Representation, Learning and Reasoning on Spatial Language for Downstream NLP Tasks

Parisa Kordjamshidi
Michigan State University
`kordjams@msu.edu`

James Pustejovsky
Brandeis University
`jamesp@cs.brandeis.edu`

Marie-Francine Moens
KU Leuven
`sien.moens@cs.kuleuven.be`

1 Description

This tutorial provides an overview over the *cutting edge* research on *spatial language understanding*. However, we cover some background material from various perspectives given that ACL community has not paid enough attention, in the last two decades, to this topic. There are a few emerging research work very recently looking back into the importance of spatial language in various NLP tasks. One of the essential functions of natural language is to express spatial relationships between objects. Linguistic constructs can encode highly complex, relational structures of objects, spatial relations between them, and patterns of motion through space relative to some reference point. Spatial language understanding is useful in many research areas and real-world applications. This topic recently has attracted the attention of various sub-communities in the intersection of Natural Language, Computer Vision and Robotics. The complexity of spatial language understanding and its importance in downstream tasks that involve grounding the language in the physical world has become to some extent evident to the NLP research community. Compared to other semantically specialized linguistic tasks, standardizing tasks related to spatial language seems to be more challenging as it is harder to obtain an agreeable set of concepts and relationships together with a formal spatial meaning representation that is domain independent (Pustejovsky et al., 2011; Kordjamshidi et al., 2010; Mani, 2009; Pustejovsky, 2017; Dan et al., 2020). For example, compare this with recent work on temporal relations within Computational Linguistics. This has made research results on spatial language learning and reasoning diverse, task-specific and, to some extent, not comparable. While formal meaning representation is a general issue for language understanding, formalizing spatial concepts and building formal reasoning

and machine learning models based on those constitute challenging research problems with a wealth of prior foundational work that can be exploited and linked to language understanding.

In this tutorial, we overview four themes: **1) Spatial Semantic Representation; 2) Spatial Information Extraction and; 3) Spatial qualitative representation and reasoning 4) Downstream applications of spatial semantic extraction and spatial reasoning including language grounding, robotics, navigation, dialogue systems and tasks that require combining vision and language.**

The semantic representation section covers the works that have attempted to arrive at a common set of basic concepts and relationships (Bateman, 2010; Hois and Kutz, 2011), as well as making existing corpora interoperable (Pustejovsky et al., 2011; Mani and Pustejovsky, 2012; Kordjamshidi et al., 2017; Kordjamshidi, 2013). We discuss the existing qualitative and quantitative representation and reasoning models that can be used for investigation of interoperabiltiy of machine learning and reasoning over spatial semantics (Cohn et al., 1997). Spatial language meaning representation includes research related to cognitive and linguistically motivated spatial semantic representations, spatial knowledge representation and spatial ontologies, qualitative and quantitative representation models used for formal meaning representation, and various spatial annotation schema and efforts for creating specialized corpora. We discuss various datasets that either focus on spatial annotations or downstream tasks that need spatial language learning and reasoning. Particularly, natural language visual reasoning data (Suhr et al., 2017, 2018). Moreover, continuous meaning representations for spatial concepts is another aspect to be highlighted in the tutorial, e.g., (Collell Talleda and Moens, 2018; Collell Talleda et al., 2018; Deruyt-

tere et al.).

We overview the state-of-the-art for extraction of spatial information from language, both the abstract semantic extraction (Kordjamshidi et al., 2011; Kordjamshidi and Moens, 2015) and extraction that is driven by various target tasks and applications. We discuss machine learning models including structured output prediction models, deep learning architectures and probabilistic graphical models that have been used in the related work.

Finally, we overview the usage of spatial semantics by various downstream tasks and killer applications including language grounding, navigation, self-driving cars, robotics (Tellex et al., 2011; Kollar et al., 2010), dialogue systems (Kelleher and Kruijff, 2006) and human machine interaction, and geographical information systems and knowledge graphs (Stock et al., 2013; Mai et al., 2020). Spatial semantics is very closely connected and relevant to visualization of natural language and grounding language into perception, central to dealing with configurations in the physical world and motivating a combination of vision and language for richer spatial understanding. The related tasks include: text-to-scene conversion; image captioning; spatial and visual question answering; and spatial understanding in multimodal settings (Rahgooy et al., 2018) for robotics and navigation tasks and language grounding (Thomason et al., 2018).

The current research using end-to-end monolithic deep models fail to solve complex tasks that need deep language understanding and reasoning capabilities (Hudson and Manning, 2019). Throughout this proposal, we will highlight the importance of combining learning and reasoning for spatial language understanding and its influence on the semantic representation and type of the learning models as well as the performance on various applications. Regarding the question of reasoning, we (a) point out the role of qualitative and quantitative formal representations in helping spatial reasoning based on natural language and the possibility of learning such representations from data to support compositionality and inference (Hudson and Manning, 2018; Hu et al., 2017); and (b) examine how continuous representations contribute to supporting reasoning and alternative hypothesis formation in learning (Krishnaswamy et al., 2019). We point to the cutting edge research that shows the influence of explicit representation of spatial entities and concepts (Hu et al., 2019; Liu et al., 2019).

The main goal of this tutorial is to combine these current related efforts from different communities and application domains into one unified treatment, to identify the challenges, problems and future directions for spatial language understanding.

2 Outline

The tutorial will cover the following syllabus:

- Spatial Representations
  - Linguistic corpora and semantic annotations
  - Spatial knowledge representation and spatial calculi models
  - Distributed representations

- Spatial Information Extraction
  - Spatial entity and relation extraction
  - Spatial ontology population
  - Considering domain knowledge and pragmatics in spatial extractions

- Spatial Semantic Grounding
  - Combining vision and language (symbolic and multimodal embeddings)
  - Capturing spatial common sense
  - Grounding language in 2D and 3D physical worlds
  - Generating referring expressions

- Spatial Reasoning
  - Overview on natural language and visual reasoning tasks and data
  - Modeling compositionality and spatial reasoning in (Deep) learning models

- Downstream tasks
  - Spatial concepts in dialogue systems
  - Spatial reasoning for QA and VQA
  - HRI, navigation and way-finding instructions
  - Corpus-based GIS systems

3 Prerequisites and reading list

Familiarity with machine learning and natural language processing will be helpful for this tutorial. Our selected reading list is as follows.

- Qualitative spatial representation and reasoning. Anthony G. Cohn, and Jochen Renz. Foundations of Artificial Intelligence 3 (2008): 551-596. http://dai.fmph.uniba.sk/~sefranek/kri/handbook/chapter13.pdf

- A linguistic ontology of space for natural language processing. John A. Bateman, Joana Hois, Robert Ross, and Thora Tenbrink. Artificial Intelligence 174, no. 14 (2010): 1027-1071. https://core.ac.uk/download/pdf/82158176.pdf

- Spatial Role Labeling: Task Definition and Annotation Scheme. Parisa Kordjamshidi, Marie-Francine Moens, Martijn van Otterlo, (2010). Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).

- The qualitative spatial dynamics of motion in language. James Pustejovsky, and Jessica L. Moszkowicz. Spatial Cognition Computation 11, no. 1 (2011): 15-44. http://www.cs-135.org/wp-content/uploads/2017/12/SCC-2011.pdf

- Interpreting Motion: Grounded Representations for Spatial Language. Inderjeet Mani and James Pustejovsky (2012), Explorations in language and space. Oxford University Press.

- Changing perspective: Local alignment of reference frames in dialogue, Simon Dobnik, Christine Howes, JD Kelleher, Proceedings of SEMDIAL (goDIAL), 24-32, 2015.

- Global machine learning for spatial ontology population. Parisa Kordjamshidi, Marie-Francine Moens, (2015). Journal of Web Semantics, 30, 3-21.

- VoxML: A Visualization Modeling Language. James Pustejovsky, and Nikhil Krishnaswamy. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 4606-4613. 2016. https://arxiv.org/pdf/1610.01508.pdf

- Do you see what I see? effects of pov on spatial relation specifications. Nikhil Krishnaswamy, and James Pustejovsky. In Proc. 30th International Workshop on Qualitative Reasoning. 2017. http://qrg.northwestern.edu/qr2017/papers/QR2017_paper_4.pdf

- ISO-Space: Annotating static and dynamic spatial information. James Pustejovsky (2017). In Handbook of Linguistic Annotation, pages 989–1024. Springer.

- Spatial role labeling annotation scheme. Parisa Kordjamshidi, Martijn van Otterlo, Marie-Francine Moens, (2017). In: Pustejovsky J., Ide N. (Eds.), Handbook of Linguistic Annotation Springer Verlag.

- Source-target inference models for spatial instruction understanding. Hao Tan and Mohit Bansal (2018). In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18) (5504-5511). https://arxiv.org/abs/1707.03804

- Acquiring common sense spatial knowledge through implicit spatial templates. Guillem Collell, Luc Van Gool and Marie-Francine Moens (2018). In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018) (pp. 6765-6772). AAAI. https://arxiv.org/abs/1711.06821

- Generating a Novel Dataset of Multimodal Referring Expressions. Nikhil Krishnaswamy, and James Pustejovsky. In Proceedings of the 13th International Conference on Computational Semantics, pp. 44-51. 2019. https://www.aclweb.org/anthology/W19-0507.pdf

4 Instructors

- **Parisa Kordjamshidi** is Assistant Professor of Computer Science Department at Michigan State University. Her research interests are in NLP and Machine learning. She has been working on spatial semantics extraction and annotation schemes, mapping language to formal spatial representations, spatial ontologies, structured output prediction models for information extraction, combining vision and language for spatial language understanding. She has been organizing/co-organizing shared tasks on Spatial role labeling, SpRL-2012, SpRL-2013 and the Space Evaluation workshop, SpaceEval-2015, in SemEval Series and Multimodal spatial role labeling workshop mSpRL at CLEF-2017 with the goal of

considering vision and language media for spatial information extraction and organized SpLU-2018 and Robonlp-SpLU collocated with NAACL-18 and NAACL-2019 respectively.

Contact information. Email: kordjams@msu.edu, Phone: +1-2174187004, Address: Engineering Building 428 S. Shaw Lane, East Lansing, MI 48824, USA. Webpage: `http://www.cse.msu.edu/~kordjams`.

- **James Pustejovsky** is the TJX Feldberg Chair in Computer Science at Brandeis University, where he is also Chair of the Linguistics Program, Chair of the Computational Linguistics MA Program, and Director of the Lab for Linguistics and Computation. He received his B.S. from MIT and his Ph.D. from UMASS at Amherst. He has worked on computational and lexical semantics for 25 years and is chief developer of Generative Lexicon Theory. Since 2002, he has been working on the development of a platform for temporal reasoning in language, called TARSQI (www.tarsqi.org). Pustejovsky is chief architect of TimeML and ISO-TimeML, a recently adopted ISO standard for temporal information in language, as well as the recently adopted standard, ISO-Space, a specification for spatial information in language. He has developed a modeling framework for representing linguistic expressions and interactions as multimodal simulations. This platform, VoxML, enables real-time communication between humans and computers or robots for joint tasks, utilizing speech, gesture, gaze, and action. He is currently working with robotics researchers in HRI to allow the VoxML platform to act as both a dialogue management system as well as a simulation environment that reveals realtime epistemic state and perceptual input to a computational agent. His areas of interest include: Computational semantics, temporal and spatial reasoning, language annotation for machine.

Contact Information. Email: pustejovsky@gmail.com, jamesp@cs.brandeis.edu, Phone: +1-781-736-2709, Address : Dept. of Computer Science, Brandeis University, 415 South Street, MS-018, Waltham, MA 02454,

USA. Web-page: `http://www.pusto.com`.

- **Marie-Francine Moens** is Full Professor at the Department of Computer Science, KU Leuven. She has a special interest in machine learning for natural language understanding and in grounding language in a visual context. She is holder of the prestiguous ERC Advanced Grant CALCULUS (2018-2023) granted by the European Research Council on the topic of language understanding. She is currently associate editor of the journal IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). In 2011 and 2012 she was appointed as chair of the European Chapter of the Association for Computational Linguistics (EACL) and was a member of the executive board of the Association for Computational Linguistics (ACL). From 2014 till 2018 she was the scientific manager of the EU COST action iVL Net (The European Network on Integrating Vision and Language).

Contact information. Email: sien.moens@cs.kuleuven.be, Phone: +32 16 32 83 53, Address: Department of Computer Science, KU Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium. Webpage: `https://people.cs.kuleuven.be/~sien.moens`

5 Acknowledgements

References

J. A. Bateman. 2010. Language and space: A two-level semantic approach based on principles of ontological engineering. *International Journal of Speech Technology*, 13(1):29–48.

Anthony G. Cohn, Brandon Bennett, John Gooday, and Nicholas M. Gotts. 1997. Representing and reasoning with qualitative spatial relations. In Oliviero Stock, editor, *Spatial and Temporal Reasoning*, pages 97–132. Springer.

Guillem Collell Talleda and Marie-Francine Moens. 2018. Learning representations specialized in spatial knowledge: Leveraging language and vision. *Transactions of the Association for Computational Linguistics*, 6:133–144.

Guillem Collell Talleda, Luc Van Gool, and Marie-Francine Moens. 2018. Acquiring common sense spatial knowledge through implicit spatial templates. In *Proceedings of Thirty-Second AAAI Conference*

on *Artificial Intelligence (AAAI-18)*, pages 6765–6772. AAAI.

Soham Dan, Parisa Kordjamshidi, Julia Bonn, Archna Bhatia, Martha Palmer, and Dan Roth. 2020. In *Proceedings of Language Resources and Evaluation Conference, LREC-2020*.

Thierry Deruyttere, Simon Vandenhende, Dusan Grujicic, Luc Van Gool, and Marie-Francine Moens. Talk2car: Taking control of your self driving car. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, EMNLP 2019. ACL.

Johana Hois and Oliver Kutz. 2011. Towards linguistically-grounded spatial logics. In *Spatial Representation and Reasoning in Language: Ontologies and Logics of Space*, number 10131 in Dagstuhl Seminar Proceedings. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

Ronghang Hu, Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Kate Saenko. 2017. Learning to reason: End-to-end module networks for visual question answering. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 804–813.

Ronghang Hu, Daniel Fried, Anna Rohrbach, Dan Klein, Trevor Darrell, and Kate Saenko. 2019. Are you looking? grounding to multiple modalities in vision-and-language navigation. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 6551–6557.

Drew A Hudson and Christopher D Manning. 2018. Compositional attention networks for machine reasoning. In *International Conference on Learning Representations (ICLR)*.

Drew A Hudson and Christopher D Manning. 2019. GQA: A new dataset for real-world visual reasoning and compositional question answering. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

John D. Kelleher and Geert-Jan M. Kruijff. 2006. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 1041–1048, Sydney, Australia. Association for Computational Linguistics.

T. Kollar, S. Tellex, D. Roy, and N. Roy. 2010. Toward understanding natural language directions. In *Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, HRI '10, pages 259–266. ACM.

Parisa Kordjamshidi. 2013. *Structured machine learning for mapping natural language to spatial ontologies*. Ph.D. thesis, KULeuven.

Parisa Kordjamshidi and Marie-Francine Moens. 2015. Global machine learning for spatial ontology population. *Web Semant.*, 30(C):3–21.

Parisa Kordjamshidi, Martijn van Otterlo, and Marie-Francine Moens. 2017. Spatial role labeling annotation scheme. In N. Ide James Pustejovsky, editor, *Handbook of Linguistic Annotation*. Springer Verlag.

Parisa Kordjamshidi, Martijn van Otterlo, and Marie-Francine Moens. 2010. Spatial role labeling: task definition and annotation scheme. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 413–420.

Parisa Kordjamshidi, Martijn van Otterlo, and Marie-Francine Moens. 2011. Spatial role labeling: towards extraction of spatial relations from natural language. *ACM - Transactions on Speech and Language Processing*, 8:1–36.

Nikhil Krishnaswamy, Scott Friedman, and James Pustejovsky. 2019. Combining deep learning and qualitative spatial reasoning to learn complex structures from sparse examples with noise. In *Proceedings of the thirty-third AAAI Conference on Artificial Intelligence*.

Yunchao Liu, Jiajun Wu, Zheng Wu, Daniel Ritchie, William T. Freeman, and Joshua B. Tenenbaum. 2019. Learning to describe scenes with programs. In *International Conference on Learning Representations*.

Gengchen Mai, Krzysztof Janowicz, Ling Cai, Rui Zhu, Blake Regalia, Bo Yan, Meilin Shi, and Ni Lao. 2020. Se-kge: A location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting. *Transactions in GIS*, 24(3):623–655.

I. Mani and James Pustejovsky. 2012. *Interpreting Motion: Grounded Representations for Spatial Language*. Explorations in language and space. Oxford University Press.

Inderjeet Mani. 2009. SpatialML: annotation scheme for marking spatial expression in natural language. Technical Report Version 3.0, The MITRE Corporation.

James Pustejovsky. 2017. Iso-space: Annotating static and dynamic spatial information. In *Handbook of Linguistic Annotation*, pages 989–1024. Springer.

James Pustejovsky, J. Moszkowicz, and M. Verhagen. 2011. ISO-Space: The annotation of spatial information in language. In *ACL-ISO International Workshop on Semantic Annotation (ISA'6)*.

Taher Rahgooy, Umar Manzoor, and Parisa Kordjamshidi. 2018. Visually guided spatial relation extraction from text. In *Proceedings of The 16th*

Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL HLT 2018.

Kristin Stock, Robert C. Pasley, Zoe Gardner, Paul Brindley, Jeremy Morley, and Claudia Cialone. 2013. Creating a corpus of geospatial natural language. In *Spatial Information Theory*, pages 279–298, Cham. Springer International Publishing.

Alane Suhr, Mike Lewis, James Yeh, and Yoav Artzi. 2017. A corpus of natural language for visual reasoning. In *ACL*.

Alane Suhr, Stephanie Zhou, Iris Zhang, Huajun Bai, and Yoav Artzi. 2018. A corpus for reasoning about natural language grounded in photographs. *arXiv preprint arXiv:1811.00491*.

S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.

Jesse Thomason, Jivko Sinapov, Raymond Mooney, and Peter Stone. 2018. Guiding exploratory behaviors for multi-modal grounding of linguistic descriptions. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.