A Linguistic Resource for Semantic Parsing of Motion Events

Kirk Roberts, Srikanth Gullapalli, Cosmin Adrian Bejan, Sanda Harabagiu

Human Language Technology Research Institute
University of Texas at Dallas
Richardson TX 75080
{kirk,svg,ady,sanda}@hlt.utdallas.edu

Abstract

This paper presents a corpus of annotated *motion events* and their *event structure*. We consider motion events triggered by a set of motion evoking words and contemplate both literal and figurative interpretations of them. Figurative motion events are extracted into the same event structure but are marked as figurative in the corpus. To represent the event structure of motion, we use the FrameNet annotation standard, which encodes motion in over 70 frames. In order to acquire a diverse set of texts that are different from FrameNet's, we crawled blog and news feeds for five different domains: sports, newswire, finance, military, and gossip. We then annotated these documents with an automatic FrameNet parser. Its output was manually corrected to account for missing and incorrect frames as well as missing and incorrect frame elements. The corpus, UTD-MOTIONEVENT, may act as a resource for semantic parsing, detection of figurative language, spatial reasoning, and other tasks.

1. The Problem

As more spatial reasoning applications incorporate natural language text, the representation and extraction of *motion events* becomes increasingly more important. The concept of motion is a linguistic primitive that allows for the concise expression of a wide range of actions. Motion events can describe literal motion such as:

(1) The car *veered* into the outside lane.

Or they can describe figurative motion such as:

(2) The voice *veered* from exasperation to incredulity.

Motion events are grounded spatially and temporally. Spatial grounding is expressed by a variety of arguments (e.g., source, destination, distance, angle). Similarly, temporal grounding is expressed by several classes of relations (e.g., frequency, duration, time). Moreover, the interpretation of motion events encompasses several forms of disambiguation. For example, a car that veers, like in sentence (1), should be interpreted as an automobile, rather than a railroad car, cable car, etc. World knowledge dictates that the car was in an inside lane before the motion, it changed lanes with a speed in some expected range, and the entire motion took place in a few seconds. However, sentence (1) only implies the world knowledge required for the full interpretation of the event. Similarly, in the figurative expression of *veer* in sentence (2), world knowledge indicates that the voice carries some emotional state, which was initially exasperation and finally incredulity. Therefore, the motion indicates a change of emotional state.

Since motion events carry such expressive power in such a compact form and are so ingrained into language, cognitive semantics has given much attention to studying them (Talmy, 1996; Talmy, 2003; Johnson, 1987).

To be able to interpret motion events, several semantic resources are available. Three commonly used semantic sources are PropBank, NomBank, and FrameNet.

PropBank (Palmer et al., 2005) is a one million word corpus that has been annotated with argument structures for verbs.

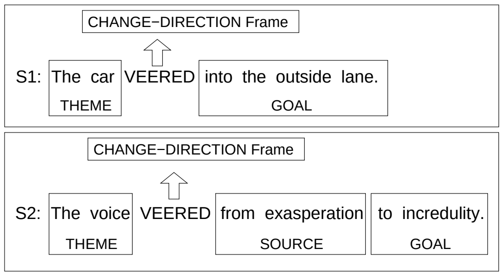

Figure 1: FrameNet parse of sentences (1) and (2).

NomBank (Meyers et al., 2004) is the equivalent for nouns. Given a specific verbal or nominal predicate, both resources assign numbered semantic arguments to the most common semantic types associated with the predicate. Typically, ARG0 is the agent, ARG1 is the theme or direct object, and ARG2 is the indirect object, benefactive, or instrument. Other argument types are specific to the predicate. For example, the verb *jump* has four specified roles: ARG1 is the thing jumping, ARG2 is the amount or distance the thing jumped, ARG3 is the starting point or state, and ARG4 is the ending point or state. Additionally, predicates can take adjunct-like arguments such as ARGM-LOC, which specifies a location.

Another resource annotated with semantic structures is FrameNet (Baker et al., 1998; Fillmore et al., 2002), which encodes lexico-semantic information according to *frame semantics* (Fillmore, 1982). FrameNet provides hundreds of schematic representations of objects, events, and scenarios. Each frame is triggered by a *lexical unit*, which may be almost any part of speech. Each frame defines a number of *frame elements* that reflect the common arguments of that frame. The advantage of using FrameNet for motion is its high degree of specificity. There are dozens of frames that describe motion, the most general of which, MOTION, enumerates 21 different elements, shown in Table 1. The semantic annotations associated with the event *veered* from (1) and (2) are shown in Figure 1.

The task of extracting semantic structures using resources such as PropBank, NomBank, and FrameNet is known as *semantic role labeling* (SRL). The goal of SRL is to iden-

Core	AREA, DIRECTION, DISTANCE, GOAL, PATH, SOURCE, THEME
Non-Core	CARRIER, CONTAINING_EVENT, DEGREE, DEPICTIVE, DURATION, FREQUENCY, ITERATION, MANNER, PATH_SHAPE, PLACE, PURPOSE, RESULT, SPEED, TIME

Table 1: FrameNet frame elements for the MOTION frame.

Domain	Source	Documents
Newswire	AP	500
Sports	Soccer by Ives	500
Gossip	TMZ	500
Financial	CNBC	500
Military	Danger Room	500

Table 2: Data sources used in corpus.

tify, for a given predicate, its semantically related phrases and the role each plays in the semantic structure. Semantic parsers (Gildea and Jurafsky, 2002; Xue and Palmer, 2004) have proven to improve the performance on a number of natural language applications such as question answering (Narayanan and Harabagiu, 2004), textual entailment (Tatu and Moldovan, 2005), and information extraction (Surdeanu et al., 2003).

To achieve higher accuracy for semantic parsing of motion events, we believe more information is required. First and foremost, more annotations are necessary. FrameNet 1.3 contains 365 annotated instances for the MOTION frame, only two of which have examples of DURATION and only five of which have DISTANCE (a core element). We believe this is not sufficient to maximize the capabilities of machine learning-based semantic parsing methods. Second, a greater variety of frame elements will aide in the identification of the semantic roles of motion events. For example, the *fly.v* lexical unit for MOTION only contains examples of objects we typically think of as capable of (non self-powered) flight such as arrows, balls, and bullets. A more diverse set of elements would allow for greater generalization of the potential participants in a motion frame, including elements used in a figurative motion frames. Third, the identification of literal and figurative events should help parsers distinguish situations where selectional constraints no longer apply, such as in sentence (2) above, where the SOURCE and GOAL arguments violate the locative selectional constraint. We later discuss the limited amount of figurative examples in FrameNet and compare this to the percentage of figurative examples we have found in our corpus. Fourth, a greater variety of training documents would aide in the creation of a more robust semantic parsing system. Notably, systems would benefit from more web documents, which are commonly used in natural language processing because web data is vast, cheap, and challenging.

To accomplish this, we have created a corpus of 2,500 documents with manually corrected FrameNet frames for motion events in order to provide more annotations for training data. These documents were drawn from a diverse set of sources available on the web to increase the variety of data (the data sources are shown in Table 2). Additionally, the motion events are annotated as literal or figurative so that a supervised system may be trained to recognize the figurative use of motion.

ARRANGING, ARRIVING, ATTACHING, AVOIDING, BOARD_VEHICLE, BODY_MOVEMENT, BRINGING, CAUSE_BEGIN_MOTION, CAUSE_CHANGE_OF_POSITION_ON_A_SCALE, CAUSE_EXPANSION, CAUSE_FLUIDIC_MOTION, CAUSE_IMPACT, CAUSE_MOTION, CAUSE_TO_AMALGAMATE, CAUSE_TO_FRAGMENT, CAUSE_TO_MOVE_IN_PLACE, CHANGE_DIRECTION, CHANGE_POSITION_ON_A_SCALE, CHANGE_POSTURE, COTHEME, DELIVERY, DEPARTING, DISEMBARKING, DISPERSAL, DODGING, ELUSIVE_GOAL, EMANATING, EMITTING, EMPTYING, ESCAPING, EVADING, EXCRETING, EXPANSION, FILLING, FLEEING, FLUIDIC_MOTION, FRICTION, GATHERING_UP, GETTING_UNDERWAY, GETTING_UP, GRINDING, HALT, HIT_TARGET, IMPACT, INTENTIONAL_TRAVERSING, LIGHT_MOVEMENT, MASS_MOTION, MOTION, MOTION_DIRECTIONAL, MOTION_NOISE, MOTION_SCENARIO, MOVING_IN_PLACE, OPERATE_VEHICLE, PATH_SHAPE, PATH_TRAVELLED, PLACING, QUITTING_A_PLACE, REDIRECTING, REMOVING, RESHAPING, RIDE_VEHICLE, ROADWAYS, SCOURING, SELF_MOTION, SENDING, SENT_ITEMS, SEPARATION, SETTING_OUT, SHOOT_PROJECTILES, SHOOTING_SCENARIO, SIDEREAL_APPEARANCE, SOUND_MOVEMENT, SOURCE_PATH_GOAL, SPEED, TRAVEL, TRAVERSING, USE_VEHICLE, VEHICLE

Table 3: FrameNet motion frames.

The remainder of this paper is organized as follows: Section 2 describes FrameNet's representation of motion events; Section 3 discusses literal and figurative motion events and our method for classifying them; Section 4 outlines the process of creating and annotating the corpus; Section 5 describes the created corpus, its current state, and some relevant statistics; and Section 6 discusses the potential uses for the corpus in semantic parsing and other fields.

2. FrameNet Motion Event Structure

Motion events in FrameNet are encoded in more than 70 frames. Table 3 contains some of these frames. Many more frames contain implicit motion information, such as the STATEMENT frame, which implies some movement of lips, tongue, and lungs (when talking) or fingers and wrists (when writing). However, we currently are only interested in events that explicitly express and describe motion and limit our study to those frames in Table 3.

While each motion frame in FrameNet may contain different elements, there is a strong consistency across many of the motion frames. Many contain a motion source (the SOURCE element type), a destination (GOAL), the object in motion (THEME), the agent that caused the motion (AGENT), the location of the motion (AREA and PLACE), the path of the motion (PATH), and several others. Table 4 lists more of the common motion elements and their descriptions. Each of the listed elements is contained in at least 15 motion frames from Table 3.

Each of these frames in FrameNet is triggered by a set of *lexical units*. For example, lexical units for the MOTION frame are shown in Table 5. The identification and disambiguation of lexical units forms an important (and difficult) first step in determining the FrameNet semantic parse for a

Element	Description
AGENT	Cause or propellant of motion
AREA	Location of motion when SOURCE and GOAL are undefined
COTHEME	A second moving object, following same or similar path as the THEME
DEGREE	Extent to which THEME crosses a boundary on route from SOURCE to GOAL
DEPICTIVE	Description of the state of the THEME during the motion
DIRECTION	Motion direction relative to the deitic center
DISTANCE	Extent of the motion (need not be numeric)
DURATION	Duration of time in which the motion takes place
GOAL	The motion's destination (need not be intentional)
MANNER	Description of manner in which the motion takes place
MEANS	Action taken that results in the motion
PATH	The complete or partial ground over which the THEME travels
PLACE	General area in which motion with specific SOURCE, PATH, and GOAL takes place
PURPOSE	State the AGENT or THEME wishes to achieve through the motion
REASON	State that leads to the motion
SOURCE	The motion's initial point
SPEED	Rate at which THEME travels
THEME	Object in motion
TIME	Time when motion occurs
VEHICLE	Mode of transportation during the motion

Table 4: Common elements for FrameNet motion frames.

blow.v, circle.v, coast.v, drift.v, float.v, fly.v, glide.v, go.v, meander.v, move.v, roll.v, slide.v, snake.v, soar.v, spiral.v, swerve.v, swing.v, travel.v, undulate.v, weave.v, wind.v, zigzag.v

Table 5: Lexical units for the FrameNet MOTION frame.

given sentence.

3. Literal and Figurative Uses of Motion

While one typically thinks of motion frames expressing literal motion, in many domains the figurative use of motion is far more common than the literal use. This is because motion is a linguistic primitive that allows us to express far more complicated events in a succinct manner. Take the following short sentence:

(3) The exam *drove* her mad.

This sentence does not express the literal motion of driving. Rather, it expresses a change of mental state and would be most properly represented by the frame CAUSE_EMOTION. However, the CAUSE_EMOTION frame does not contain the lexical unit *drive.v* and thus cannot be directly identified as such given the FrameNet specification. Using a lexical unit and typical syntactic structure from CAUSE_EMOTION, sentence (3) can be re-stated in a way that could be recognized by an automatic FrameNet parser:

(4) She was madly *offended* by the exam.

Clearly, sentence (3) is more succinct than sentence (4), which in part explains the common preference for the use of motion to express non-motion events.

In FrameNet's 365 sentences for the MOTION frame, less than ten percent [1] of the frames were used in the figurative sense. We shall see this is far from the case in our corpus. We seek an annotation distribution that is more realistic for commonly used domains. This is one of the main motivations behind choosing a diverse set of domains to form the corpus: we expect the distribution of literal and figurative motion to vary significantly from domain to domain.

One may argue that a motion corpus should include only literal events. Since the motion frames seem designed for physical motion, attempting to fit figurative motion into the frame semantics of motion sometimes produces awkward results. For instance, the most appropriate type for "*mad*" from sentence (3) is DISTANCE, as this sentence uses the same syntactic construction as:

(5) Bob *drove* her five miles.

Since "*mad*" does not fit the conventional selectional constraints for a distance, describing it as such may seem illogical.

We respond to this argument on two levels: empirically and theoretically. Empirically, not only does the FrameNet data contain figurative motion events (even if not very many), but many motion events are difficult to classify as literal or figurative, whether by human or machine. Consider the following three examples:

(6) Camera flashes *followed* him all the way to the entrance.
(7) The news *spread* around the room.
(8) Reyna *left* the team mid-season.

In all three cases, the THEME is not physically moving in the exact manner described, yet each displays many of the semantic properties of a motion event. This leads to our second argument.

Theoretically, if figurative motion displays many of the same syntactic and semantic properties, then we may still be able to perform (limited) spatial reasoning. Each of the figurative examples above displays some properties of literal motion. If we were performing textual entailment and were given sentence (7) as our background, a spatial reasoner could reject a hypothesis such as "*The news is next door*," even if it does seem like a nonsensical proposition. In (Roberts, 2009), we created a corpus for textual entailment that requires a system to perform spatial reasoning on figurative text. Part of our (empirical) motivation for including figurative frames is to better classify the motion events as literal or figurative.

We now discuss our methodology for annotating events as literal or figurative. If we were interested in detecting the use of metaphor, we would limit our definition of literal motion to physical motion with a concrete theme, source, destination, etc. But since we are interested in spatial reasoning, we define a literal motion as that which is best realized by a motion frame. In other words, if a motion frame

[1] This of course depends on your standard for literal and figurative. We used the same standard as for the documents in our corpus. Evaluation was performed on FrameNet 1.3.

Task	Score
Frame Disambiguation (Accuracy)	76.71
FE Boundary Precision (F1-measure)	79.80
FE Classification (Accuracy)	88.93

Table 6: (Bejan and Hathaway, 2007) scores for the SEMEVAL-2007 FrameNet task.

is being used because it is more succinct than another, non-motion, frame, then the motion is figurative. For example, sentence (3) is best realized by a CAUSE_EMOTION frame, while sentence (8) is about an individual leaving an organization and could be realized by the QUITTING frame [2]. Since sentence (6) has no non-motion alternative, and since sentence (7) concentrates more on the spread of information and less about talking (i.e., the CHATTING frame), they would both be considered literal in our corpus.

4. A Corpus of Semantically Annotated Motion Events

We created the UTD-MOTIONEVENT corpus in five major steps: (1) acquisition of data, (2) automatic FrameNet parsing, (3) manual correction of FrameNet parses, (4) annotation of literal/figurative frames, and (5) automatic consistency checking. Each of these processes is described below:

- STEP 1: *Acquisition of data*. The websites from Table 2 were crawled. Thousands of pages were downloaded for each source to allow us to skip pages without motion events or lexical units that might be linked to motion events. Each HTML page was stripped of its markup and a number of NLP tools were run across each new document. These tools include a tokenizer, sentence segmenter, part-of-speech tagger (Klein and Manning, 2003), named entity recognizer[3], and a full syntactic parser (Bikel, 2002). All these annotations were necessary for the next phase of data preparation.

- STEP 2: *Automatic FrameNet parsing*. We used the FrameNet parser described in (Bejan and Hathaway, 2007) to provide an initial semantic parse for the documents. This parser was the best performing system in the SENSEEVAL-3 evaluation and the second-best performing system in the SEMEVAL-2007 evaluation. Its results on the SEMEVAL-2007 FrameNet task are shown in Table 6.

  The reason for using an automatic parser as a first-pass is to aid the annotation process by pre-annotating easier lexical units and motion elements as well as alerting the annotators to common mistakes made by the system. For example, three common mistakes made by the parser are (i) disambiguation of multi-word lexical units (especially if one of the words in the lexical unit is a lexical unit for another frame), (ii) distinguishing between "*go*" being part of a future-tense predicate (e.g., "*going to eat tacos*") versus being a present progressive tensed verb (e.g., "*going to the store*"), and (iii) identification of less common element types (e.g., the PURPOSE and DEPICTIVE elements for the MOTION frame).

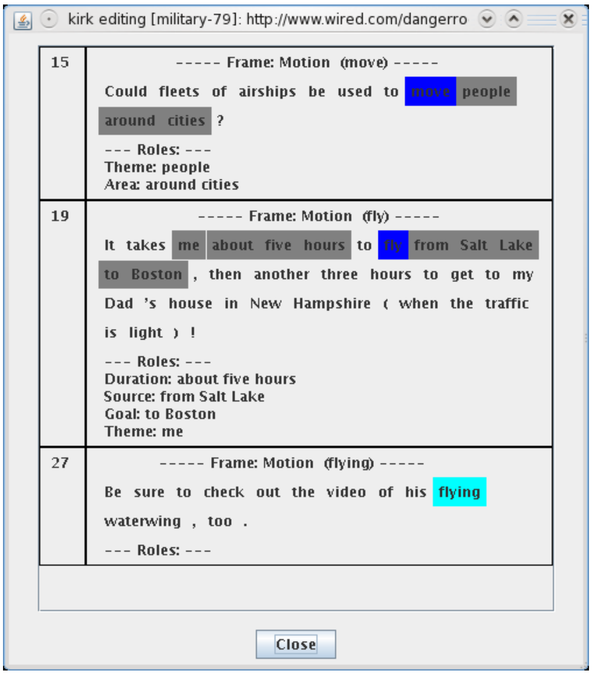

Figure 2: Annotating examples of the MOTION frame.

  In addition to automatically annotating frames, all lexical units for each motion frame were marked as potential frames. This allows the annotators to determine the frames that the automatic system missed. Not doing this would constrain the capability of any system trained on the UTD-MOTIONEVENT corpus to the performance of the automatic system for frame disambiguation.

- STEP 3: *Manual correction*. Two annotators (the first two authors) manually corrected the automatic FrameNet output. Annotation proceeded one frame at a time for each sub-corpus. This allowed annotators to maximize the consistency across the annotations and identify typical errors made by the semantic parser on different frame types.

  This was done using the purpose-built graphical user interface shown in Figure 2. This interface distinguished between frames annotated by the automatic parser and the potential frames marked on unannotated lexical units. It allows the annotator to select from the range of frame elements for a given frame and does not allow for frame element overlap.

- STEP 4: *Literal/Figurative Annotation*. The next stage of the annotation process was annotating whether a given frame was used in the literal or figurative sense. The annotators inspected all manually corrected frames in the document with a graphical user interface similar to the one shown in Figure 2. Annotation was performed for all frames in the document at once (instead of on a per-frame basis similar

[2] As of FrameNet 1.3, *leave.v* is not a lexical unit for the QUITTING frame. The online version of FrameNet has been updated to include *leave.v*, but is still without any annotated examples as of the time of this writing.

[3] http://www.surdeanu.name/mihai/bios/

to frame annotation). This allowed for additional context that proved helpful in quickly identifying whether the frame was being used in the literal sense.

- STEP 5: *Consistency Checking*. FrameNet enforces certain requirements on its annotations that may be checked automatically. Additionally, we employed several heuristics to help find incorrectly or inconsistently annotated frames or elements. These can be seen as an extension of (Scheffczyk and Ellsworth, 2006). Whereas they were more concerned about the structure of FrameNet itself, our consistency checks are targeted at resolving annotation errors and inconsistencies. Some of the automatic checks include:

  1. *Frame overlap*. No two frames may share the same lexical unit span, yet this is a common annotation mistake. For example, the verb "*move*" is a lexical unit for such motion frames as MOTION, CAUSE_MOTION, and CHANGE_POSITION_ON_A_SCALE, among others. It is common for an annotator to mark it as multiple frames given our one-frame-at-a-time approach, and the interface intentionally does not prevent them from doing so.
  2. *Role type inconsistency*. In the course of our annotating, the adverb "*quickly*" was annotated as both SPEED and MANNER. Similarly, "*wild*" was annotated as both MANNER and DEPICTIVE. While it is certainly possible for a textual expression to have different elements in separate events, it is a likely source of annotator inconsistency.
  3. *Relative clause labeling*. FrameNet has a special method for dealing with frames triggered inside relative clauses. According to (Ruppenhofer et al., 2006), when a target occurs inside a relative clause, both the constituent that contains the relativizer and its antecedent are assigned to separate frame elements with the same label. For example, given the sentence:

     (9) Everyone that *left* was noticed.

     Both "*Everyone*" and "*that*" are marked as separate THEME elements within the DEPARTING event triggered by "*left*". It can easily and automatically be checked that when frame elements are followed by relative words such as *that*, *whose*, and *which*, the relative word is marked as that same frame element type.

5. Corpus Statistics

The corpus consists of 2,500 documents containing both positive and negative examples of frames. For each lexical unit that maps to a motion frame in FrameNet, it is marked as positive (an instance of that frame) or negative (not an instance). When the automatic FrameNet parser correctly marks a frame, its elements are manually corrected. When the parser misses a frame, all of its elements are annotated manually. An alternative method, since (Bejan and Hathaway, 2007) works in a pipeline structure, would have been to annotate the lexical units first. After those were manually corrected, the parser would automatically annotate the frame elements and another round of corrections would occur. However, we did not feel this method would save annotators a significant amount of time.

	$P(A)$	$P(E)$	k
Frame Disambiguation	0.97	0.67	0.91
Literal/Figurative	0.93	0.64	0.81

	Accuracy
FE Boundary & Label	0.85

Table 7: Inter-annotator agreement on frame disambiguation, literal/figurative, and frame element annotations.

Currently, there are 3,389 manually corrected frames and 2,631 lexical units that have been verified as not triggering a particular frame. While only eight frame types have been annotated thus far, we have concentrated on some of the most common frames such as MOTION and ARRIVING. We intend to version the corpus, releasing more frames along with corrections from the previous versions as they become available. The current version is v0.1 [4].

Inter-annotator agreement was computed for frame disambiguation (whether the lexical unit evokes the frame or not), complete frame element agreement (all frame element boundaries and labels), and literal/figurative agreement. We used Cohen's Kappa agreement coefficient (Cohen, 1960) to measure agreement for frame disambiguation and literal/figurative classification. That score is computed by:

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (1)$$

where $P(A)$ is the percent of actual agreement and $P(E)$ is the agreement due to chance. While literal/figurative classification is a binary decision, frame disambiguation allows each lexical unit to choose from one or more frames that it triggers, plus a null option. However, since the annotator is only able to choose from a small number of frames, we model this as a binary decision as well.

Frame element boundary and label agreement, however, is not easily modeled with Kappa. Since we are doing a token-level boundary, the agreement due to chance of selecting the same start and end tokens is quite small, and thus not very informative of actual annotator agreement. For this, we present a simple accuracy assessment of element boundary and label agreement. In set notation, a single element is a 3-tuple consisting of a start offset, end offset, and label. Then R_1 and R_2 are the sets of elements for the first and second annotator, respectively. The overall accuracy is:

$$\frac{2|R_1 \cap R_2|}{|R_1| + |R_2|} \quad (2)$$

50 documents from the newswire sub-corpus were chosen for both annotators to work on. Both the MOTION and ARRIVING frames were annotated. The results are shown in Table 7.

The distribution of MOTION frames is shown in Table 8, as well as the number of lexical units that do not evoke the

[4] Available at http://www.utdallas.edu/∼kirk/projects/.

Domain	Positive LUs	Negative LUs	Avg. Tokens
Sports	617	175	957
Newswire	299	89	609
Financial	430	110	505
Military	490	290	577
Gossip	188	82	183

Table 8: Positive (evoked) and negative (un-evoked) MOTION frame counts, plus average document size (in tokens) from the UTD-MOTIONEVENT corpus.

MOTION frame. Gossip documents have the highest density of MOTION frames, while newswire documents have the lowest. This is consistent with our discussion of the use of motion from the beginning of the paper: since motion is a compact method of communication, one would expect less formal language use to contain motion more often than formal language use (newswire documents clearly being more formal than gossip documents). The distribution of literal/figurative frames varies from sub-corpus to sub-corpus as well. The only two domains that are currently well-annotated with literal and figurative frames are the sports and newswire domains. While 41.5% of all annotated frames in the newswire domain are marked as literal, only 20.1% of frames in the sports domain are literal. This is again consistent with the theory that more formal documents (e.g., newswire) use motion more strictly than less formal documents (e.g., sports).

6. The Utility of the UTD-MOTIONEVENT Corpus

The obvious first utility of the corpus is to inform and enhance semantic parsing for motion events. Improving semantic parsing consists of obtaining superior results for three FrameNet sub-tasks: (1) frame disambiguation, (2) frame element boundary detection, and (3) frame element labeling. Additionally, semantic parsing can be enhanced through the recognition of figurative events.

6.1. Frame Disambiguation

While the UTD-MOTIONEVENT corpus does not provide annotations for every frame, a frame disambiguation system trained on this data would have a close to accurate account of the distribution of motion and non-motion frames for certain domains. This is because our corpus annotates complete documents, indicating when a motion-evoking word actually evokes a motion frame and when it does not. Admittedly, when combined with the FrameNet data, this would not lead to a balanced corpus, though FrameNet by itself does not purport to be a balanced corpus of frame instances.

6.2. Frame Element Boundary Detection

Boundary detection is the task of identifying the argument boundaries for a given frame. Many FrameNet systems train boundary information across the entire corpus (Gildea and Jurafsky, 2002; Bejan and Hathaway, 2007), but it is known that boundaries can be very frame specific, since frames vary by syntactic structure (e.g., some frames are based on verbal predicates, others nominal predicates, and even others have prepositional or adjectival predicates) and specificity (e.g., some frames have elements for almost every adjective and adverb that may appear, while others have two or three elements in total and would ignore such modifiers or include them in a larger element). However, the alternative, training a boundary detector for each frame, suffers from a lack of training data. The inclusion of additional data for motion frames may make it possible to overcome this limitation and improve the learning of boundaries that are motion-specific.

Frame Element	FrameNet	UTD-MOTIONEVENT
Core Elements		
AREA	27	38
DIRECTION	19	71
DISTANCE	5	17
GOAL	68	567
PATH	160	169
SOURCE	37	60
THEME	274	907
Non-Core Elements		
CARRIER	9	9
CONTAINING_EVENT	2	2
DEGREE	1	20
DEPICTIVE	10	13
DURATION	1	20
FREQUENCY	0	0
ITERATION	0	0
MANNER	41	96
PATH_SHAPE	1	0
PLACE	15	29
PURPOSE	11	138
RESULT	3	5
SPEED	5	5
TIME	29	153

Table 9: Counts of unique frame elements in the original FrameNet data and in the UTD-MOTIONEVENT corpus.

6.3. Frame Element Labeling

Given a specific frame and a set of arguments, the task of frame element labeling is to assign argument types to each argument according to its semantic function. With the limited amount of training data available in each FrameNet frame specification, it can be difficult for a supervised system to generalize the potential realizations of a frame element. By providing a resource with far more frames (and thus frame elements) annotated at the document level, we can provide a distribution of frame elements that better approximates the data for that domain. Table 9 contains unique frame element counts for both FrameNet and the UTD-MOTIONEVENT corpus for the MOTION frame.

With the exception of the PATH_SHAPE element, every MOTION unique frame element realization has at least as many instances in UTD-MOTIONEVENT as it does in FrameNet. This should allow for better recognition of these elements. For instance, the FrameNet data contains only one DISTANCE realization that is actually a measurable distance ("*two miles*") and only one unique DURATION, which is not an explicit measured time quantity ("*for days*"). UTD-MOTIONEVENT contains 17 and 20 unique realizations for these elements, respectively, including "*48 hours*", "*200-day*", and "*more than 100 miles past Minneapolis*".

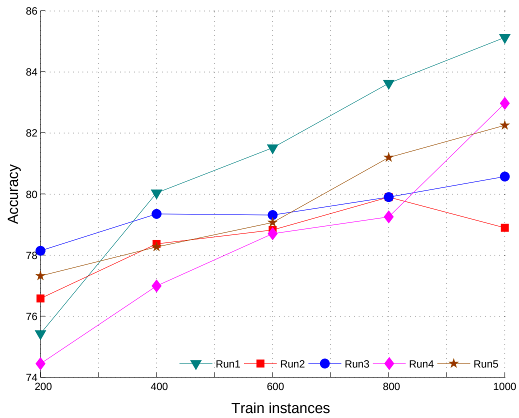

Figure 3: Performance of five runs on literal/figurative classification. Each run uses a new random train set.

6.4. Literal/Figurative Classification

The literal and figurative annotations available in the corpus may be used to train or evaluate a system for classifying motion events as literal or figurative. We have implemented a simple maximum entropy model in order to provide a baseline for such a system and to measure the impact of increased annotations. The model uses word, lemma, part-of-speech, entity, frame element, and WordNet hypernym features. We performed five experiments, each training on a new random sample of 1,000 frames and testing on 417 frames. Figure 3 shows the results of these models on increasing amounts of training data. Interestingly, the accuracy continues to increase for most runs, especially in the final segment. This suggests that even 1,000 training instances is not enough to reach the classic machine learning "plateau". Future versions of UTD-MOTIONEVENT should continue to improve results on this task.

7. Conclusion

We have developed a corpus of motion events and their participants using the FrameNet specification. The corpus is comprised of five diverse domains from web sources such as newswire feeds and blogs. An automatic FrameNet parser was used to annotate initial frames, and a pair of annotators manually corrected its output and added frames missed by the parser. Frames were then marked as literal or figurative. The corpus may serve as a resource for researchers working in semantic parsing, detection of figurative language, spatial reasoning, and other fields.

8. References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*.

Cosmin Adrian Bejan and Chris Hathaway. 2007. UTD-SRL: A Pipeline Architecture for Extracting Frame Semantic Structures. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval)*.

Daniel Bikel. 2002. Design of a Multi-lingual, Parallel-processing Statistical Parsing Engine. In *Proceedings of the conference on Human Language Technology*.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Charles J. Fillmore, Collin F. Baker, and Hiroaki Sato. 2002. The FrameNet Database and Software Tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*.

Charles J. Fillmore. 1982. Frame semantics. *Linguistics in the Morning Calm*, pages 111–137.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245–288.

Mark Johnson. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago.

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430.

Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. 2004. The NomBank Project: An Interim Report. In *Proceedings of HLT-EACL Workshop: Frontiers in Corpus Annotation*.

Srini Narayanan and Sanda Harabagiu. 2004. Question Answering Based on Semantic Structures. In *COLING-2004*.

Martha Palmer, Paul Kingsbury, and Daniel Gildea. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Kirk Roberts. 2009. Building an Annotated Textual Inference Corpus for Motion and Space. In *Proceedings of the ACL/IJCNLP Workshop on Applied Textual Inference*.

Josef Ruppenhofer, Michael Ellsworth, Miriam Petruck, Christopher Johnson, and Jan Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*.

Jan Scheffczyk and Michael Ellsworth. 2006. Improving the Quality of FrameNet. In *Proceedings of the LREC Workshop on Quality assurance and quality measurement for language and speech resources*.

Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. 2003. Using Predicate-Argument Structures for Information Extraction. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*.

Leonard Talmy. 1996. Fictive Motion in Language and "Ception". *Language and Space*.

Leonard Talmy. 2003. *Toward a Cognitive Semantics*. The MIT Press.

Marta Tatu and Dan Moldovan. 2005. A Semantic Approach to Recognizing Textual Entailment. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 371–378.

Nianwen Xue and Martha Palmer. 2004. Calibrating Features for Semantic Role Labeling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*.