

Representing Spatial Relations in FrameNet

Miriam R. L. Petruck Michael Ellsworth

International Computer Science Institute
Berkeley, California

{miriamp, infinity}@icsi.berkeley.edu

Abstract

While humans use natural language to express spatial relations between and across entities in the world with great facility, natural language systems have a facility that depends on that human facility. This position paper presents **FrameNet**'s¹ approach to representing spatial relations in language, and advocates its adoption for representing the meaning of spatial language. This work shows the importance of axis-orientation systems for capturing the complexity of spatial relations, which FrameNet encodes with semantic types.

1 Introduction

While humans use natural language to express spatial relations across entities in the world with great facility, natural language systems have a facility that depends on that human facility. (See (Mikolov et al., 2013) for a different perspective.) Natural Language Processing (NLP) applications such as robotic systems responding to commands about objects in a scene require accurate information on the spatial relations among those objects. In addition to determining *what* information to provide is the challenge of determining *how* to represent such information. This work presents the Frame Semantics view on representing spatial language, specifically as given in FrameNet (FN).

The rest of this position paper is organized as follows; Section 2 presents basic information about FN, including its current status; Section 3 provides a brief overview of related work; Section 4 covers the different kinds of spatial information that FN has recorded, including semantic types for characterizing spatial relation language, two of which constitute innovations over prior work; Section 5 shows how employing FN's spatial information can benefit NLP; and Section 6 briefly

¹<http://framenet.icsi.berkeley.edu>

discusses FrameNet's plans for future work on the language of spatial relations. Importantly, expanding FN's coverage for representing spatial relations is possible given existing FN infrastructure, i.e. frames, frame elements, and frame-to-frame relations, as well as semantic types.

2 Background to FrameNet

This section provides a very brief overview of FN, with information about its foundational principles and its relatively recent attention to basic linguistic phenomena that pose challenges to NLP, including the language of spatial relations, as well as details about its current status.

2.1 Frame Semantics and FrameNet

Frame Semantics (Fillmore, 1985) is the theoretical basis of **FrameNet** (Ruppenhofer et al., 2016), a knowledge base building effort, whose product, the FN database, is useful in NLP applications.

Central to the theory is the **semantic frame** (Fillmore, 1975), a schematic representation of a scene, whose frame elements (**FEs**), or semantic roles, identify participants and other conceptual entities, and whose underlying conceptual structure humans access for both encoding and decoding purposes. FrameNet adopted the lexical unit (**LU**) as the focus of analysis, defining an LU as a pairing of a lemma and a frame (Cruse, 1986).

FrameNet also distinguishes core and non-core frame elements. Thus, **core FEs** uniquely define a frame: BUYER, SELLER, MONEY, and GOODS² uniquely define frames that constitute the `Commercial_transaction`³ family of frames. In contrast, **non-core FEs** are relevant to events or situations in general; all events and situations occur at a time and in a place. The non-core

²Frame Element names appear in SMALL CAPS.

³Frame names appear in typewriter font.

FE PLACE is of particular importance for spatial relations (and is discussed further below).

FrameNet defines `Spatial_contact` as a scene in which a `FIGURE` is located in contact with a `GROUND`. With some words, the `FIGURE` is also asserted as fully or partially supported by the `GROUND` (e.g. *on*), while in others a support relation is denied (e.g., **TO**, as in *She put her hand **TO** the wall*), or unspecified (e.g. *against*). Some LUs assert a direction in which to find the `FIGURE` from the `GROUND` (e.g., *atop*).

Consider the two example sentences below that instantiate the `Spatial_contact` frame, where each realizes the `FIGURE` and the `GROUND` FEs.

1. Then [Maria `FIGURE`] fell and lay **ON** [the floor `GROUND`].
2. There were [a hat and feathers `FIGURE`] **ATOP** [the lid `GROUND`].

Contrast *on* and *atop*: *on* allows any direction of contact (*on the {ceiling, wall, ground}*), while *atop* specifies a particular direction of contact, i.e., above the `GROUND`. FN encodes such differences in a set of semantic types that specify axis systems and directions, based on these axis systems.

2.2 Current Status of FrameNet

At the time of this writing, the FN database holds over 1,220 frames, 13,640 LUs, and nearly 202,230 annotated sentences. Of importance here, FrameNet has defined 29 spatial language frames, covering 409 LUs that describe spatial relations, and approximately 4,200 annotated sentences, along with six semantic types for distinguishing spatial relation LUs.

3 Related Work

Linguists, computational linguists, and NLP researchers in particular, have studied spatial relations in language, and for the sake of developing annotation schema and NLP systems that take such information into consideration.

For example, (Dorr and Voss, 1993), addressed spatial relations for defining the relation between an interlingua and a system for representing knowledge in machine translation. Pursuing machine translation (Voss et al., 1998) investigated *how* the semantics of a spatial expression is allocated lexically.

(Jackendoff, 1996) considered how language users talk about what they see, addressing how

the mind might encode spatial information and linguistic information, as well how it might communicate between the two. That work also laid out some of the "boundary conditions for a satisfactory answer to these questions" (1996:3), and defined an approach to spatial representation. In a somewhat similar vein (as a contribution to cognitive semantic theory of conceptual structure), albeit from a different perspective, (Talmy, 2003) presented an approach to spatial representation that encompasses spoken and signed language.

More practically-oriented recent work (Kipper et al., 2004) expanded a verb lexicon (Kipper et al., 2000) using prepositions, i.e., linguistic material that encodes spatial information, extrapolating information about classes of verbs and their syntactic frames from (Levin, 1993). The annotation of spatial relations in language (Pustejovsky et al., 2011) constituted the focus of a workshop on interoperable semantic annotation, and included work on spatial role labeling with an eye toward extracting spatial information from corpora (Kordjamshidi et al., 2011) that also led to multimodal spatial role labeling (Kordjamshidi et al., 2017).

4 Spatial Information in FrameNet

This section describes the kind of information that FN provides about spatial relations, i.e., frames that characterize spatial relations, non-core FEs that indicate location of an event or an entity, frame-to-frame relations that link the relevant frames, and semantic types that give specific semantic information beyond a frame description or a LU definition.

4.1 Non-Core Frame Elements

An advantage of FrameNet as a resource for spatial language is that FN also models non-spatial language. This feature is especially important since spatial and non-spatial language are not completely separable. Most frames in FrameNet include one or more spatial FEs, the most common of which are PLACE, present in all frames that inherit from `Event`, as in # 3, and `LOCATION_OF_PROTAGONIST`, available in all frames with a causal entity (e.g. `CAUSE`) as in # 4, or a perceiver (e.g. `EXPERIENCER`).

3. The hiker **DIED** [in Antarctica `PLACE`].
4. She **TESTED** the bomb [from a safe distance `LOCATION_OF_PROTAGONIST`].

4.2 Frames and Frame Relations

Frames represent situations and states of affairs at a level of generalization that recognizes the commonalities within and across sets of semantically related lexical items. FN records several frame-to-frame relations to indicate how frames relate to each other in its hierarchy of frames; **Inheritance** and **Using** are the relevant ones for spatial relations language. Frames that *inherit* `Locative_relation` capture the lexical material for spatial relations in English.⁴

Inheritance exists between a parent frame and a child frame under specific circumstances: for each FE, frame relation, and semantic characteristic in the parent, the same or a more specific corresponding entity in the child exists, as in the relationship between `Locative_relation` and `Interior_profile_relation`. `Using` is a relationship between a child frame and parent frame in which only some of the FEs in the parent have a corresponding entity in the child; if such exist, they are more specific. `Using` holds between `Interior_profile_relation` and `Bounded_region`.

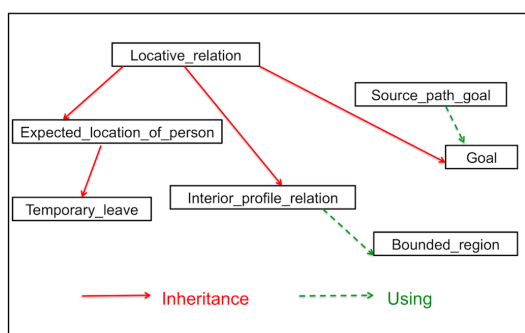


Figure 1: Inheritance and Using Relations

Figure 1 depicts some frames related to the `Locative_relation` frame via `Inheritance`, some of which also employ the `Using` relationship. Note that a frame may inherit one frame and use another: `Goal` inherits `Locative_relation` and uses `Source_path_goal`.⁵

The static spatial relations frames inherit from `Locative_relation`, which defines the basic situation where the `FIGURE` entity has a location that is determined by means of a relation to the

⁴<https://tinyurl.com/y7jpt9hd>. FN team members are well-aware that the work has only begun.

⁵The careful reader will note the “incorrect” direction of the arrows in Figure 1, which follows conventions that FrameNet uses.

`GROUND`, another entity. These static spatial relations all share this basic structure; moreover, each specific frame also holds a `Using` relation to an image schema⁶ that defines the relation between the `FIGURE` and the `GROUND`.

FrameNet models the lexical unit *in* as a member of the `Interior_profile_relation` frame (which inherits `Locative_relation`). Its frame elements include `FIGURE`, the located entity and `GROUND`, the basis of the location. `Interior_profile_relation` uses the `Bounded_region` image schema, which defines a boundary, an inside, and an outside. Part of `Using` specifies that the `FIGURE` identifies the inside region, and the `GROUND` identifies the boundary. FN distinguishes among other LUs by defining them in different related frames in this family (of frames) and via semantic types that cross-cut frame distinctions.

4.3 Semantic Types

Linguists, anthropologists, and computer scientists have studied the cognitive, cultural, linguistic, and computational aspects of space and spatial relations for decades (Herskovitz, 1987; Bowerman and Pederson, 1992; Regier, 1996; Levinson, 2003). FrameNet has defined a cognitively-inspired set of semantic types for spatial LUs to indicate (1) with respect to which axis-system(s) (Talmy, 2000) a given LU is defined, and (2) which direction(s) from these axes the active zone a given LU selects.

	Semantic Type	Example
Basic Axis System	absolute viewpoint-based motion-based ground-based	to the east of Pam to the left of Sue ahead of Paul to Chuck’s left
FN Added	Near_absolute Flexible	atop the tree in front of her

Table 1: Semantic Types for Spatial Relations

As Table 1 shows, the basic axis systems include four types: **absolute** (*to the east of X*); **viewpoint-based** (*to the left of X*); **motion-based** (*ahead of X*); and **ground-based** (*to X’s left*). FrameNet has defined a semantic type for each of these four possibilities. Besides semantic types

⁶Image schemas (Lakoff and Johnson, 1980) are cognitive models, such as of containment, oppositional forces, and verticality, which language users apply to understand and reason about the world. FN characterized image schemas as frames.

named with the terminology of the basic axis systems, FN has defined two other **new** semantic types: **Near_absolute** (*atop*) and **Flexible** (*in front*). These two semantic types innovate on previous work (Talmy, 2000), and derive from FN’s fairly recent work on spatial relations.

Using semantic types for each direction in each axis system would seem like a simple enough modeling choice. However, LUs exhibit patterns whereby a default axis system is overridden under specific conditions. Thus, for example, some LUs inflexibly select an absolute direction (e.g., *east*); some normally select an absolute direction, but allow a ground-based one (*atop*); and some default to a ground-based direction, but allow viewpoint-based or motion-based direction (*in front*). FrameNet’s semantic types specify the pattern of axis ambiguity a LU exhibits.

5 Operationalization

FrameNet’s models of spatial language consist of frames, frame relations, and semantic types, all static and abstract. However, using FrameNet’s models for visual scene understanding requires grounded and flexible implementations. As such, the machinery needed to match a spatial description like *the cow IN FRONT of the train* to an image requires the following: (1) object recognition of the GROUND (train); (2) image parsing for each axis system centered on the train (since *in front* is a flexible lexical unit); and (3) recognition of the FIGURE (cow) in the forward-pointing vector for each axis system.

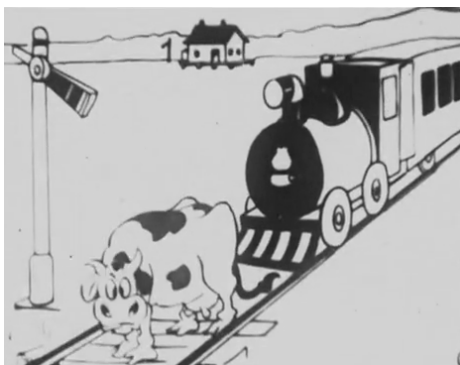


Figure 2: the cow in front of the train

FrameNet contributes in three critical ways to this matching process, by providing the following:

1. training data showing the manifestation of the FIGURE and GROUND roles in language;

2. an inventory of frames for spatial situations that any system must recognize (e.g. Containment, Contact);
3. an inventory of semantic types for axis systems and their vectors.

Crucially, FrameNet’s semantic types distinguish the flexible LU *in front* from a Motion-based LU (*ahead*), where only the motion-based forward zone of the train is scanned.

6 Future Work

This position paper has described FrameNet’s work on static spatial relations. It has shown that FN provides critical information for certain NLP applications that require input for the processing of spatial relations language.

Going forward and with sufficient resources, FrameNet plans to analyze other types of spatial relations language, including the following:

- Dynamic spatial relations language, e.g. *to*, *from*, as in:

She went **TO** the lake **FROM** the house.

- Pseudo-dynamic spatial relations, e.g. *across*, as in:

She lives **ACROSS** the bridge.

- Constructions (Kay and Fillmore, 1999; Fillmore, 2013) that license static spatial relations to be construed as GOALS, as in:

I went **UNDER** the bridge.

Preliminary studies of the other types of spatial language indicate that FN’s existing system of frames, frame elements, frame-to-frame relations, and semantic types will serve as a solid foundation for future work.

References

- Melissa Bowerman and Eric Pederson. 1992. Topological relations picture series. In Stephen C. Levinson, editor, *Space Stimuli Kit 1.2*. Max Planck Institute for Psycholinguistics, Nijmegen.
- D.A. Cruse. 1986. *Lexical Semantics*. Cambridge University Press, Cambridge, UK.
- Bonnie J. Dorr and Clare R. Voss. 1993. Machine translation of spatial expressions: Defining the relation between an interlingua and a knowledge representation system. In *Proceedings of the 11th National Conference on Artificial Intelligence*. Washington, DC, USA, July 11-15, 1993., pages 374–379.

- Charles J. Fillmore. 1975. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1:123–131.
- Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 6(2):222–254.
- Charles J. Fillmore. 2013. Berkeley construction grammar. In Thomas Hoffmann and Graeme Trousdale, editors, *Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford.
- Annette Herskovitz. 1987. *Language and Spatial Cognition: An interdisciplinary study of the prepositions in English*. Studies in Natural Language Processing. Cambridge University Press, Cambridge.
- Ray Jackendoff. 1996. The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 1–30. MIT Press, Cambridge.
- Paul Kay and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what’s x doing y? construction. *Language*, 75(1):1–33.
- Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 691–696. AAAI Press.
- Karin Kipper, Benjamin Snyder, and Martha Palmer. 2004. Using prepositions to extend a verb lexicon. *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*, pages 23–29.
- Kordjamshidi, Parisa, 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):4:1–4:36.
- Parisa Kordjamshidi, Taher Rahgooy, Marie-Francine Moens, James Pustejovsky, Umar Manzoor, and Kirk Roberts. 2017. Clef 2017: Multimodal spatial role labeling (msprl) task overview. In Jones G. et al., editor, *Lecture Notes in Computer Science*, vol. 10456, pages 367–377. Springer, Cham.
- George Lakoff and Mark Johnson. 1980. *Metaphors we Live by*. University of Chicago Press, Chicago.
- Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.
- Stephen C. Levinson. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Language Culture and Cognition. Cambridge University Press.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.
- James Pustejovsky, Jessica L Moszkowicz, and Marc Verhagen. 2011. Iso-space: The annotation of spatial information in language. *Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, 6:1–9.
- Terry Regier. 1996. *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. ICSI: Berkeley.
- Leonard Talmy. 2000. *Toward a Cognitive Semantics*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- Leonard Talmy. 2003. The representation of spatial structure in spoken and signed language: A neural model. *Language and Linguistics*, 4(2):207–250.
- Clare R. Voss, Bonnie J. Dorr, and M. Ulku Sencan. 1998. Lexical allocation in interlingua-based machine translation of spatial expressions. In Patrick Olivier and Klaus-Peter Gapp, editors, *Representation and Processing of Spatial Expressions*, pages 133–148. Psychology Press.