different perceptions humans have of the meaning of word, phrase, and sentence (Navigli, 2018).

The premise in the Natural Semantic Metalanguage (NSM) (Wierzbicka, 1980) approach to semantics is that any complex meaning can be decomposed without circularity and without residue into a combination of discrete other meanings called *semantic primes*. Unlike other semantic approaches such as Semantic Role Labeling (SRL) (Màrquez et al., 2008) and Abstract Meaning Representation (AMR) (Banarescu et al., 2013), where semantic features can have peculiarities of a particular language (Zhu et al., 2019), the irreducible semantic core considered by NSM and composed by the *semantic primes*[1], composes a universal

---

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

mini-language, i.e., the primes are universal to many, if not all, languages. The hypothesis is that meaning could be reconstructed into basic elements or semantic primes, one of which is MOVE representing the meaning of motion. At a cognitive level, almost all life forms beyond plants and simple organisms recognize motion. In natural language, motion can describe different movement types (e.g., rotational, transactional, internal), it can relate to changes in a concept or abstract thing when it is *figurative motion* (e.g., *"Her voice twisted from incredulity to astonishment"*), or it can describe the movement of a physical object when it is *literal motion* (e.g., *"The player twisted his leg before kicking the ball"*), which makes motion analysis and detection challenging.

Because motion is a linguistic primitive (Wierzbicka, 1996) that allows us to express complicated events more concisely, it has been considered extensively in theoretical linguistic analysis. Many linguists have agreed that motion is a semantic fundamental (Talmy, 1985; Goddard, 1997), thus identifying motion in text and its features could improve NLU tasks. Because of the nature of motion and its importance in natural language, we selected it as the first semantic prime to explore. However, and to the best of our knowledge, there is no publicly available dataset for motion detection in natural language that enables researchers to empirically investigate and understand the implications of motion in natural language.

This paper describes the LiMiT dataset, a large and annotated dataset of literal and no literal motion sentences in English. We share the dataset publicly and show results on classification models built to detect physical motion in text and models to identify the physical entities in motion. We describe applications of the LiMiT dataset, and delineate future research directions using the dataset.

2 Background

This section describes related work of motion in natural language, including motion elements or features, the MOVE semantic prime, and previous attempts to create a motion corpus.

2.1 Motion in Natural Language

Motion in natural language can be present as *literal* or *figurative* (Talmy, 2000; Roberts, 2009; Pustejovsky and Yocum, 2013). When motion involves the movement of a physical object, it is considered *literal* motion (e.g., *"The boy ran in the park"*), while *figurative* motion is when there is no physical occurrence of motion (e.g., *"The fence goes from the plateau to the valley"*). We are interested in the *literal* description of motion events, i.e., those produced by a physical object or entity.

2.1.1 Motion Elements

Contemporary linguistics identifies four elements of a basic motion event: (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). However, there exist instances where not all of the motion elements are present in natural language. For example, different verbs of motion, such as *wiggle* or *wave*, describe the manner of motion of an object but do not imply traversal of a path, and some other verbs such as *dance* may not specify a path. We focused our attention on the occurrence of physical motion in text, and the identification of physical entities in motion (i.e., figure or participant) as the main motion element.

2.2 The MOVE Semantic Prime

The meaning of *motion*, or *movement*, is represented by the semantic prime MOVE in NSM, and its description goes deeper than a 'change of place' (Wierzbicka, 1980). The primitive MOVE represents a meaning that is neutral concerning the distinction between 'internal motion' (e.g., *"The girl waved at the boy"*) and translational motion' (e.g., *"The train traveled from NYC to Newark."*) (Goddard, 1997). Thus, the semantic prime MOVE appears in sentences where there is no change in place such as the following ones: *"I see something is moving"*, or *"My lips trembled"*, where the place is not necessarily relevant for the movement.

There are challenges associated with the identification of motion in natural language. As with many other semantic primes, the MOVE semantic prime might be expressed by variant forms in different combinatorial or positional contexts (known as *allolexy* in NSM), and common terms that are equivalents of the MOVE semantic prime can be prone to polysemy. Thus, the identification of the equivalent terms for MOVE in natural language is not a straightforward task.

2.3 Related Work

Previous work introduced some guidelines for creating a textual inference corpus that included literal and figurative motion in a text (Roberts, 2009). The study by (Roberts et al., 2010) described the creation of the UDT-MotionEvent dataset, including 2, 500 sentences, where motion events and their structure were encoded using the FrameNet[2] annotation standard. Another motion-related corpus is MotionBank (Pustejovsky and Yocum, 2013). In MotionBank, both fictive and literal motion instances were considered, including several motion elements related to location (e.g., place, path) and non-location features (e.g., spatial entity, event, spatial signal, measure, and relationships).

Previously, the Spatial Role Labeling (SpRL) (Kordjamshidi et al., 2011) task[3] at Semantic Evaluation (SemEval) (Pustejovsky et al., 2015) included the identification of trajectors (e.g., entities in motion), landmark, spatial indicator, motion indicator (e.g., verb), path, direction, distance, and spatial relation. Later the SpaceEval (SpaceEval) task was proposed in SemEval as an extension of the SpRL task. In SpaceEval, the main task was to extract and classify spatial information in text such as toponyms, spatial nominals, location, and movement along paths using as the annotation the ISO-Space (ISO-Space) standard and the SpaceBank corpus (Pustejovsky and Yocum, 2013; Pustejovsky et al., 2015).

Although previous corpora consider the description of literal motion and its features, including the entity in motion (i.e. figure or trajector), our focus is more on capturing the physical occurrence of motion only. Therefore, we provide a larger and diverse set of data examples for literal

---

[2] https://framenet.icsi.berkeley.edu/fndrupal/
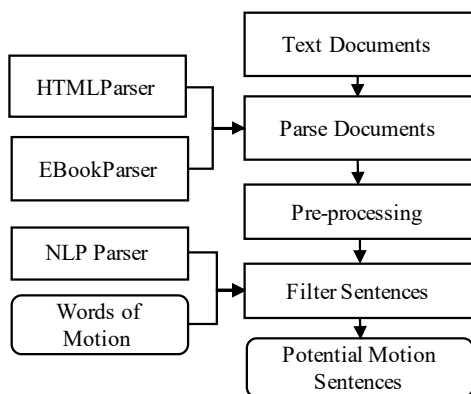[3] https://www.cs.york.ac.uk/semeval-2013/task3/
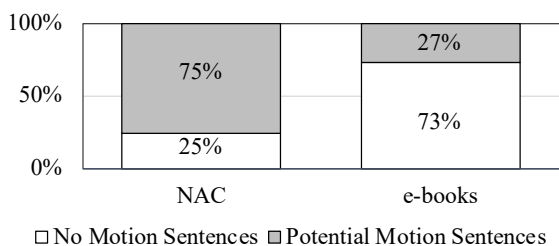
Figure 1: Data processing for text documents.

Figure 2: Potential motion sentences by data source.

motion instances in natural language with their entities in motion. For example, for SpaceEval only four motion verbs (e.g., arrive, leave, drive, and walk) were considered for motion signal identification, compared to $1,200$ motion verbs that we considered a potential signal to motion sentences in the LiMiT dataset (see Section 3). Also, the size (number of sentences) of the LiMiT dataset, is 13 times larger than the size of previous corpora. The larger size in our corpus was possible thanks to the crowdsourcing strategy we employed. The main differences between previous work and our work are: (1) Our dataset contains positive and negative examples of literal motion of physical objects; (2) Our dataset includes labeled entities in motion, i.e., subject(s) or object(s) in motion for motion sentences; (3) Our dataset contains around $15,000$ motion sentences, in which around 48% of sentences have more than one labeled entity in motion; (4) Our dataset sentences were drawn from two different sources including fiction e-books and video captions; (5) We release the dataset publicly to enable the research and applications of literal motion using the LiMiT Dataset. Section 6 describes how the LiMiT dataset could be used in different tasks to investigate further the semantics of motion and its impact on other NLP tasks.

3 Annotation Methodology

In this section we describe the methodology used for the crowdsourcing annotation of physical entities in motion. We describe the annotation job that allows workers to understand and participate in the annotation tasks i.e., identifying whether sentences describe motion of physical entities and annotating entities in motion in sentences with physical occurrence of motion. We studied different annotation task designs to collect judgments from crowdsourced workers. The analysis of different annotation task designs allowed us to select the design that provided the best quality of annotation data and achieve a good agreement among annotation workers.

3.1 Data for Annotation

To build our dataset of English sentences with the physical occurrence of motion, we used sentences extracted from two main sources: fiction e-books and novels in English from online websites, and sentences from the Net Activity Captions (NAC) dataset (Krishna et al., 2017). Text documents were pre-processed as shown in Figure 1. For selecting online e-books, we created a list of 207 authors, downloaded their e-book titles, and filtered those categorized as fiction or novel and available in the English language. Selected fiction e-books were out of copyright notice in U.S., and only e-books of the 19th century or after were selected. We collected a total of $1,169$ fiction e-books to extract sentences from. We extracted $2,204,880$ sentences from the collected e-books and $54,429$ sentences from the NAC dataset from which we took one caption sentence per video in the dataset.

Filtering Potential Motion Sentences. We define a potential motion sentence as one containing a *motion* verb. To identify and extract sentences describing motion, we built a filtering list of motion verbs that included verbs that could be used for describing the motion of physical objects. The filtering list of motion verbs was extracted from the list of 235 motion verbs (Levin, 1993), plus additional 965 verbs that we manually identified as verbs used in the description of physical motion events. For instance, verbs such as *patrol, transfer*, and *sink* were not included in the Verbs of Motion classification by (Levin, 1993), but we included

Motion Type	Example	Example Type
Literal motion	*"The girl jumped over the chair."*	Positive
Fictive motion	*"The street light threw its shadow on the road."*	Negative
Uncertain motion	*"If the birds fly away, I will open the window."*	Negative
Command motion	*"please go and bring it!, he said"*	Negative
Imaginative motion	*"I dreamed about an eagle flying close to us."*	Negative

Table 1: Motion types: positive and negative examples.

them in our motion verbs list to filter out potential motion sentences. Figure 2 shows the proportion of potential motion sentences versus no motion sentences that we found for each data source. For online fiction and novel e-books, around 27% of sentences had a motion verb from our filtering list, while for NAC dataset, there were around 75% potential motion sentences in the dataset.

3.2 Annotation Platform and Jobs

We used the Appen Platform[4] to annotate the sentences in our dataset.

Annotator workers. We collected annotations from both internal and external annotation workers. Internal workers were company volunteers including the three authors of this paper and additional 13 workers. External workers were preselected from the group of workers with higher experience in annotation jobs within the platform (Level 3) and with the highest performance in our annotation task, which in total added up to 250 workers. We ensured workers were paid a fair wage by paying each worker the hourly rate according to the annotation platform guidelines and considering the estimated annotation time for a sentence.

Annotation Job. An annotation job includes several pages of work with annotation rows according to an annotation task design. A row in our annotation task included a text representing a sentence, and a set of questions about the sentence to collect labels. The number of rows, or sentences, in the annotation job and the size of a work page defined the number of annotation work pages available for a worker. For every sentence available for annotation, we collected a total of nine judgments for each sentence. All annotation jobs included guidelines for the annotation task that described the label classes and positive and nega-

Figure 3: Annotation Job Design I. Annotation of physical entities in motion.

tive examples of the physical occurrence of motion in text. Positive examples showed sentences describing the literal motion of physical entities. In contrast, negative examples show workers examples of sentences describing fictive/apparent motion, uncertain motion, command, or imaginative motion as shown in Table 1.

3.2.1 Motion in Text Annotation Tasks

Here we present two different annotation task designs we studied for the annotation of physical occurrence of motion in text.

Design I: Identify Physical Entities in Motion in the text. Figure 3 shows the basic annotation task design, which included the following two general questions:

1. Given a sentence, we asked the worker to answer whether or not the presented text described the motion of a physical entity (e.g., a person, a vehicle).

2. When the answer to the first question was yes, we asked the worker to identify and

---

4[4]https://appen.com/

Step 1: Review the following text

The man stops and grabs a lamp post in the yard and turns towards the camera.

Is there a moving entity (e.g., animals, persons, or other objects) in the sentence shown above? (required)
- ● Yes
- ○ No

Step 2: INANIMATE entity -- Highlight text and assign the 'inanimate_entity_in_motion' class. Please highlight all INANIMATE entities in motion. If there are multiple INANIMATE entities in motion please highlight all of them.

If there is not an inanimate entity in motion in the text, please highlight another ANIMATED entity in motion in the text and assign the class 'animated_entity_in_motion' to the text (Step 3 below).

`inanimate_entity_in_motion`

Remember to highlight all the characters of the word for the 'animated' or 'inanimated' entity in motion, and not just part of it, as it is shown in the example below.

Find 123 Main Street ✗   Find 123 Main Street ✗   Find 123 Main Street ✓

Step 3: ANIMATE entity -- Highlight text and assign the 'animate_entity_in_motion' class: Please highlight all ANIMATE entities in motion. If there are multiple ANIMATE entities in motion please highlight all of them.

`animate_entity_in_motion`

Term Definitions

1. inanimate_entitiy_in_motion

Physical objects that are in motion but are not alive or have no life.

Examples
- He took **the book** from the shelf.
- The **oak beam** gradually tilted away from the wall.

2. animate_entitiy_in_motion

Physical objects that are in motion and have life or are alive.

Examples
- Eddie bends her **head**.
- **He** took the book from the shelf.
- **William** jumped off the cliff.

Figure 4: Annotation Job Design II. Annotation of physical *inanimate* and *animate* entities in motion.

highlight from the text the physical entity(ies) in motion.

We launched two annotation jobs using this annotation task design. Analyzing the annotation results, we noticed that workers generally agreed to annotate *animate* entities in motion and less often *inanimate* entities in motion. For instance, for the following sentence:

"**She** lifted her **wine glass**, but **each guest** laid a **hand** over theirs . . . "

where bold tokens indicate different entities in motion, *animate* entities include *she, each guest*, and *hand*, and *inanimate* entities include *wine glass*. Using Design I for the annotation task, we noticed that workers were not motivated to identify and select different physical entities in motion. Thus, the majority of labeled entities were *animate*, and few were *inanimate* entities in motion. To avoid bias towards *animate* entities only, we created a second annotation task design to encourage workers to annotate *inanimate* entities as well when possible.

Design II: Identify Physical Animate and Inanimate Entities in Motion in the text. To improve the diversity of the type of labeled entities in motion in our dataset, we designed our second annotation task to ask workers to identify the physical occurrence of *inanimate* and *animate* entities

in motion. We described to workers the difference between physical *animate* entities (i.e., alive entities such as persons or animals) and *inanimate* entities (i.e., lifeless entities such as ball or stick). In this annotation task, for sentences identified as having physical occurrence of motion, we asked users first to look for *inanimate* entities in motion in the text, and then identify *animate* entities in motion as shown in Figure 4. This job design increased the average number of labeled entities per sentence: labeled inanimate entities increased 70%, and labeled animate entities increased about 17% compared to the initial design.

3.2.2 Annotations Evaluation

For every annotation job, we did two phases of annotation evaluation: First, every time a new worker participated in an annotation job (s)he was required to complete a quiz of the same size of a page of work; Second, workers that passed the quiz, by achieving a score of 80% or higher, were tested subsequently in every work page by a test row selected at random from the annotation job test set. Workers that failed the quiz were not allowed to continue with the annotation job. Test questions followed the same format as rows in the annotation task, and each worker's answer to a test question was validated by the platform with the provided truth answers. A work page in our annotation jobs had a size of five rows. Each row in a work page consisted of a sentence followed by two or more questions according to the annotation task design, as shown in Figures 3 and 4. Although using a small number of rows per work page entails creating more test rows for an annotation job, we noticed that using five rows per page allowed us to test workers more frequently and thus identify bad workers quicker in order to stop them. In total, the time for annotation of the dataset took approximately 42 non-continuous worker hours.

3.2.3 Judgment Processing

Once annotation jobs were finalized, i.e., the jobs collected all judgments for all sentences from workers, the collected data need to be postprocessed. Figure 5 shows the process we followed to analyze the judgments from the report of finalized annotation jobs. For every annotation job launched, we obtained an annotation report from the annotation platform with judgments from annotator workers for all the job sentences. Judgments for every sentence included a *motion label*,
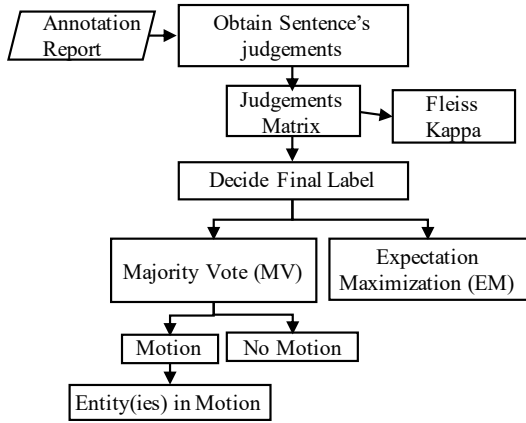
Figure 5: Judgments analysis and label extraction.

Data Source	Literal Motion	No Motion
Fiction e-books	7,582	6,951
NAC	7,764	2,262
Total	15,346	9,213

Table 2: Total number of sentences in the LiMiT dataset.

Labeled Entities	E-books	NAC	Total
One	4,557	3,489	8,046
Two	2,519	3,795	6,314
More than two	506	480	986
Total	7,582	7,764	15,346

Table 3: Frequency of sentences by number of labeled entities in the LiMiT dataset.

specifying whether a sentence described the motion of a physical entity, and an *entity_in_motion* label for the tokens that belong to the labeled entity(ies) in motion in the text if the worker identified the sentence as describing the movement of a physical object.

3.2.4 Label Selection from Data Judgements

To select the final labels for every sentence in our dataset, we analyzed the judgments collected for the *motion* label and the *entity_in_motion* label for every sentence using a majority vote strategy and an Expectation-Maximization (EM) strategy.

The majority vote strategy considered the majority label among judgments for both *motion* and the *entity_in_motion* labels. In cases where judgments had labeled the same entity but spanning different tokens at the same location in the text, we identified the shortest common span of the labeled entities to compute the majority vote for the entity. For instance, for the sentence *"Pencroft and Neb also pushed with the cart as far as the vein of coal . . . "* one judgment had an entity in motion label *the cart:34* and another judgment had *cart:38*, then *cart:38* was selected as the final entity in motion.

To improve over annotator workers' errors, we experimented with an EM strategy (Dawid and Skene, 1979; Raykar et al., 2009) to select the best motion label from annotation judgments. We found that using EM as a strategy to select final labels could improve the motion label selection. To measure how EM can improve the label selection performance, we used an expert labeled evaluation set of 640 sentences with ground truth motion labels provided by the authors of the paper. Before using EM, we obtained a F1-score of 0.747, and

after using EM to select the final motion labels, we obtained a F1-score of 0.783, about 4% improvement over the majority vote strategy.

3.3 Inter-Annotator Agreement

We computed the Fleiss Kappa measure of agreement among judgments collected from annotators workers for every sentence in the LiMiT dataset, as shown in section 3.2.3. The level of agreement among annotation workers was 0.66 Fleiss kappa for sentences from fiction e-books, and 0.77 Fleiss kappa for sentences from the NAC dataset. The computed Fleiss kappa for the combined sentences was 0.71, which indicates a substantial agreement among the workers participating in the annotation job.

4 The Literal Motion in Text Dataset

In this section, we present the Literal-Motion-in-Text (LiMiT)[5] dataset and its characteristics.

Table 2 shows the total number of sentences in the LiMiT dataset, which includes sentences extracted from e-books, with an average sentence length of 118 characters, and sentences from the NAC dataset of video descriptions, having an average sentence length of 70 characters. Having these two data sources results in a diversified set of 24, 559 positive and negative examples of literal

---

[5]https://github.com/ilmgut/limit_dataset

996

motion sentences in the dataset. All the sentences in the LiMiT dataset cover a vocabulary of $29,021$ unique words, with an average sentence length of 98.81 characters.

For sentences labeled as motion, i.e., describing the movement of a physical entity, Table 3 shows the frequency of sentences, for each data source, having a given number of labeled entities in motion. Although most sentences have only one entity in motion, there are about 40% of sentences with two or more entities in motion.

5 Motion in Text Baseline Models

For the train and test sets we used the data splits shown in table 4.

5.1 Motion in Text Classification

We built a model to classify a sentence as a literal motion sentence, i.e., a sentence describing the physical occurrence of motion such as *"John took the book from the shelf."*, or a sentence without literal motion such as *"A road goes through the desert."*.

Model. We used the Watson Natural Language Classifier (WNLC) on the IBM Cloud to build a text classification model for literal motion. The WNLC model include multiple Support Vector Machines (SVMs) and a Convolutional Neural Network (CNNs), using IBM's Deep Learning-as-a-Service (DLaaS)[6].

Setup. To build the WNLC motion classifier, we used the train and test data splits shown in Table 4, using only the labels for the presence of physical motion in each sentence: `"yes"` for literal motion sentences, and `"no"` for sentences not describing literal motion.

Results. With the motion in text classification model baseline we were able to achieve close to 78% precision, 77% recall, F1-score of 77%, and accuracy of 77% for identifying literal motion in text.

5.2 Entity in Motion Tagging

To identify the physical entities in motion in a sentence, we built a Motion Entity Tagging (MET) model. For an input sentence, the MET model will

Split	Data	Motion	No Motion	Total
	NAC	7,435	2,117	9,552
Train	e-books	7,225	6,782	14,007
	All	14,660	8,899	23,559
	NAC	329	145	474
Test	e-books	357	169	526
	All	686	314	1,000

Table 4: LiMiT Dataset Splits.

predict and tag the 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.

Models. We used two state of the art Deep Learning (DL) model architectures to built the MET classifiers. The first model was based on the Bidirectional Encoder Representations from Transformers (BERT) architecture (Devlin et al., 2019; Vaswani et al., 2017). The second model was built using a Bidirectional Long Short Term Memory (Bi-LSTM) model over Em-beddings from Language Models (ELMO) (Peters et al., 2018; Huang et al., 2015).

Setup. For each of the sentences in the LiMiT dataset, we transformed the text and its labeled entities in motion to the IOB tagging scheme, where each token was tagged with one of three unique token tags that belong to the MET label set: `I-mot` tag, for tokens inside the entity in motion; `O` tag, for tokens outside the entity in motion, and `B-mot` tag for tokens marking the beginning of the entity in motion. For instance, for the tokens sequence $[Daniel, Camil, ran, to, the, car]$, the IOB tag sequence with MET labels is $[B\text{-}mot, I\text{-}mot, O, O, O, O]$.

For the BERT-based MET model, we used the $BERT_{BASE}$ case (whole word masking) language model[7], which takes a maximum 512 input word piece token sequence $X = [x_1; x_2; ::::; x_T]$ and uses a $L = 12$ layer Transformer network (with 12 attention heads and 768 embedding dimensions) to output a sequence of contextualized token representations $H_L = [h_1^L; h_2^L; ::::; h_T^L]$. For the BERT-based MET model, we used the representation of

---

[6]https://www.ibm.com/watson/services/natural-language-classifier/

[7]Experiments with BERT's large model showed minor improvements.

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
BiLSTM	All	73.62	65.00	69.04
	NAC	70.54	71.30	70.91
	e-books	77.40	60.20	67.73

Table 5: Performance of Motion Entity Tagging (MET) Models.

Dataset	#UniqueSent	#MotSent	%Motion
SQuAD 2.0 (dev)	6,411	910	14.2
SQuAD 2.0 (train)	93,768	6617	7.1
SNLI 1.0 (train)	629,518	331,897	**52.7**
SNLI 1.0 (dev)	13,138	6,964	**53**
SNLI 1.0 (test)	13,137	6,857	**52.2**
SICK (train)	5,034	2,581	**51.3**
SICK (test)	5,002	2,551	**51**
MSRPar (train)	7,923	1,419	17.9
MSRPar (test)	3,440	649	18.9

Table 6: Motion Text Proportion in NLP Datasets.

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 5 shows the evaluation results for the MET classifiers using the BERT-based MET model and the ELMO+Bi-LSTM model described above. We used sequence labeling evaluation considering exact matches, calculating macro recall, precision, and F1-scores. We show the results of both models on different splits of the test data: considering all the data in the test set, considering only test sentences from the NAC data source, and considering test data from e-books only. On literal motion sentences from NAC, the recall is better than on e-books; but precision is better on e-Books literal motion sentences than on NAC sentences. We think this is because sentences from NAC are often shorter and contain fewer entities in motion than e-books sentences. Also, e-books sentences have a more complex use of natural language (i.e., more descriptive and verbose) than NAC sentences, making 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 reported due to a mismatch to the surface string, e.g., *'bowling ball'* was predicted as the entity in motion, but *'ball'* was the labeled golden entity in motion. Overall, the BERT model achieves better performance than the ELMO+Bi-LSTM model.

6 Applications And Research Directions

We envision that different NLP tasks can benefit from models built using the LiMiT dataset. This section presents a preliminary study on the proportion of motion sentences in some NLP tasks' datasets and future research directions. Also, we list some applications where the LiMiT dataset

could be used.

6.1 Motion in NLP Tasks

We conducted a preliminary study to compute the proportion of potential motion data in some well-known datasets of several established NLP tasks. For each dataset, we selected unique sentences and run our Motion Text Classification model to identify the proportion of motion sentences in the data. Table 6 shows the results of the motion in text classification for the SQuAD 2.0 (Rajpurkar et al., 2018), SNLI (Bowman et al., 2015), SICK (Marelli et al., 2014), and MSR Paraphrase (Dolan and Brockett, 2005) datasets. From the motion in text classification results, we see that some NLP datasets, such as SNLI and SICK, have more than 50% motion text and therefore can be used as candidates for analyzing how motion-related features could impact the performance of their related NLP tasks.

6.2 LiMiT Dataset Research Directions

The LiMiT dataset could be used to investigate how motion features can impact NLU tasks empirically. Because of the high proportion of motion text in NLP tasks such as text entailment and semantic textual similarity, we plan to investigate how motion features can improve their performance. Empirically investigating the impact of the MOVE semantic prime for NLU is another research direction. Also, the LiMiT dataset could be augmented with additional motion features such as path, place, to complement the description of the motion events described in the dataset and further analyze the impact of additional motion features on NLU.

6.3 Applications

Several applications in different domain areas can benefit from models built using the LiMiT dataset.

Computer Vision (CV) related tasks such text-to-scene or text-to-3D generation, where scene animations are generated based on text descriptions with implications in arts, education and robotics (Krishnaswamy and Pustejovsky, 2016; Chang et al., 2015) can use motion in text models to understand better whether motion is occurring and to identify the entities performing the motion action. Approaches to action recognition, text summarization, and others using multi-modal data, where recognition and prediction are built on several types of input data such as video, audio, and text, can benefit from motion in text models that provide fine-grained action understanding of the physical motion events in place (Li et al., 2019).

7 Conclusions and Future Research

We presented the Literal-Motion-in-Text (LiMiT) dataset and described the crowdsourcing methodology employed to collect and consolidate annotations of literal motion sentences that describe physical entities's motion. We also presented preliminary results of two classifiers: one for the task of identifying literal motion sentences, and another for the task of identifying physical entities in motion in sentences. We highlighted potential applications and future research directions for using the LiMiT dataset.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186. Association for Computational Linguistics.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

Angel Chang, Will Monroe, Manolis Savva, Christopher Potts, and Christopher D. Manning. 2015. Text to 3D scene generation with rich lexical grounding. In *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 53–62. Association for Computational Linguistics.

A. P. Dawid and A. M. Skene. 1979. Maximum likelihood estimation of observer error-rates using the em algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):20–28.

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.

Cliff Goddard. 1997. The semantics of coming and going. *Pragmatics; Vol 7, No 2 (1997)*, 7.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv e-prints*.

Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. 2011. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process.*, 8(3).

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)*.

Nikhil Krishnaswamy and James Pustejovsky. 2016. VoxSim: A visual platform for modeling motion language. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 54–58.

Beth Levin. 1993. *English Verb Class and Alternations: A Preliminary Investigation*. University of Chicago Press.

H. Li, J. Zhu, C. Ma, J. Zhang, and C. Zong. 2019. Read, watch, listen, and summarize: Multi-modal summarization for asynchronous text, image, audio and video. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):996–1009.

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.

Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. Special issue introduction: Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.

Roberto Navigli. 2018. Natural language understanding: Instructions for (present and future) use. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*, pages 5697–5702. International Joint Conferences on Artificial Intelligence Organization.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke" Zettlemoyer. 2018. Deep contextualized word representations. 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.

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 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.

Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don't know: Unanswerable questions for squad. *arXiv preprint arXiv:1806.03822*.

Vikas C. Raykar, Shipeng Yu, Linda H. Zhao, Anna Jerebko, Charles Florin, Gerardo Hermosillo Valadez, Luca Bogoni, and Linda Moy. 2009. Supervised learning from multiple experts: Whom to trust when everyone lies a bit. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 889–896. Association for Computing Machinery.

Kirk Roberts. 2009. Building an annotated textual inference corpus for motion and space. In *Proceedings of the 2009 Workshop on Applied Textual Inference*, TextInfer '09, page 48–51, USA. Association for Computational Linguistics.

Kirk Roberts, Srikanth Gullapalli, Cosmin Adrian Bejan, and Sanda Harabagiu. 2010. A linguistic resource for semantic parsing of motion events. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA).

Leonakd Talmy. 1985. *Language Typology and Syntactic Description*, 2 edition, volume 3. Cambridge University Press.

Leonard Talmy. 2000. Toward a cognitive semantics, vol. 1: Concept structuring systems. language, speech, and communication. *MIT Press, Language Arts and Disciplines*.

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.

Anna Wierzbicka. 1996. Semantics: Primes and universals.

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.