



Published in final edited form as:

LREC Int Conf Lang Resour Eval. 2020 May ; 2020: 2251–2260.

Rad-SpatialNet: A Frame-based Resource for Fine-Grained Spatial Relations in Radiology Reports

Surabhi Datta¹, Morgan Ulinski², Jordan Godfrey-Stovall¹, Shekhar Khanpara³, Roy F. Riascos-Castaneda³, Kirk Roberts¹

¹School of Biomedical Informatics, The University of Texas Health Science Center at Houston

²Department of Computer Science, Columbia University

³McGovern Medical School, The University of Texas Health Science Center at Houston

Abstract

This paper proposes a representation framework for encoding spatial language in radiology based on frame semantics. The framework is adopted from the existing SpatialNet representation in the general domain with the aim to generate more accurate representations of spatial language used by radiologists. We describe Rad-SpatialNet in detail along with illustrating the importance of incorporating domain knowledge in understanding the varied linguistic expressions involved in different radiological spatial relations. This work also constructs a corpus of 400 radiology reports of three examination types (chest X-rays, brain MRIs, and babygrams) annotated with fine-grained contextual information according to this schema. Spatial trigger expressions and elements corresponding to a spatial frame are annotated. We apply BERT-based models (BERT_{BASE} and BERT_{LARGE}) to first extract the trigger terms (lexical units for a spatial frame) and then to identify the related frame elements. The results of BERT_{LARGE} are decent, with F1 of 77.89 for spatial trigger extraction and an overall F1 of 81.61 and 66.25 across all frame elements using gold and predicted spatial triggers respectively. This frame-based resource can be used to develop and evaluate more advanced natural language processing (NLP) methods for extracting fine-grained spatial information from radiology text in the future.

Keywords

spatial relations; radiology; frame semantics

1. Introduction

Radiology reports are one of the most important sources of medical information about a patient and are thus one of the most-targeted data sources for natural language processing (NLP) in medicine (Pons et al., 2016). These reports describe a radiologist’s interpretation of one or more two- or three-dimensional images (e.g., *X-ray*, *computed tomography*, *magnetic resonance imaging*, *ultrasound*, *positron emission tomography*). As a consequence, these reports are filled with spatial relationships between medical findings (e.g., *tumor*,

pneumonia, inflammation) or devices (e.g., *tube, stent, pacemaker*) and some anatomical location (e.g., *right ventricle, chest cavity, T4, femur*). These relations describe complex three-dimensional interactions, requiring a combination of rich linguistic representation and medical knowledge in order to process.

Table 1 shows an example radiology report.¹ The example demonstrates the large number of medical findings and the relationships with various anatomical entities often found within radiology reports.

A flexible way of incorporating fine-grained linguistic representations with knowledge is through the use of frames for spatial relations (Petrucci and Ellsworth, 2018). Notably, SpatialNet (Ulinski et al., 2019) extends the use of FrameNet-style frames with enabling the connection of these frames to background knowledge about how entities may interact in spatial relationships.

In this paper, we propose an extension of SpatialNet for radiology, which we call Rad-SpatialNet. Rad-SpatialNet is composed of 8 broad spatial frame types (as instantiated by the relation types), 9 spatial frame elements, and 14 entity types. We also describe an initial study in the annotation of the Rad-SpatialNet schema on a corpus of 400 radiology reports comprising of 3 different imaging modalities from the MIMIC-III clinical note corpus (Johnson et al., 2016). The annotated dataset will soon be made publicly available. There are 1101 sentences that contain at least one spatial relation in the corpus. 1372 spatial trigger expressions (lexical units for spatial frames) are identified in total with the most frequently occurring spatial relation types being ‘Containment’, ‘Description’, and ‘Directional’. Finally, we describe a baseline system for automatically extracting this information using BERT (Devlin et al., 2019) based on deep bi-directional transformer architecture.

2. Related Work

A range of existing work has focused on extracting isolated radiological entities (e.g., findings/locations) utilizing NLP in radiology reports (Hassanpour and Langlotz, 2016; Hassanpour et al., 2017; Cornegruta et al., 2016; Bustos et al., 2019; Annarumma et al., 2019). There exists recently published radiology image datasets labeled with important clinical entities extracted from corresponding report text (Wang et al., 2017; Irvin et al., 2019). Most of these studies have been targeted toward attribute extraction from text without focusing on recognizing relations among these entities. Only a few studies have focused on relation extraction from radiology text (Sevenster et al., 2012; Yim et al., 2016; Steinkamp et al., 2019). However, the datasets were limited to specific report types (e.g., hepatocellular carcinoma) and the relations extracted do not capture spatial information.

Extracting spatial relations has been studied in the general-domain for various purposes such as geographic information retrieval (Yuan, 2011), text-to-scene generation (Coyne et al., 2010; Coyne and Sproat, 2001), and human-robot interaction (Guadarrama et al., 2013).

¹Unlike the data annotated in this paper, the example in Table 1 is publicly available without restriction from openi.nlm.nih.gov (image ID: CXR1000_IM-0003-1001).

One of the schemas developed for representing spatial language in text is spatial role labeling (Kordjamshidi et al., 2010; Kordjamshidi et al., 2017). In the medical domain, some studies have extracted spatial relations from text including biomedical and consumer health text (Kordjamshidi et al., 2015; Roberts et al., 2015). In radiology, two prior works have identified spatial relations between radiological findings and anatomical locations, however, the reports were specific to appendicitis and did not extract other common clinically significant contextual information linked to the spatial relations (Roberts et al., 2012; Rink et al., 2013).

Frame semantics provide a useful way to represent information in text and has been utilized in constructing semantic frames to encode spatial relations (Petrucci and Ellsworth, 2018). The Berkeley FrameNet project (Baker, 2014) contains 29 spatial frames with a total of 409 spatial relation lexical units. A recent work (Ulinski et al., 2019) proposed a framework known as SpatialNet to map spatial language expressions to actual spatial orientations by utilizing resources such as FrameNet (Baker, 2014) and VigNet (Coyne et al., 2011). In this paper, we aim to extend the SpatialNet framework to encode spatial language in radiology and utilize the framework to develop a dataset of reports annotated with detailed spatial information.

3. Rad-SpatialNet Description

Rad-SpatialNet provides a framework description to represent fine-grained spatial information in radiology reports by converting the linguistic expressions denoting any spatial relations to radiology-specific spatial meanings. For this, we extend the core design proposed in the general domain SpatialNet (Ulinski et al., 2019), which is based on FrameNet and VigNet, and tailor the framework specifically to encode spatial language in the domain of radiology. We have presented an overview of radiological spatial relations and the main participating entities in Figure 1. SpatialNet describes spatial frames by linking surface language to lexical semantics and further mapping these frames to represent the real spatial configurations. We update SpatialNet in this work with the aim to disambiguate the various spatial expressions used by radiologists in documenting their interpretations from radiographic images.

To build Rad-SpatialNet, we first utilize the language in radiology reports to construct a set of *spatial frames* leveraging linguistic rules or valence patterns. The fundamental principle in forming the radiology spatial frames is the same used for constructing frames in FrameNet/SpatialNet. However, the main difference is that the target words (*lexical units*) of the frames are the spatial trigger words which are more common in radiology and are usually prepositions, verbs, and prepositional verbs. We then construct a list of *spatial vignettes* to transform these high-level spatial frames into more fine-grained frame versions by incorporating semantic, contextual or relation type constraints as well as radiology domain knowledge. The fine-grained frames reveal the true meaning of the spatial expressions from a radiology perspective. We describe these final frames containing the actual spatial configurations as the *spatio-graphic primitives*. Unlike SpatialNet that uses the VigNet ontology to map the different lexical items into semantic categories, we utilize

the publicly available radiology lexicon, RadLex, to map different radiological entities mentioned in the reports to standard terminologies recommended in radiology practice.

The main components involved in the proposed Rad-SpatialNet framework are described in the following subsections.

3.1. Ontology of Radiology Terms

We leverage RadLex (Langlotz, 2006) to map the various radiological entities along with other clinically important information in the report text to standard unified vocabularies to facilitate standardized reporting and decision support in radiology practice as well as research. RadLex consists of a set of standardized radiology terms with their corresponding codes in a hierarchical structure. Rad-SpatialNet utilizes the RadLex ontology to map all possible contextual information with reference to any spatial relation in the reports to the broader RadLex classes which capture the semantic types of the various information. For example, ‘*Endotracheal tube*’ which is a TUBE and a type of IMPLANTABLE DEVICE is mapped to the broad RadLex class MEDICAL DEVICE. Similarly, ‘*Ground-glass opacity*’ belongs to the RadLex class OPACITY which falls under the broad class IMAGING OBSERVATION. Thus, terms such as ‘*opacity*’, ‘*opacification*’, and ‘*Ground-glass opacity*’ are mapped to IMAGING OBSERVATION. We also link the various modifier or descriptor entities to standard RadLex descriptors. For instance, in ‘*roundedparenchymal opacity*’, ‘*rounded*’ is mapped to the MORPHOLOGIC DESCRIPTOR class of RadLex, which is one of the categories of RadLex descriptors. Unlike VigNet, RadLex does not contain any graphical relations representing spatial configurations. So, we utilize Radlex mainly to map the terms in reports to radiology-specific semantic categories and not for creating spatio-graphic primitives. The mapped entities are utilized in the following steps to construct the spatial frames and subsequently the radiology-specific spatio-graphic primitives.

3.2. Spatial Frames

The spatial frames organize information in a radiology report sentence containing any spatial relation between common radiological entities (e.g., imaging observations and anatomical structures) according to the frame semantic principles, similar to FrameNet and SpatialNet. All the spatial frames created are inherited from the SPATIAL-CONTACT frame in FrameNet. We adopt similar valence patterns as defined in SpatialNet by specifying various lexical and syntactic constraints to automatically identify the frame elements from a sentence. However, there are differences in the set of frame elements in Rad-SpatialNet compared to SpatialNet. For example, besides FIGURE and GROUND, some of the other common elements in the Rad-SpatialNet frames are HEDGE, DIAGNOSIS, DISTANCE, and RELATIVE POSITION.

To do this, we identify the most frequent words or phrases expressing spatial relations in radiology. Such a word/phrase also forms the lexical unit for a spatial frame. The type or the sense of the spatial trigger is also recognized to include the spatial relation type information in the frame. For example, in the sentence describing the exact position of a medical device - ‘*The umbilical venous catheter tip is now 1 cm above the right hemidiaphragm*’, ‘*above*’ is the spatial trigger having a directional sense. This instantiates a spatial frame with **Directional** as the relation type and *above.prep* as the lexical unit. Some other common

spatial relation types in Rad-SpatialNet are **Containment**, triggered by lexical units such as *in.prep*, *within.prep*, and *at.prep*; **Descriptive**, triggered by lexical units such as *shows.v*, *are.v* and *with.prep*; and **Spread** triggered by lexical units such as *extend (into).prep*, *through-out.prep*, and *involving.v*. The spatial relation types are shown in Table 2 and the elements identified for the spatial frames are described in Table 3.

The semantic type of the **FIGURE** and **GROUND** elements are identified for each spatial frame constructed using the RadLex ontology. For the example above related to the positioning of the umbilical venous catheter tip, the semantic type of **FIGURE** is **MEDICAL DEVICE** and the type of **GROUND** is **ANATOMICAL LOCATION**. All this information-combining the semantic types and the spatial relation type-are used to further refine the spatial frame **Figure-Directional-Ground-SF** as **Medical Device-Directional-Anatomical Location-SF**.

3.3. Spatial Vignettes

The main idea of spatial vignettes are also adopted from SpatialNet. In this paper, we develop vignettes primarily to resolve the ambiguities involved in using the same spatial trigger words or phrases to describe different radiological contexts. In other words, the same spatial expressions might have different spatial configurations based on the context or the radiological entities associated.

The vignettes connect the spatial frames to spatio-graphic primitives utilizing the RadLex ontology, semantic/relation type constraints as well as domain knowledge to generate more accurate spatial representations of radiology language. Consider the following two sentences having the same spatial trigger ‘*extends into*’:

1. There is interval increase in the right pleural effusion which **extends into** the fissure.
2. There is an NG tube which **extends into** the stomach.

The first sentence contains a radiologist’s description of a fluid disorder ‘*pleural effusion*’ with respect to an anatomical reference–fissure in the pleural cavity (a closed space around the lungs), whereas the second sentence describes the positioning of a feeding tube. It is difficult to interpret the actual spatial meaning of the same prepositional verb ‘*extends into*’ in these two different contexts solely from the lexical information. The spatial vignettes map the spatial frames corresponding to these sentences to different spatio-graphic primitives representing the actual spatial orientation of ‘*extends into*’ utilizing radiology domain knowledge.

Semantic constraints are applied to the **FIGURE** and **GROUND** frame elements, whereas a spatial relation type constraint is also added to the relation sense. If the semantic category of the **FIGURE** is **MEDICAL DEVICE** and the relation type is **DIRECTIONAL**, with the lexical unit *extends (into).prep*, then a spatial vignette will generate the spatio-graphic-primitive **MEDICAL DEVICE-TERMINATES INTO-ANATOMICAL LOCATION-SGP** from the spatial frame **MEDICAL DEVICE-DIRECTIONAL-ANATOMICAL LOCATION-SF** (corresponding to example 2 above). Another vignette will produce the spatio-graphic primitive **DISORDER-EXTENDS INTO-ANATOMICAL LOCATION-SGP** if the semantic category of the **FIGURE** is **DISORDER**

instead of **MEDICAL DEVICE** and particularly refers to fluid-related disorders like ‘*pleural effusion*’ (corresponding to example 1 above). The vignettes here determine the spatial meanings of radiology sentences based on both the semantic types of **FIGURE** element and specific properties of the **DISORDER** type (for example, *fluidity* in this case). Similar ambiguities are observed for spatial expressions containing lexical units such as *projects (over).prep* and *overlying.v* as they often occur in both device and observation or disorder-related contexts. Thus, the vignettes differentiate the real orientation of the spatial expression when it is used in context to a medical device versus any other radiographic finding. Consider another set of examples below where both sentences are related to the tip position of two medical devices:

1. The umbilical arterial catheter tip **projects over** the T8–9 interspace.
2. The tip of the right IJ central venous line **projects over** the upper right atrium.

Here, the spatial vignettes produce a more specific spatial representation of the prepositional verb ‘*projects over*’ based on the details of anatomical location. A spatial vignette will produce the spatio-graphic primitive **MEDICAL DEVICE TIP-IN FRONT OF-ANATOMICAL LOCATION-SGP** for the first sentence. **IN FRONT OF** is derived as the anatomical structure is corresponding to the ‘*T8-9 interspace*’ in the spine which is often used as the level of reference to indicate the position of catheters and tubes and these tubes and catheters are in front of the spine. Another vignette will produce **MEDICAL DEVICE TIP-TERMINATES AT LEVEL OF-ANATOMICAL LOCATION-SGP** for the second sentence as the IJ venous line lies within the internal jugular vein and might go upto the ‘*right atrium*’.

Consider the following sentences containing ‘*overlying.v*’ as the lexical unit:

1. There is a less than 1 cm diameter rounded nodular opacity **overlying** the 7th posterior rib level.
2. The left IJ pulmonary artery catheter’s tip is currently **overlying** the proximal SVC.

If the **FIGURE** is **IMAGING OBSERVATION** and the **GROUND** is particularly associated with anatomical locations such as ribs, a spatial vignette will produce the primitive **IMAGING OBSERVATION-PROJECTS OVER-ANATOMICAL LOCATION-SGP** for the first sentence. Ribs are also used the same way as spine to describe the level of objects or pathology. However, for the second sentence, the spatio-graphic primitive will be **MEDICAL DEVICE TIP-TERMINATES AT LEVEL OF-ANATOMICAL LOCATION-SGP** as the **FIGURE** is **MEDICAL DEVICE** and the **GROUND** (anatomical location) is ‘*SVC*’.

The following examples contain the spatial trigger ‘*of*’ connecting two anatomical structures or parts of anatomical structures:

1. There is increased signal identified within the pons extending to the right side **of** the midline.
2. The UA catheter tip overlies the left pedicle **of** the T9 vertebral body.

In the above cases with respect to the spatial preposition ‘*of*’, the semantic types of both FIGURE and GROUND are referring to ANATOMICAL LOCATIONS. However, there is a difference in the interpretation of the same trigger word ‘*of*’. The vignette adds a constraint that if the spatial relation type is DESCRIPTIVE and both FIGURE and GROUND have semantic type as ANATOMICAL LOCATION, then the meaning of the preposition is determined based on the words of the FIGURE element. If the words are either ‘*side*’ or ‘*aspect*’, then ‘*of*’ refers to a subarea/side of the GROUND anatomical location. Whereas for other words, ‘*of*’ refers to a specific identifiable anatomical structure contained within the GROUND element, similar to ‘*pedicle*’ present at each vertebra in Example 2 above. The spatial vignettes corresponding to the three spatial expressions described above - *extends (into).prep*, *projects (over).prep*, and *of.prep* are illustrated in Figure 2.

4. Annotation Process

We annotated a total of 400 radiology reports—Chest X-ray reports (136), Brain MRI reports (127), and Babygram reports (137)—from the MIMIC III clinical database (Johnson et al., 2016). The language used in MIMIC reports is more complex and the report lengths as well as sentence lengths are long compared to other available datasets such as open chest X-ray reports (Demner-Fushman et al., 2016). We filtered the babygram-related reports following the ‘*babygram*’ definition, that is, an X-ray of the whole body of an infant (usually newborn and premature infants). Since babygram reports have frequent mentions of medical device positions and involve multiple body organs, we incorporate this modality mainly with the intention to build a corpus with balance between two major spatially-grounded radiological entities—imaging observations/clinical findings and medical devices. We pre-processed the reports to de-identify some identifiable attributes including dates and names and also removed clinically less important contents. All the sentences in the reports containing potential spatial relations were annotated by two human annotators. The annotation was conducted using Brat. The annotations were reconciled three times—following the completion of 50, 200, and 400 reports. Since identifying the spatial frame elements involves interpreting the spatial language from the contextual information in a sentence, the annotations of the first 50 reports (i.e. our calibration phase) differed highly between the annotators. Some of the major disagreements related to spatial relation sense were discussed and the annotation guidelines were updated after first two rounds of annotation. Examples of sample annotations are provided in Figure 3.

To provide more insights into some of the complexities encountered in the annotation process, we highlight a few cases here. **First**, oftentimes two anatomical locations are present in a sentence and in such scenarios, an intermediate anatomical location (first location occurrence) is chained as a FIGURE element in context to the broader anatomical location (second location occurrence and the GROUND element). However, this chaining is not valid if the two locations are not connected. Consider the sentences below:

1. There are few small air fluid levels in [**mastoid air cells**]_{FirstLocation} within the left [**mastoid process**]_{SecondLocation}.
2. Area of increased signal adjacent to the left lateral [**ventricle**]_{FirstLocation} at the level of [**corona radiata**]_{SecondLocation}.

For the first sentence, ‘*mastoid air cells*’ is contained within ‘*mastoid process*’ and these anatomical locations are connected. Therefore, ‘*mastoid air cells*’ is annotated as **FIGURE** element and ‘*mastoid process*’ as the **GROUND** in context to the spatial frame formed by the lexical unit ‘*within.prep*’. Note that ‘*mastoid air cells*’ is the **GROUND** element associated with the **FIGURE** ‘*air fluid levels*’ through the spatial trigger ‘*in*’. However, for the second sentence, ‘*ventricle*’ and ‘*corona radiata*’ are two separate anatomical references and are not connected. Hence, two separate spatial relations are formed with the radiographic observation ‘*signal*’, one between ‘*signal*’ and ‘*ventricle*’ described through ‘*adjacent to*’ and the other between ‘*signal*’ and ‘*corona radiata*’ described through ‘*at*’ and the **RELATIVE POSITION** ‘*level of*’. **Second**, some instances require correct interpretation of whether prepositions such as ‘*of*’ are **SPATIAL TRIGGERS** or are part of location descriptors. Note the examples below:

1. PICC line with its tip located at the **junction of** superior vena cava and left brachiocephalic vein.
2. Abnormal signal in posterior portion **of** spinal cord.

In the first sentence, ‘*junction of*’ is annotated as the **RELATIVE POSITION** describing the connection between ‘*superior vena cava*’ and ‘*brachiocephalic vein*’, whereas in the second sentence, ‘*of*’ is a **SPATIAL TRIGGER** connecting the **FIGURE**–‘*portion*’ and the **GROUND**–‘*spinal cord*’. **Third**, although location descriptor words such as *anterior*, *lateral*, and *superior* are usually annotated as **RELATIVE POSITION**, in a few cases they are annotated as **SPATIAL TRIGGER**. For example, note the following two sentences:

1. There is an area of high signal intensity extending into the [**anterior**]**RELATIVE POSITION** mediastinum.
2. There is a 6 mm lymph node [**anterior to**]**SPATIAL TRIGGER** the carina.

In the first sentence, ‘*anterior*’ is used to describe ‘*mediastinum*’ and is annotated as a **RELATIVE POSITION**, whereas in the second sentence, ‘*anterior*’ contributes in perceiving the actual spatial sense and hence ‘*anterior to*’ is annotated as a **SPATIAL TRIGGER**.

5. Results

5.1. Annotation statistics

The inter-annotator agreement results are shown in Table 4. We calculate the overall F1 agreement for annotating the spatial relation types, the main entities, and the spatial frame elements. The agreement measures are particularly low (around 0.4) for **FINDING/OBSERVATIONS** as often there are higher chances of boundary mismatch in the process of separating the descriptor-related words from the main finding or observation term. There are very few instances of **PROCESS** entities and cases where the disorder terms act as **REASONS** in the corpus which have resulted in the agreement measures being zero. Ultimately, Rad-SpatialNet requires an extremely knowledge-intensive annotation process, so low agreement at this stage is not unreasonable. Future work will include additional quality checks to ensure the semantic correctness of the annotations.

5.2. Corpus statistics

358 (89.5%) of the reports contain spatial relations. The reconciled annotations contain a total of 1101 sentences with mentions of spatial triggers. There are 1372 spatial trigger terms in total (average of 3.8 triggers per report). The frequencies of the entity types, the types of spatial relations, and the spatial frames are presented in Table 5. The predominant spatial trigger types to instantiate spatial frames are ‘Containment’, ‘Description’, and ‘Directional’. 81 of the 330 DEVICE entities are described using various descriptor terms, 327 of the 436 OBSERVATION entities, 240 of the 367 CLINICAL FINDINGS, 190 of the 390 DISORDERS, and 564 of the 1492 ANATOMICAL LOCATIONS contain descriptors. The distribution of various types of descriptors (consistent with RadLex) across the main radiological entities are shown in Table 6. Note that the figures in Table 5 and Table 6 are considering only those entities and their descriptors which are involved in a spatial relation. There are 1328 frame instances which correspond to the main radiological entities as demonstrated in Table 5. Among the remaining 44 spatial triggers, 11 are related to Port/Lead of devices and 33 describes how a specific part of an anatomical structure is linked to the main part.

5.3. Baseline system performance

As an initial step of extracting the spatial trigger terms (lexical units of spatial frames) and the associated spatial information (spatial frame elements) in report sentences, we utilize both BERT_{BASE} and BERT_{LARGE} pre-trained language models as our baseline systems. We formulate both the tasks of lexical unit identification and frame elements extraction as sequence labeling task. We initialize the model parameters obtained by pre-training BERT_{BASE} and BERT_{LARGE} on MIMIC-III clinical notes (Si et al., 2019) for 300K steps and fine-tune the models on our constructed corpus. For fine-tuning, we set the maximum sequence length at 128, learning rate at 2e-5, number of training epochs at 4 and use cased version of the models.

10-fold cross validation is performed to evaluate the model’s performance with 80–10–10% training, validation, and test splits of the reports. First, the spatial triggers are extracted in a sentence. Second, we extract the common frame elements with respect to the trigger. For element extraction, we evaluate the system performance using both the gold spatial triggers and the predicted triggers on the test set. The results are shown in Table 7, Table 8, and Table 9.

The results demonstrate that BERT_{LARGE} performs better than BERT_{BASE} for both spatial trigger prediction and frame elements prediction. However, in case of spatial trigger, the recall is low for both the models (Table 7). For element extraction, BERT_{BASE}’s overall F1 combining all the frame elements is 79.48 (using gold spatial triggers) and 48.55 (using predicted triggers), whereas, for BERT_{LARGE}, the difference in the overall F1 between using gold (81.61) versus predicted triggers (66.25) is much lower (15.4 vs 30.9). The F1 values are zeroes for REASON and ASSOCIATED PROCESS because of few occurrences in the dataset.

6. Discussion

The Rad-SpatialNet framework attempts to capture all possible contextual information in radiology reports from a spatial perspective. The aim is to utilize linguistic information as well as domain knowledge to accurately represent spatial language in radiology. The proposed examples of spatial vignettes described above have been validated by a practicing radiologist (SK). We are currently in the process of developing the valence patterns and vignettes by taking input from radiologists.

The results of the baseline system illustrates that it is difficult to identify the spatial trigger expression. This might be because of their wide variation in the reports and many of these appear as multi-word expressions. We only use the developed corpus to extract the core frame elements in a spatial frame. The results indicate that there is enough scope for improving the predictions by developing more advanced methods in the future.

In this work, we only consider intra-sentence spatial relations. Oftentimes, we encounter inter-sentence relations and scenarios where the differential diagnoses are documented in the sentence following the spatial relation or even far apart in the ‘Impression’ section (around 12.75% of the reports in our corpus). Other complex information in context to a spatial trigger could be considered for later work. Consider the sentence - ‘*The tip of the catheter has a **mild rightward curve**, suggesting that it may be directed into a portal vein.*’ Here, information about ‘*intermediate position change*’ might also be annotated as a frame element. Further, some phrases such as ‘*needs repositioning*’ can be differentiated from POSITION STATUS as POSITION RECOMMENDATION. Thus, further refinements to Rad-SpatialNet are still necessary.

7. Conclusion

This work aims to develop a resource for encoding fine-grained spatial information in radiology reports. For this, we extend an existing spatial representation framework in the open-domain, spatialNet, to accurately represent radiology spatial language. We describe the components of Rad-SpatialNet. Construction of linguistic rules and the spatial vignettes are currently under development. We annotate 400 radiology reports with important spatial information and apply a baseline model based on BERT to extract the spatial triggers and the related elements. The results demonstrate that the task of extracting fine-grained spatial information is challenging and the corpus constructed in this work can serve as an initial resource to develop methods with improved performance in the future.

Acknowledgments

This work was supported in part by the National institute of Biomedical imaging and Bioengineering (NIBIB: R21EB029575), the U.S. National Library of Medicine (NLM: R00LM012104), the Patient-Centered Outcomes Research Institute (PCORI: ME-2018C1-10963), and the Cancer Prevention Research Institute of Texas (CPRIT: RP160015).

8. Bibliographical References

- Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, and Montana G (2019). Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology*, 291(1):196–202. [PubMed: 30667333]
- Baker CF (2014). FrameNet : A Knowledge Base for Natural Language Processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore*, number 1968, pages 1–5.
- Bustos A, Pertusa A, Salinas J-M, and de la Iglesia-Vayá M. (2019). PadChest: A large chest x-ray image dataset with multi-label annotated reports. arXiv preprint arXiv:1901.07441.
- Cornegruta S, Bakewell R, Withey S, and Montana G (2016). Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. In *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis*, pages 17–27.
- Coyne B and Sproat R. (2001). WordsEye: An automatic text-to-scene conversion system. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 487–496.
- Coyne B, Sproat R, and Hirschberg J (2010). Spatial relations in text-to-scene conversion. In *Computational Models of Spatial Language Interpretation, Workshop at Spatial Cognition*, volume 620, pages 9–16.
- Coyne B, Bauer D, and Rambow O (2011). VigNet: Grounding Language in Graphics using Frame Semantics. In *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*, pages 28–36.
- Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, Thoma GR, and McDonald CJ (2016). Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310. [PubMed: 26133894]
- Devlin J, Chang M-W, Lee K, and Toutanova K (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- Guadarrama S, Riano L, Golland D, Gohring D, Jia Y, Klein D, Abbeel P, and Darrell T (2013). Grounding spatial relations for human-robot interaction. In *IEEE International Conference on Intelligent Robots and Systems*, pages 1640–1647.
- Hassanpour S and Langlotz CP (2016). Information extraction from multi-institutional radiology reports. *Artificial Intelligence in Medicine*, 66:29–39. [PubMed: 26481140]
- Hassanpour S, Bay G, and Langlotz CP (2017). Characterization of Change and Significance for Clinical Findings in Radiology Reports Through Natural Language Processing. *Journal of Digital Imaging*, 30(3):314–322. [PubMed: 28050714]
- Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpan-skaya K, Seekins J, Mong DA, Halabi SS, Sandberg JK, Jones R, Larson DB, Langlotz CP, Patel BN, Lungren MP, and Ng AY (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In *AAAI Conference on Artificial Intelligence*.
- Johnson AE, Pollard TJ, Shen L, Wei H, Lehman L, Feng M, Ghassemi M, Moody B., Szolovits P, Celi LA, and Mark RG. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035. [PubMed: 27219127]
- Kordjamshidi P, Otterlo MV, and Moens M-F (2010). Spatial Role Labeling : Task Definition and Annotation Scheme. In *Proceedings of the Language Resources & Evaluation Conference*, pages 413–420.
- Kordjamshidi P, Roth D, and Moens M-F (2015). Structured learning for spatial information extraction from biomedical text: Bacteria biotopes. *BMC Bioinformatics*, 16(1):1–15. [PubMed: 25591917]
- Kordjamshidi P, Rahgooy T, and Manzoor U (2017). Spatial Language Understanding with Multimodal Graphs using Declarative Learning based Programming. In *Proceedings of the 2nd Workshop on Structured Prediction for Natural Language Processing*, pages 33–43.

- Langlotz CP (2006). RadLex: a new method for indexing online educational materials. *Radiographics*, 26(6):1595–1597. [PubMed: 17102038]
- Petruck MR and Ellsworth M (2018). Representing Spatial Relations in FrameNet. In *Proceedings of the First International Workshop on Spatial Language Understanding*, pages 41–45.
- Pons E, Braun LM, Hunink MM, and Kors JA (2016). Natural Language Processing in Radiology: A Systematic Review. *Radiology*, 279(2).
- Rink B, Roberts K, Harabagiu S, Scheuermann RH, Toomay S, Browning T, Bosler T, and Peshock R (2013). Extracting actionable findings of appendicitis from radiology reports using natural language processing. In *AMIA Joint Summits on Translational Science Proceedings*, volume 2013, page 221.
- Roberts K, Rink B, Harabagiu SM, Scheuermann RH, Toomay S, Browning T, Bosler T, and Peshock R (2012). A machine learning approach for identifying anatomical locations of actionable findings in radiology reports. In *AMIA Annual Symposium Proceedings*, volume 2012, pages 779–788.
- Roberts K, Rodriguez L, Shooshan S, and Demner-Fushman D (2015). Automatic Extraction and PoSt-coordination of Spatial Relations in Consumer Language. In *AMIA Annual Symposium Proceedings*, volume 2015, pages 1083–1092.
- Sevenster M, Van Ommering R, and Qian Y (2012). Automatically correlating clinical findings and body locations in radiology reports using MedLEE. *Journal of Digital Imaging*, 25(2):240–249. [PubMed: 21796490]
- Si Y, Wang J, Xu H, and Roberts K (2019). Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, pages 1–8. [PubMed: 30590540]
- Steinkamp JM, Chambers C, Lalevic D, Zafar HM, and Cook TS (2019). Toward Complete Structured Information Extraction from Radiology Reports Using Machine Learning. *Journal of Digital Imaging*, 32(4):554–564. [PubMed: 31218554]
- Ulinski M, Coyne B, and Hirschberg J (2019). SpatialNet: A Declarative Resource for Spatial Relations. In *Proceedings of the Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP)*, pages 61–70.
- Wang X, Peng Y, Lu L, Lu Z, Bagheri M, and Summers RM (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471.
- Yim W-W, Denman T, Kwan SW, and Yetisgen M (2016). Tumor information extraction in radiology reports for hepatocellular carcinoma patients. In *AMIA Joint Summits on Translational Science Proceedings*, volume 2016, pages 455–64.
- Yuan Y (2011). Extracting spatial relations from document for geographic information retrieval. In *Proceedings - 2011 19th International Conference on Geoinformatics*. IEEE.

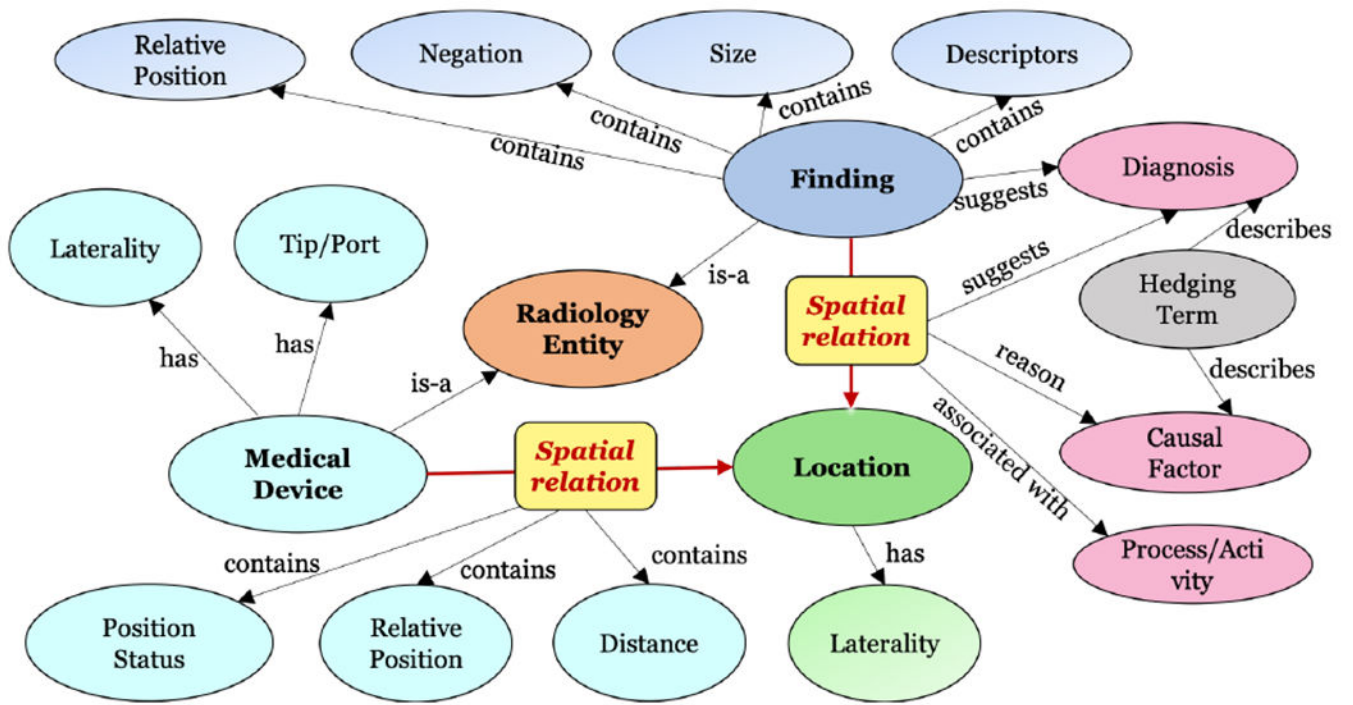


Figure 1:
Relationship between entities in Rad-SpatialNet

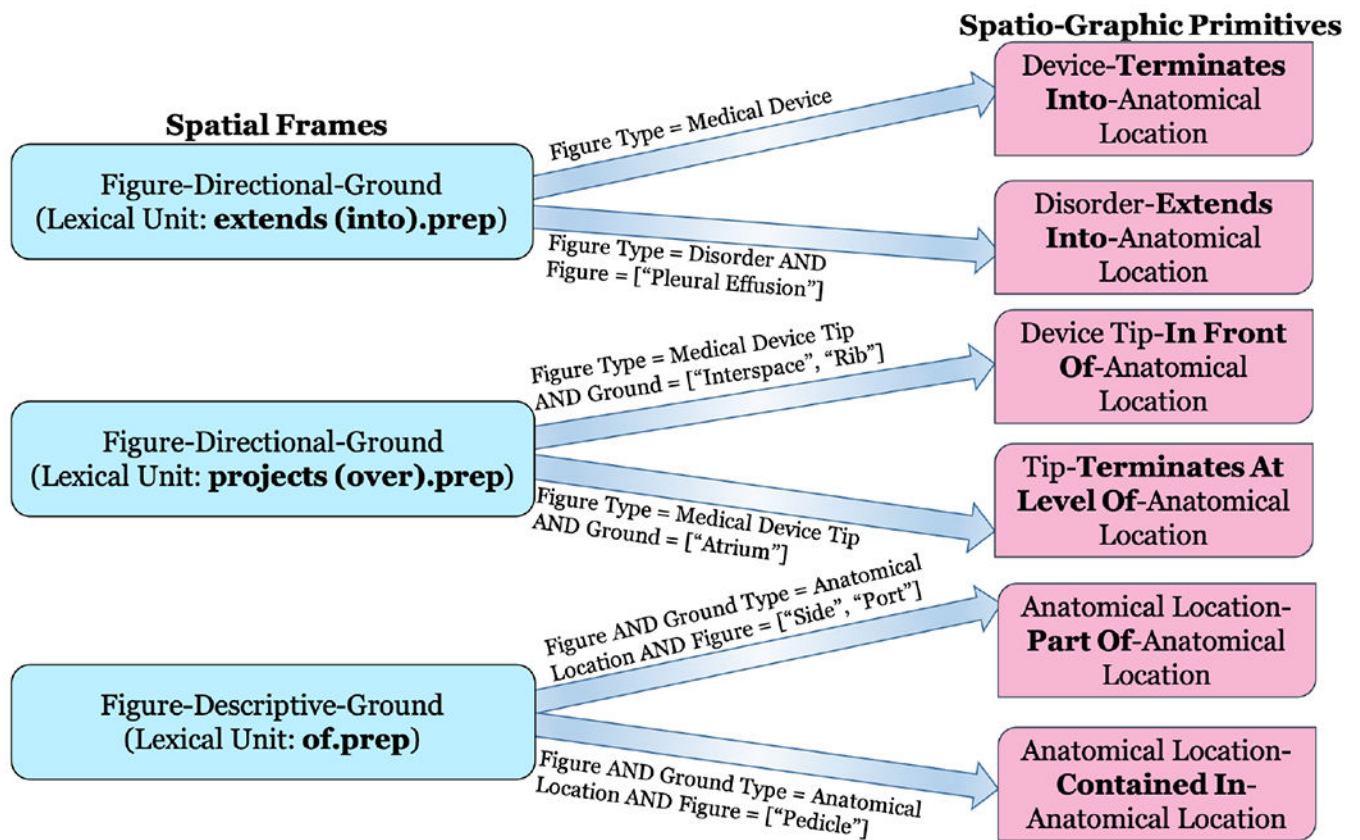


Figure 2: Examples of spatial vignettes for differentiating the spatial meanings of three commonly found spatial expressions in radiology reports.

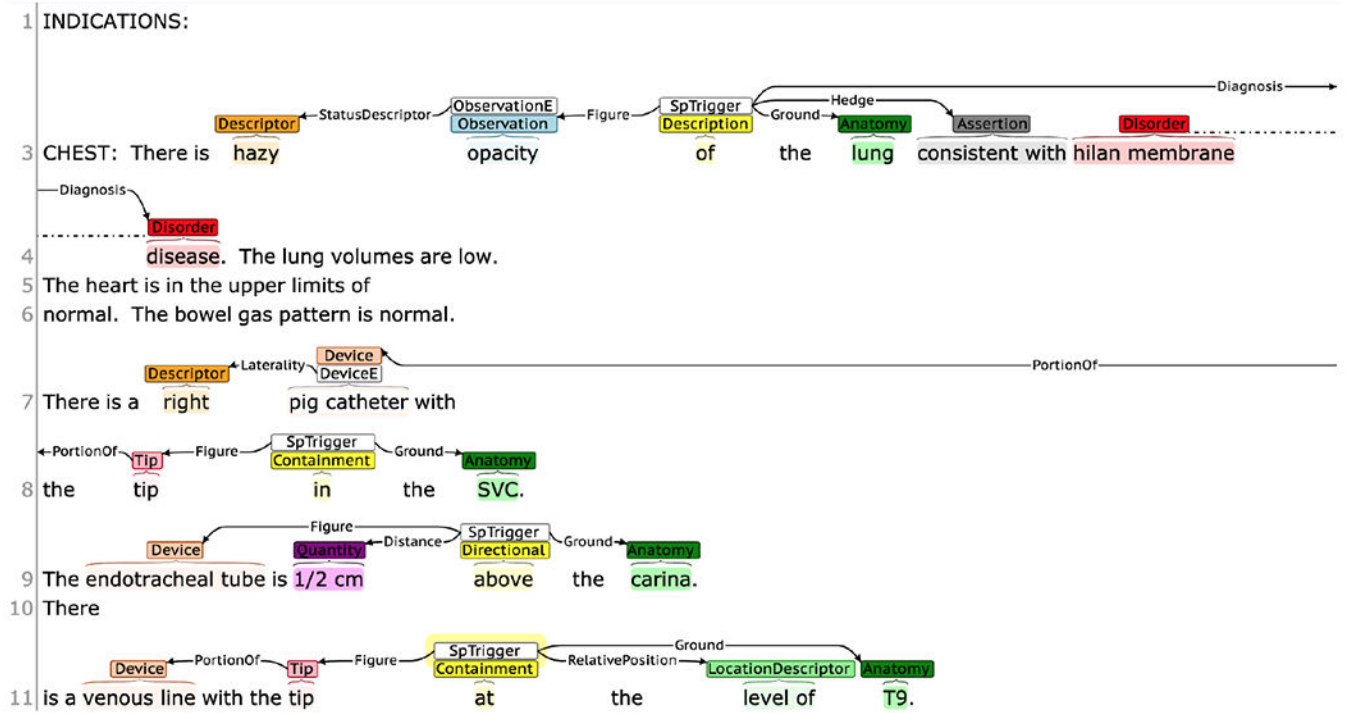


Figure 3:
Examples of annotations.

Table 1:

Example radiology report containing spatial language. Findings are in green, anatomical locations are in blue, while the spatial expressions are in cyan. *Note:* XXXX corresponds to phrases stripped by the automatic de-identifier.

Comparison: XXXX PA and lateral chest radiographs

Indication: XXXX-year-old male, XXXX.

Findings: There is XXXX increased opacity within the right upper lobe with possible mass and associated area of atelectasis or focal consolidation. The cardiac silhouette is within normal limits. XXXX opacity in the left midlung overlying the posterior left 5th rib may represent focal airspace disease. No pleural effusion or pneumothorax. No acute bone abnormality.

Impression: 1. Increased opacity in the right upper lobe with XXXX associated atelectasis may represent focal consolidation or mass lesion with atelectasis. Recommend chest CT for further evaluation.
2. XXXX opacity overlying the left 5th rib may represent focal airspace disease.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Broader categories of spatial relations in radiology

Table 2:

Relation Type	Description
Containment	Denotes that a finding/observation/device is contained within an anatomical location (“ <i>There is again seen high T2 signal within the mastoid air cells bilaterally</i> ”)
Directional	Denotes a directional sense in which a radiological entity is described wrt location (“ <i>An NGT has its tip below the diaphragm</i> ”)
Contact	Denotes an entity is in contact with an anatomical structure (“ <i>NGT reaches the stomach</i> ”)
Encirclement	Denotes a finding is surrounding an anatomical location or another finding (“ <i>Left temporal hemorrhage with surrounding edema is redemonstrated</i> ”)
Spread	Denotes traversal of an entity toward an anatomical location (“ <i>An NG tube extends to the level of the diaphragms.</i> ”)
Description	Denotes an anatomical location being described with any abnormality or observation (“ <i>There is also some opacification of the mastoid air cells.</i> ”)
Distance	Denotes a qualitative distance between a radiographic finding and an anatomical location (“ <i>There are areas of T2 hyperintensity near the lateral ventricles.</i> ”)
Adjacency	Denotes a radiographic finding is located adjacent to a location (“ <i>There is a small amount of hypodensity adjacent to the body of the right lateral ventricle.</i> ”)

Table 3:

Frame elements in Rad-SpatialNet

Element	Description
Elements with respect to a spatial trigger	
FIGURE	The object whose location is described through the spatial trigger (usually refers to finding/location/disorder/device/anatomy/typ/port)
GROUND	The anatomical location of the trajector described (usually an anatomical structure)
HEDGE	Uncertainty expressions used by radiologists (e.g., ‘could be related to’, ‘may concern for’ etc.)
DIAGNOSIS	Clinical condition/disease associated with finding/observation suggested as differential diagnoses, usually appears after the hedge related terms
REASON	Clinical condition/disease that acts as the source of the finding/observation/disorder
RELATIVE POSITION	Terms used for describing the orientation of a radiological entity wrt to an anatomical location (e.g., ‘posteriorly’ in “Blunting of the cosmittlehrehrenic sulci posteriorly is still present”, ‘high’ in “The LV line tip is high in the right atrium.”)
DISTANCE	The actual distance of the finding or device from the anatomical location (e.g., ‘1 cm’ in “ETT tube is 1 cm above the carina.”)
POSITION STATUS	Any position-related information, usually in context to a device (e.g., ‘terminates’ in “A right PIC catheter terminates in the mid SVC.”)
ASSOCIATED PROCESS	Any process/activity associated with a spatial relation (e.g., ‘intubation’ in “may be related to recent intubation”)
Elements with respect to a radiological entity	
STATUS	Indicating status of entities (e.g., ‘stable’, ‘normal’, ‘mild’)
MORPHOLOGIC	Indicating shape (e.g., ‘rounded’)
DENSITY	Terms referring to densities of findings/observations (e.g., ‘hypodense’, ‘lucent’)
MODALITY	Indicating modality characteristics (e.g., ‘attenuation’)
DISTRIBUTION	Indicating distribution patterns (e.g., ‘scattered’, ‘diffuse’)
TEMPORAL	Indicating any temporality (e.g., ‘new’, ‘chronic’)
COMPOSITION	Indicating composition of any radiological observation (e.g., ‘calcified’)
NEGATION	The negated phrase related to a finding/observation (e.g., ‘without evidence of’)
SIZE	The actual size of any finding/observation (e.g., ‘14-mm’ describing the size of a lytic lesion)
LATERALITY	Indicating side (e.g., ‘left’, ‘bilateral’)
QUANTITY	Indicating the quantity of any radiological entity (e.g., ‘multiple’, ‘few’)

Table 4:

Annotator agreement.

Item	First 50	Next 150	Last 200
Relation Types			
CONTAINMENT	0.73	0.81	0.81
DIRECTIONAL	0.29	0.73	0.46
CONTACT	0.44	0.1	0.03
ENCIRCLEMENT	0.67	0.57	0
SPREAD	0.24	0.55	0.45
DESCRIPTION	0.42	0.54	0.58
DISTANCE	0	0.5	0.4
ADJACENCY	0	0.25	0.44
Main entities			
SPATIAL TRIGGER	0.58	0.81	0.78
ANATOMY	0.48	0.68	0.76
DEVICE	0.35	0.82	0.83
TIP	0.38	0.98	0.97
FINDING/OBSERVATION	0.28	0.43	0.38
DESCRIPTORS	0.29	0.64	0.71
Frame Elements			
FIGURE	0.33	0.58	0.62
GROUND	0.42	0.67	0.70
DIAGNOSIS	0	0.51	0.54
HEDGE	0.19	0.48	0.45
REASON	0	0.38	0
RELATIVE POSITION	0.07	0.48	0.58
DISTANCE	0.4	0.86	0.71
POSITION STATUS	0	0.62	0.42
ASSOCIATED PROCESS	0.57	0	0

Table 5:

General corpus statistics.

Item	Freq
General	
Average sentence length	17.6
Spatial triggers	1372
Sentences with 1 spatial trigger	874
Sentences with more than 1 spatial trigger	227
Entity Types	
MEDICAL DEVICE	330
TIP OF DEVICE	142
DEVICE PORT/LEAD	11
IMAGING OBSERVATION	436
CLINICAL FINDING	367
DISORDER	390
ANATOMICAL LOCATION	1492
DESCRIPTOR	1548
ASSERTION	326
QUANTITY	67
LOCATION DESCRIPTOR	398
POSITION INFO	167
PROCESS	19
Spatial Frames (SFs)	
CONTAINMENT	642
DESCRIPTION	387
DIRECTIONAL	168
SPREAD	69
CONTACT	54
ADJACENCY	32
ENCIRCLEMENT	14
DISTANCE	6
Most frequent Lexical Units - Containment SF	
'in.prep'	410
'within.prep'	134
'at.prep'	86
Most frequent Lexical Units - Description SF	
'of.prep'	277
'are.prep'	37
'with.prep'	17
Most frequent Lexical Units - Directional SF	

Item	Freq
'above.prep'	43
'projecting (over).prep'	24
'below.prep'	18
Spatial Frames based on semantic types	
MEDICAL DEVICE-related	194
MEDICAL DEVICE TIP-related	142
IMAGING OBSERVATION-related	436
CLINICAL FINDING-related	344
DISORDER-related	212
Spatial frame elements	
FIGURE	1491
GROUND	1537
HEDGE	249
DIAGNOSIS	190
REASON	33
RELATIVE POSITION	398
DISTANCE	45
POSITION STATUS	167
ASSOCIATED PROCESS	21

Table 6:

Descriptive statistics of the radiological entities in the annotated corpus. (Obs - Observation, Fndg - Finding, Dis - Disorder, DEVC - Device, Aty - Anatomy)

ELEMENT	OBS	FNDG	DIS	DEVC	ATY
STATUS	231	121	77	1	22
QUANTITY	50	31	14	5	30
DISTRIBUTION	43	14	6	0	2
MORPHOLOGIC	37	20	6	0	6
SIZE DESC.	33	19	31	1	9
NEGATION	32	50	20	0	1
TEMPORAL	23	26	50	14	0
LATERALITY	20	3	22	68	499
SIZE	19	1	3	0	0
COMPOSITION	5	8	2	0	2
DENSITY	5	0	0	0	0
MODALITY	2	0	0	0	0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7:

10 fold CV results for spatial trigger extraction. P - Precision, R - Recall.

Model	P (%)	R (%)	F1
BERT _{BASE} (MIMIC)	92.20	43.04	57.52
BERT _{LARGE} (MIMIC)	93.72	67.13	77.89

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 8: 10 fold CV results for extracting spatial frame elements using BERT_{BASE} (MIMIC). P - Precision, R - Recall.

Main Frame Elements	GOLD SPATIAL TRIGGERS			PREDICTED SPATIAL TRIGGERS		
	P (%)	R (%)	F1	P (%)	R (%)	F1
FIGURE	75.14	84.10	79.35	63.26	40.48	48.00
GROUND	84.96	88.90	86.87	68.83	42.95	51.64
HEDGE	64.31	77.46	69.78	56.91	31.84	38.69
DIAGNOSIS	53.99	79.54	63.93	38.89	26.48	29.18
RELATIVE POSITION	81.27	75.56	78.12	67.01	39.91	48.81
DISTANCE	86.64	87.50	84.87	77.17	73.67	74.29
POSITION STATUS	62.29	64.07	62.74	56.86	52.05	52.39
ASSOCIATED PROCESS, REASON	0.0	0.0	0.0	0.0	0.0	0.0
OVERALL	76.53	82.69	79.48	63.98	40.73	48.55

Table 9: 10 fold CV results for extracting spatial frame elements using BERT_{LARGE} (MIMIC). P - Precision, R - Recall.

Main Frame Elements	GOLD SPATIAL TRIGGERS			PREDICTED SPATIAL TRIGGERS		
	P(%)	R(%)	F1	P(%)	R(%)	F1
FIGURE	77.42	85.75	81.35	65.51	65.44	65.12
GROUND	88.92	91.57	90.22	73.31	70.21	71.51
HEDGE	67.20	77.94	71.59	60.43	57.26	57.82
DIAGNOSIS	51.96	79.81	62.31	47.06	57.64	50.76
RELATIVE POSITION	81.31	78.42	79.57	66.02	67.76	66.33
DISTANCE	86.83	87.50	87.00	86.50	90.24	88.05
POSITION STATUS	65.73	65.83	65.38	58.59	63.63	60.37
ASSOCIATED PROCESS, REASON	0.0	0.0	0.0	0.0	0.0	0.0
OVERALL	78.83	84.64	81.61	66.63	66.28	66.25