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

Parisa Kordjamshidi, Michigan State University, USA, kordjams@msu.edu

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

James Pustejovsky, Brandeis University, USA, jamesp@cs.brandeis.edu

The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP-2020)

Nov 20th, 2020

Learning Reasoning

Heterogeneous

Table of Content

- Challenges and Motivating Applications
- Spatial Representations
- Spatial Reasoning
- Spatial Information Extraction
- Downstream tasks
  - (Visual) Question Answering
  - Navigation and Instruction Following
  - Dialogue Systems
  - Talking to Self-driving Cars

Spatial Language Challenges

"Hi! You are just on time! Please get me a piece of cake.

It's in the kitchen. Go out to the hall; you will see a door with a table on it. It's on the kitchen's table. A plate is under the counter, in the drawer. Utensils are next to it. There are also tissue papers above the table.

Be careful! there will be a vase on the ground on your left … .

…

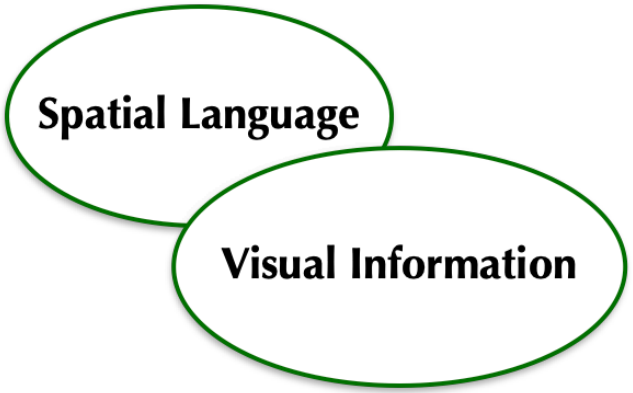
Great! You are on top of it!"

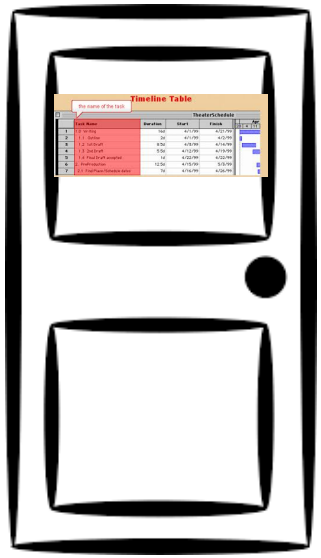
- Lexical variability
- Structural variability
- Polysemy
- Ambiguity
- Discourse and Common Sense

Spatial Language Challenges

"Hi! You are just on time! Please get me a piece of cake. It's in the kitchen. Go out to the hall; you will see a door with a table on it. It's on the kitchen's table.

A plate is under the counter, in the drawer. Utensils are next to it. There are also tissue papers above the table. Be careful! there will be a vase on the ground on your left … . Great! You are on top of it!"



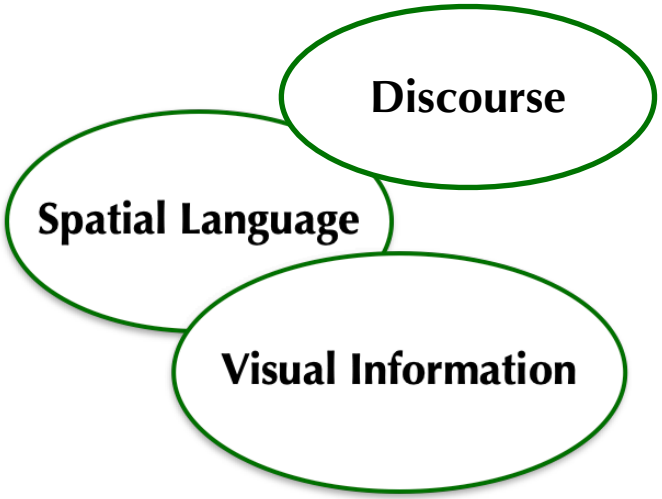


Spatial Language Challenges

"Hi! You are just on time! Please get me a piece of cake. It's in the kitchen. Go out to the hall; you will see a door with a table on it. It's on the kitchen's table.

A plate is under the counter, in the drawer. Utensils are next to it.

There are also tissue papers above the table. Be careful! there will be a vase on the ground on your left … . Great! You are on top of it!"



Spatial Language Challenges

Complex Linguistic Utterances

I: Complex locative statements

The vase is in the living room, on the table under the window.

II: Sequential scene descriptions

Behind the shops is a church, to the left of the church is the town hall, in front of the town hall is a fountain.

III: Path and route descriptions

The man came from between the shops, ran along the road and disappeared down the alley by the church.

[Barclay, Michael & Galton, Antony. (2008). A Scene Corpus for Training and Testing Spatial Communication Systems.]

Spatial Language Challenges

Implicit spatial semantics

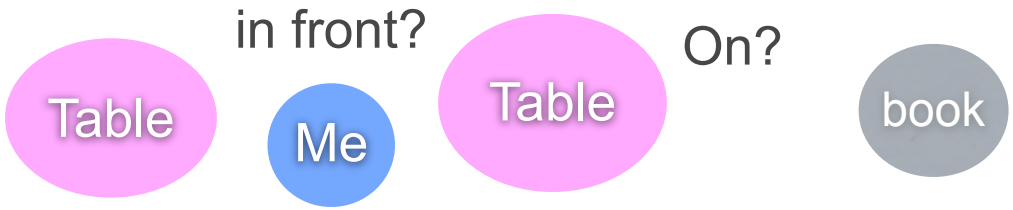
Put the milk in the coffee vs. Put the milk in the refrigerator

Fly a kite vs. Carry a kite

Spatial Language Applications

Navigation Instruction Following

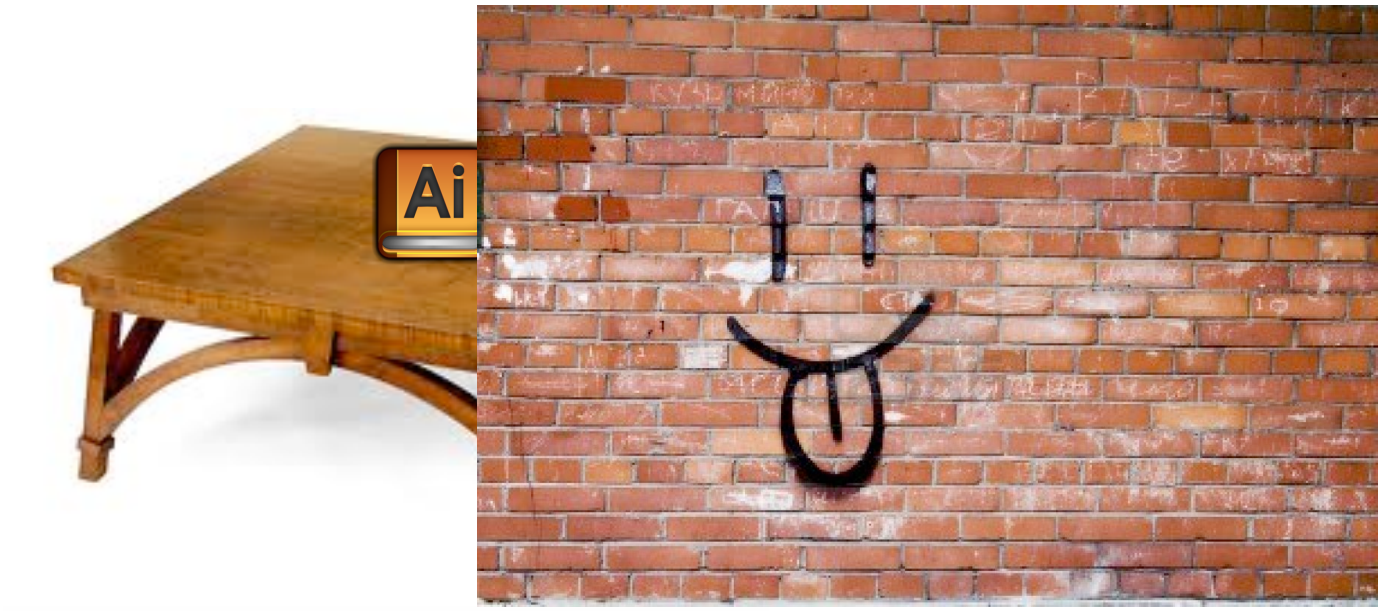
"Give me the book on AI on the big table in front of you!"



Spatial Language Applications

Text to Scene conversion (Visualization)

“The book on AI is on the big table behind the wall.”



Spatial Language Applications

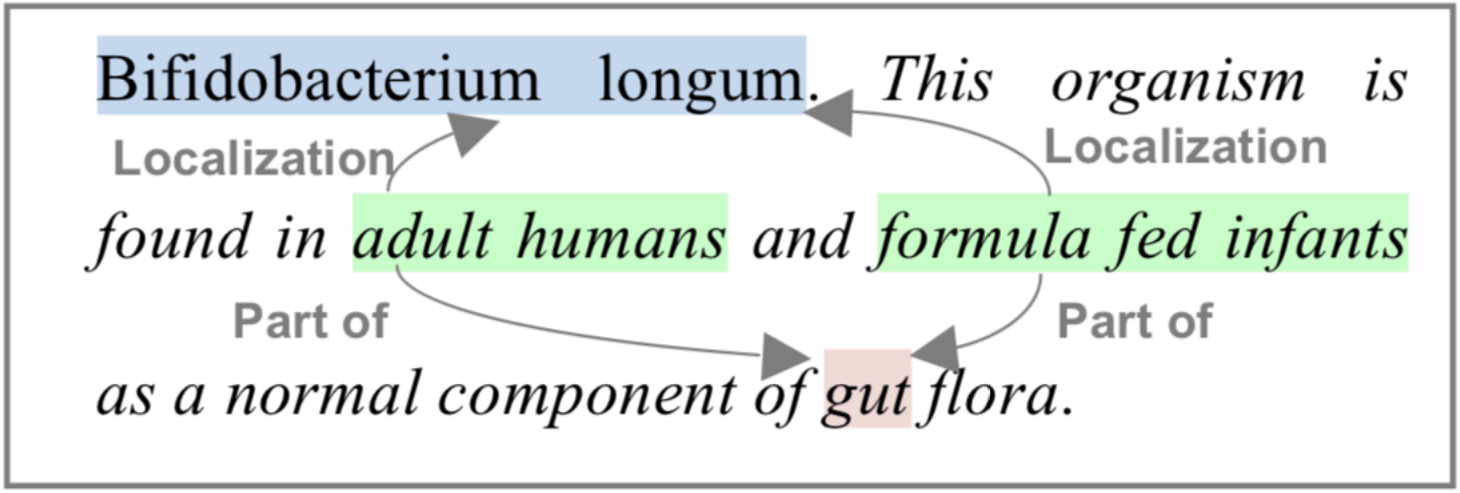
Scene to Text conversion (Image captioning)

ref: Google images, dpreview.com

Spatial Language Applications

Scientific text: Biomedical Domain

- Whether Bacteria X can live in human body?
- What are the habitats of Bacteria Y?
- What kind of Bacteria can be found in home made Yogurt that do not live in commercial Yogurt?



[Kordjamshidi, Roth, Moens,. BMC-BioInformatics. Structured learning for spatial information extraction from biomedical text: bacteria biotopes, 2015.]

Table of Content

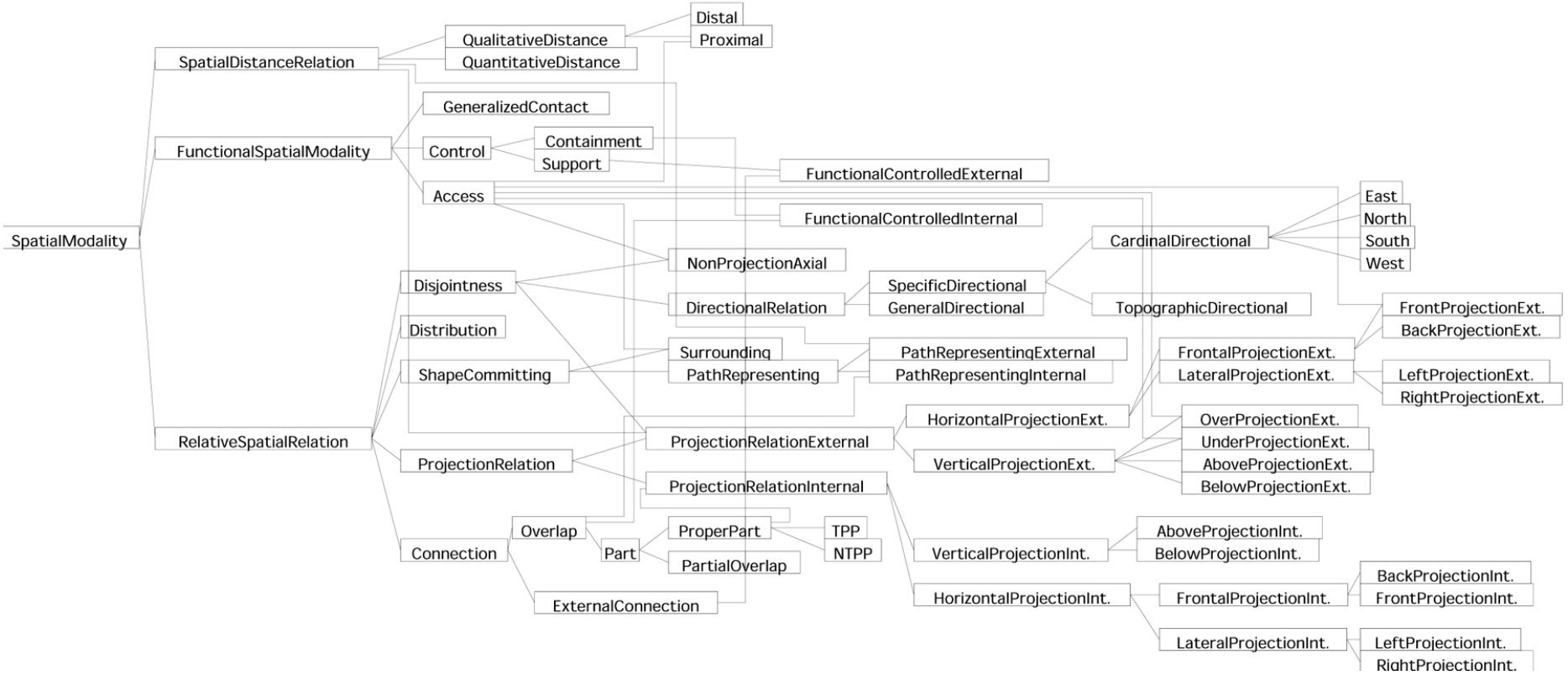
Spatial Representation

- Symbolic Semantic Representations
  - Cognitive Linguistic Conceptualizations
  - Spatial Knowledge Representation and Reasoning

- Continuous Representations
  - Learning Representations (corpora and sources of supervision)
  - Can be also cognitive linguistically motivated

Linguistically motivated representations

General Upper Model (GUM) ontology

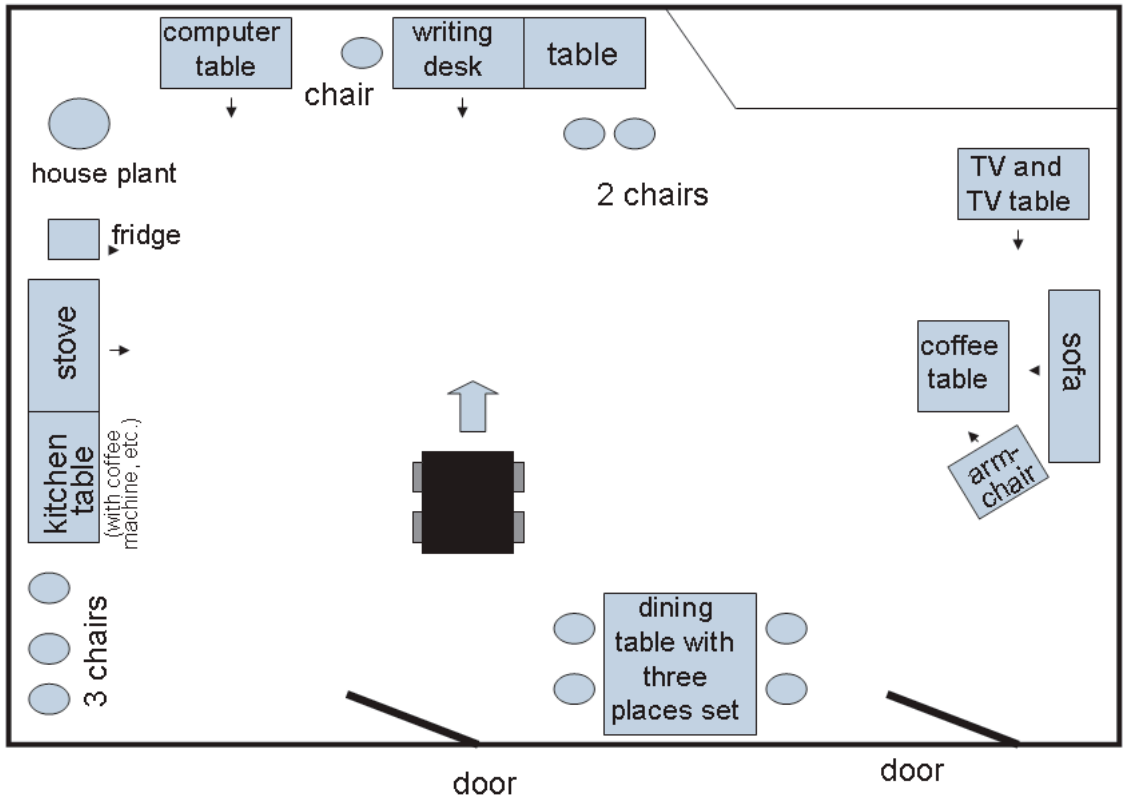


[J. A. Bateman, J. Hois, R. Ross, and T. Tenbrink. A linguistic ontology of space for natural language processing. Artificial Intelligence, 174(14):1027–1071, 2010.]

[L. Talmy, The fundamental system of spatial schemas in language, in: From Perception to Meaning: Image Schemas in Cognitive Linguistics, Mouton de Gruyter, Berlin, 2006, pp. 37–47.]

[M. Bierwisch, How much space gets into language, in: P. Bloom, M.A. Peterson, L. Nadel, M.F. Garrett (Eds.), Language and Space, MIT Press, Cambridge,MA, 1999, pp. 31–76.]

Linguistically motivated representations



1. so from here exactly opposite is my desk.

2. and next to that left of that is my computer, perhaps a meter away.

3. (breathing) ähm.

4. next to that at the wall is my kitchen, first there is my fridge all the way to the right.

[J. Bateman, T. Tenbrink, and S. Farrar. The role of conceptual and linguistic ontologies in discourse. Discourse Processes , 44(3):175–213, 2007.]

Linguistically motivated representations

so from here exactly opposite is my desk…and next to that left of that is my computer, perhaps a meter away…

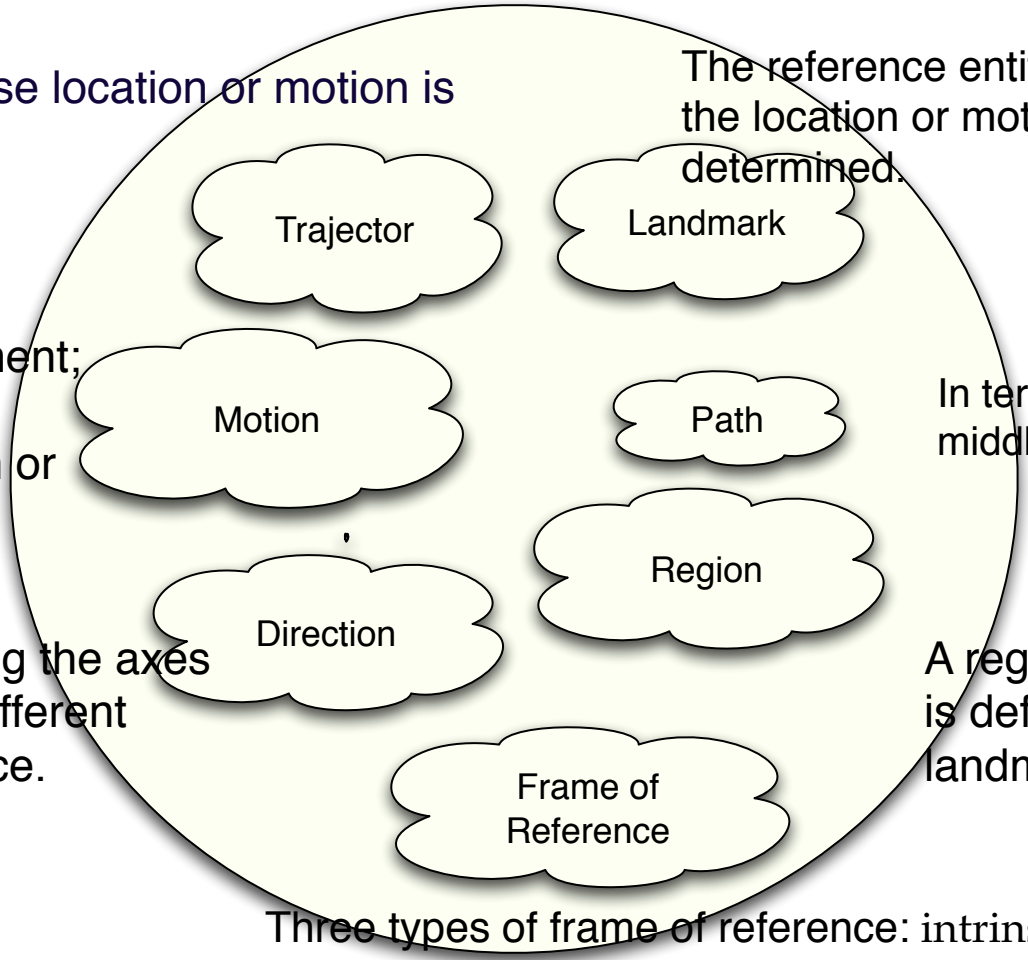
Utterance	*Locatum*	*Relatum*	*GUM Category*
1	Desk	Self	NonprojectionAxial: opposite
2	Computer	Desk	LeftProjectionExt [distance: 1m]
4	Kitchen	Computer	HorizontalProjectionExt: next
4	Kitchen	Wall	ExternalConnection: at
4	Fridge	Kitchen	RightProjectionInt: rightmost
4	Fridge	Corner	Containment: in
5	Houseplant	Corner	Containment: in
6	Stove	“There”	ExternalConnection: at
6	{Stove, kitchen table}	Fridge	HorizontalProjectionExt: side of
6–7	{Stove, kitchen table}	Fridge	LeftProjectionExt
9	Entrance	Self	BackProjectionExt
10	Dining	table	Self RightProjectionExt

SFB/TR 8 SPATIAL COGNITION

Holistic Spatial Semantics

The entity whose location or motion is of relevance.

The reference entity in relation to which the location or motion of the trajector is determined.



A binary component; whether there is perceived motion or not.

In terms of its beginning, middle and end.

The direction along the axes provided by the different frames of reference.

A region of space which is defined in relation to a landmark.

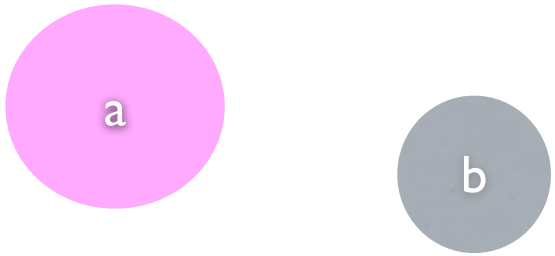
Three types of frame of reference: intrinsic, relative or absolute.

[J. Zlatev. Spatial semantics. In D. Geeraerts and H. Cuyckens, editors, The Oxford Handbook of Cognitive Linguistics , pages 318–350. Oxford Univ. Press, 2007.]

What representation is needed for spatial reasoning?

Spatial Knowledge Representation

Formal representation of the meaning! (Symbolic representations)



Disconnected?

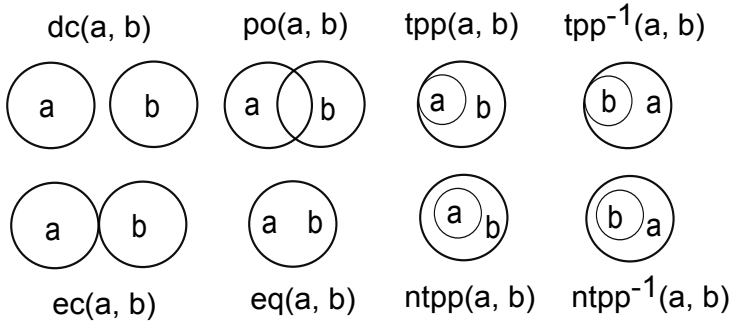
Touch?

Overlap?

Within?

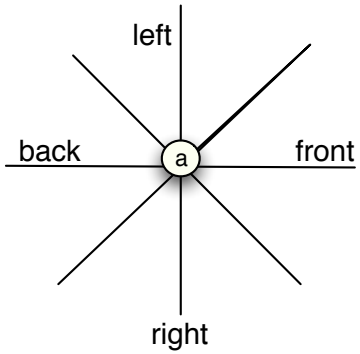
Qualitative Representation and Reasoning

- Topology (Region Connection Calculus)
- Orientation/Directions
- Distances, Sizes and Shapes



The RCC-8 relations.

Answering GIS queries: Retrieve all toxic waste dumps which are within 10 miles of an elementary school and located in Penobscot County and its adjacent counties.



[Cohn, Anthony & Hazarika, Shyamanta. (2001). Qualitative Spatial Representation and Reasoning: An Overview. Fundamenta Informaticae, 46. 1-29]

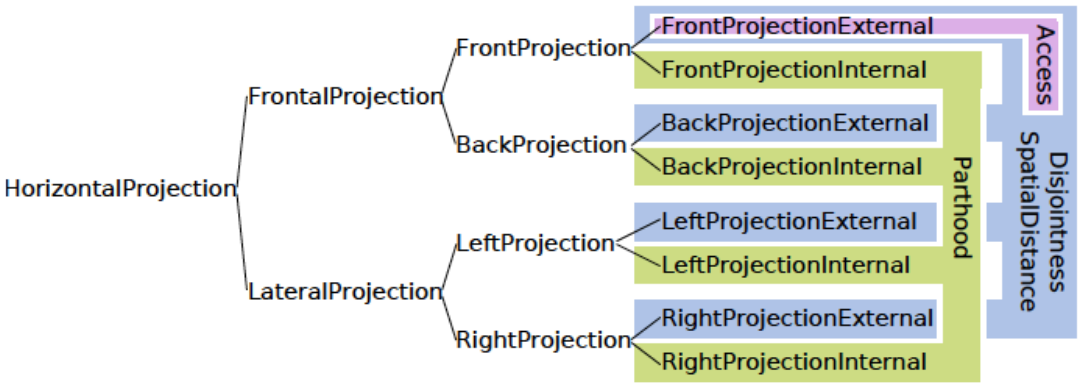
[Andrew U. Frank, Qualitative Spatial Reasoning: Cardinal Directions as an Example, Geographical Information Systems 10(3):269-290, 1996]

[Max J. Egenhofer and Robert D. Franzosa, Point-set topological spatial relations, International Journal of Geographical Information Systems, 1991]

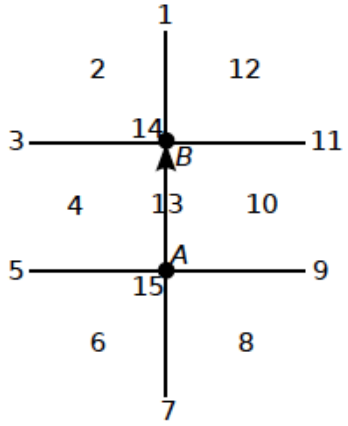
From Language to Spatial Calculi

Connections between linguistic representations and logical theories of space

- Connecting linguistically motivated ontologies like GUM to a projective spatial relations formalism, double-cross calculi.



Projective horizontal relations in GUM



DCC's 15 qualitative orientation relations

[J. Hois and O. Kutz. Natural language meets spatial calculi. In C. Freksa, N. S. Newcombe, P. Gärdenfors, and S. Wölfl, editors, Spatial Cognition VI. Learning, Reasoning, and Talking about Space , volume 5248 of LNCS, Springer, 2008.]

Moving to Corpus-based Models and Machine Learning

Corpus-based Learning and Reasoning

SpatialML:

Focused on geographical locations, annotating directional and topological relations.

[Inderjeet Mani, et, al. (2009) SpatialML: Annotation Scheme, Resources, and Evaluation, MITRE Corporation.]

Spatial Role Labeling (SpRL):

Based on holistic spatial semantics also trying to connect to multiple spatial calculi models

[Kordjamshidi, P., van Otterlo, M., Moens, M. F. (2010). Spatial role labeling: Task definition and annotation scheme. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).]

ISO-Space:

More comprehensive by considering dynamics of motion verbs and detailed properties of spatial entities.

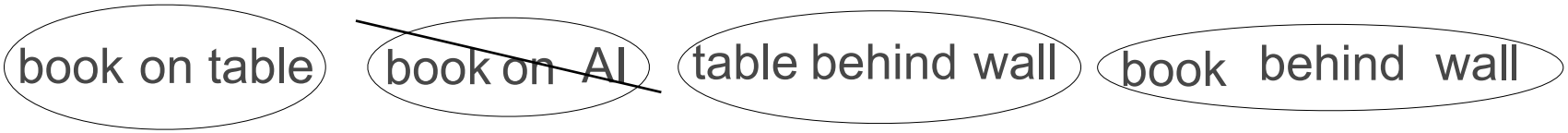
[J. Pustejovsky and J. L. Moszkowicz. Integrating motion predicate classes with spatial and temporal annotations. In Donia Scott and Hans Uszkoreit, editors, COLING 2008: Companion volume D, Posters and Demonstrations , pages 95–98, 2008.]
[J. Pustejovsky and J. L. Moszkowicz. The role of model testing in standards development: The case of iso-space. In Proceedings of LREC'12 , pages 3060–3063. European Language Resources Association (ELRA), 2012.]
[Handbook of linguistic annotation, N Ide, J Pustejovsky, Springer, 2017.]

And MORE…

Information Extraction Perspective

Extraction of spatial information to obtain a formal representation of the spatial meaning of text.

"Give me the book on AI on the big table behind the wall."



Information Extraction/ Formal Semantics

Two Layers of Semantics:

Based on cognitive linguistic elements and multiple calculi.

1. **SpRL**: Spatial role labeling

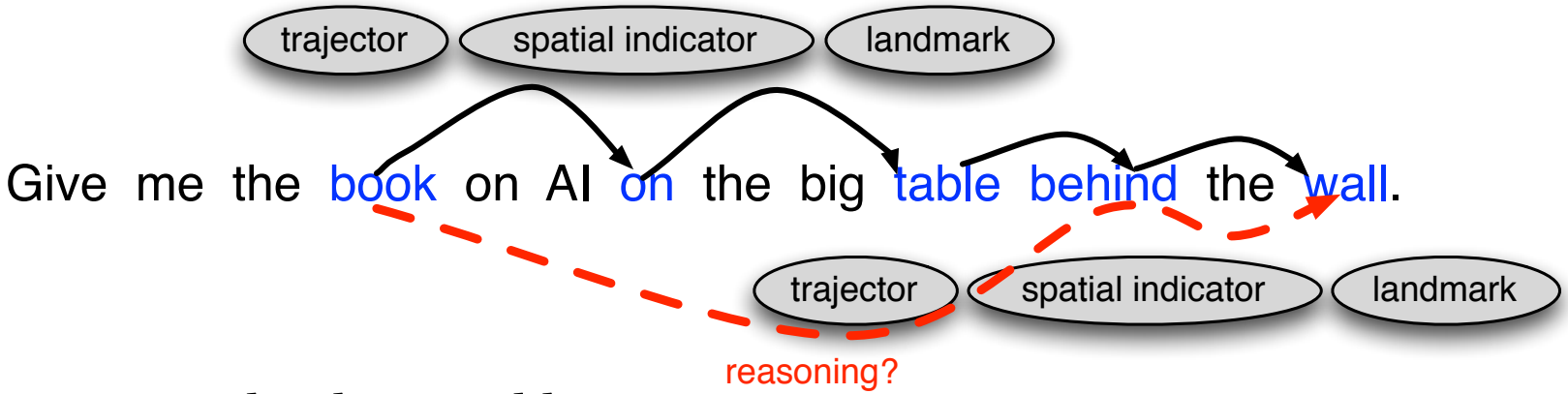
   -Identifying objects, roles and relations

2. **SpQL**: Spatial qualitative labeling

   -Identifying types of relations based on spatial calculi models

[Kordjamshidi et.al, 2012, Learning to interpret spatial natural language in terms of qualitative spatial relations Series Explorations in Language and Space.]

Spatial Role Labeling (SpRL)



$<on_{SP}\ book_{TR}\ table_{LM}>$

$<behind\ book\ wall>$

$<behind_{SP}\ table_{TR}\ wall_{LM}>$

Come over here!

Implicit roles?

$<over_{SP}\ undefined_{TR}\ here_{LM}>$

[P Kordjamshidi, M Van Otterlo, MF Moens, Spatial role labeling: Towards extraction of spatial relations from natural language ACM-Transactions in speech and language processing, 2011]

[P Kordjamshidi, P Frasconi, M Van Otterlo, MF Moens, L De Raedt, Relational learning for spatial relation extraction from natural language; International Conference on Inductive Logic Programming, ILP proceedings, LNCS, 2012]

Spatial Qualitative Labeling (SpQL)

Based on multiple calculi models

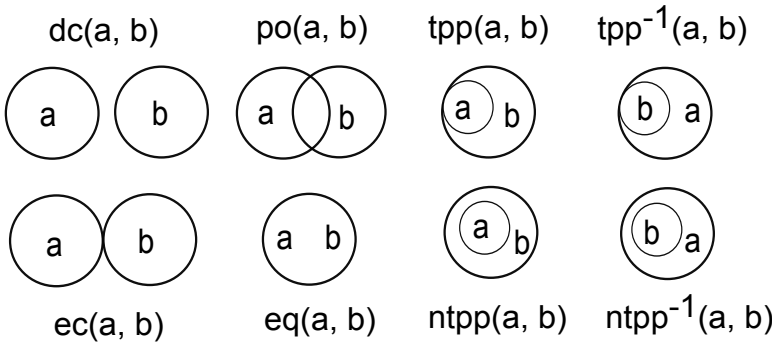
- **Topological**

{EQ, DC, EC, PO, PP}

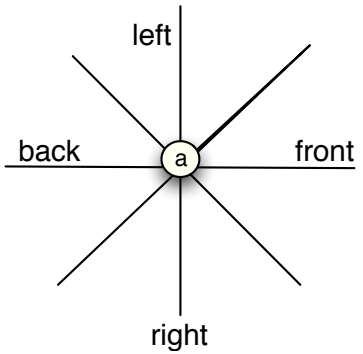
- **Directional**

{Right, Left, Above, Below, Front, Back}

- **Distal**



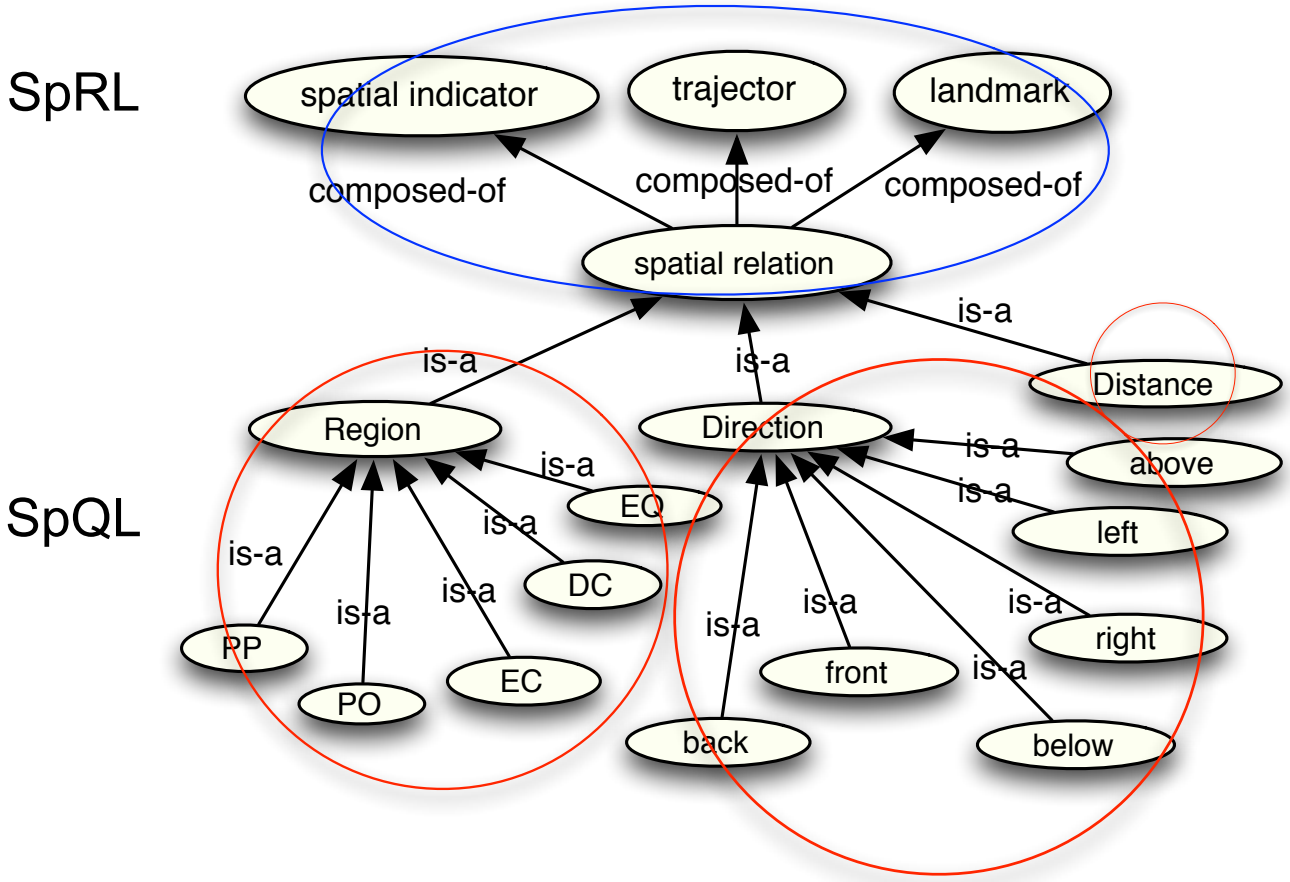
The RCC-8 relations.



[Kordjamshidi, P., van Otterlo, M., Moens, M.F.. From language towards formal spatial calculi. Computational Models of Spatial Language Interpretation Workshop (COSLI-2010) at COSIT.]

Spatial Ontology

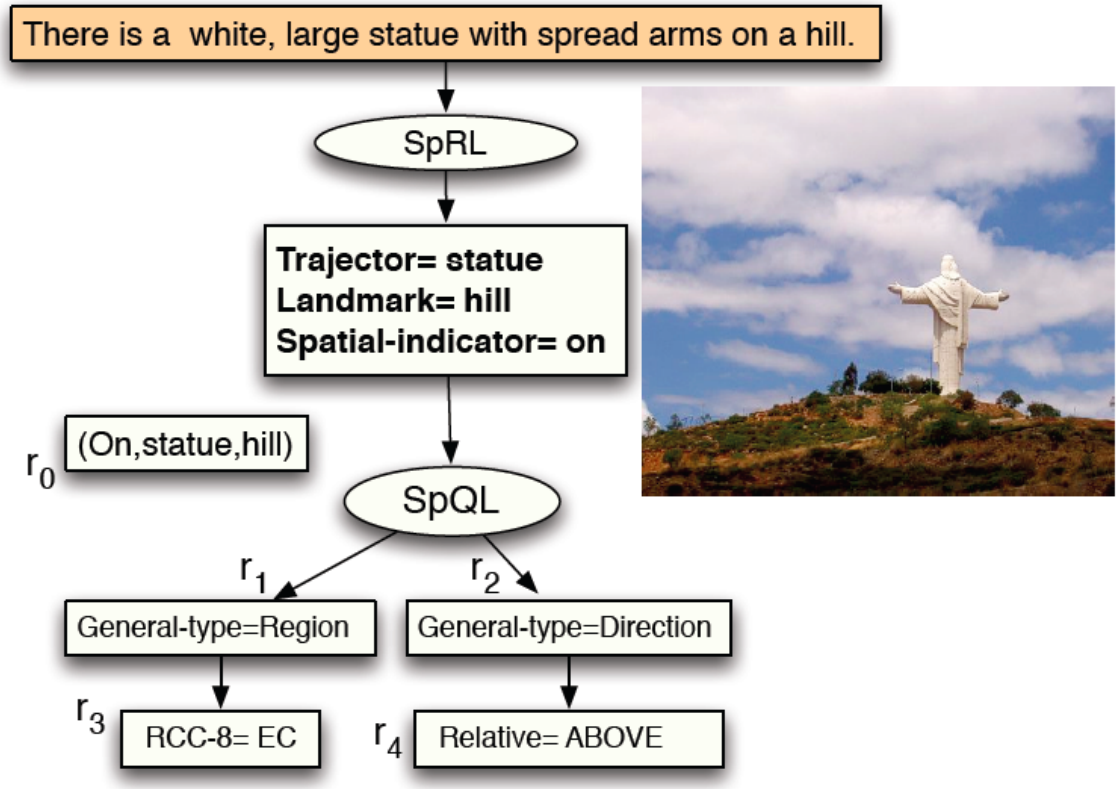
Based on cognitive linguistic elements and multiple calculi.



[Kordjamshidi, P., van Otterlo, M., Moens, M. F., Spatial role labeling: Task definition and annotation scheme. LREC-2010).]

[Kordjamshidi, P., Hois, J., van Otterlo, M., Moens, M. F., Learning to interpret spatial natural language in terms of qualitative spatial relations. Series Explorations in Language and Space. 2011.]

SpRL data



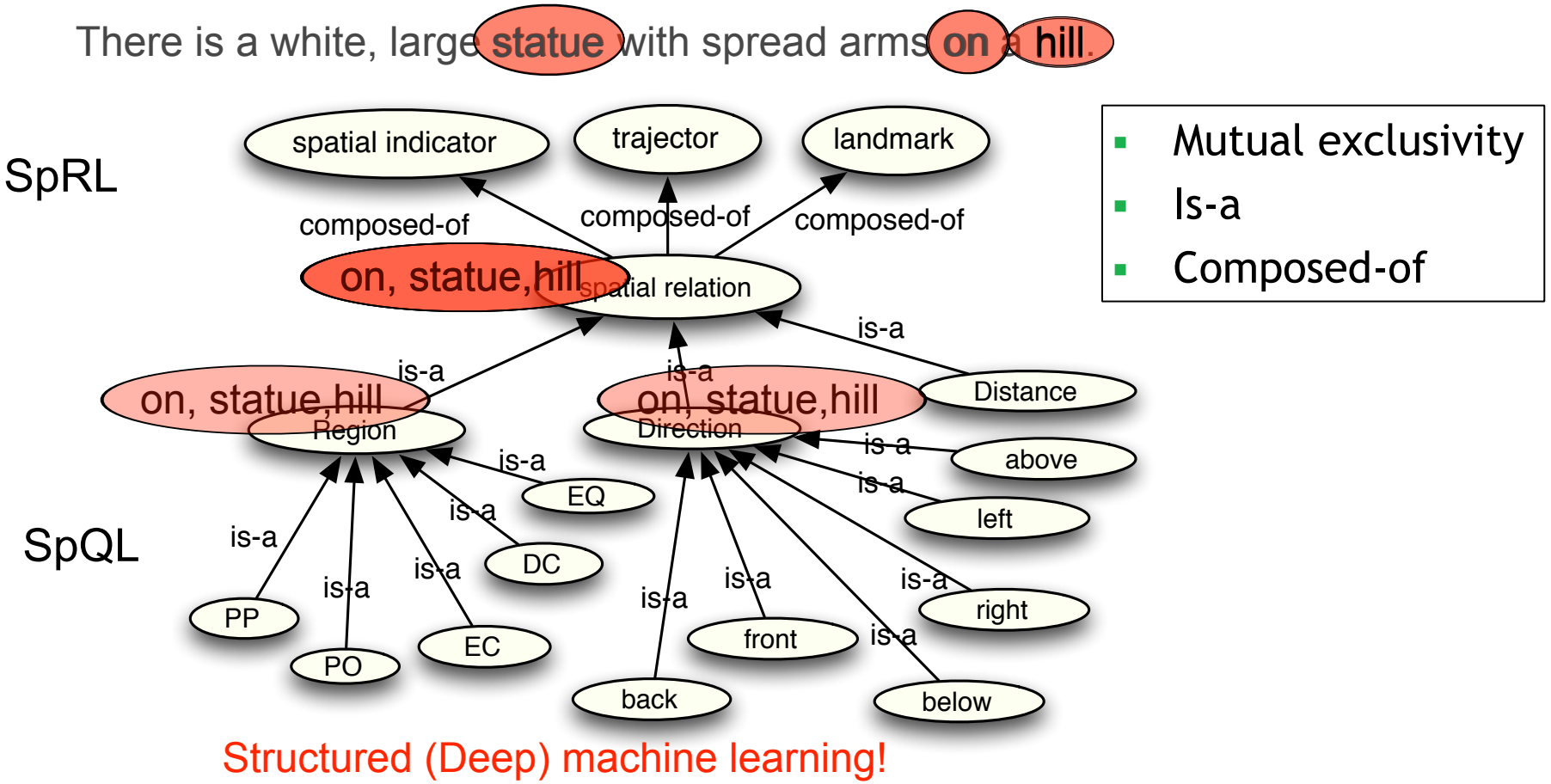
SemEval-2012/2013/2015 and CLEF/mSpRL-2017 benchmarks.

[Kordjamshidi et al. SemEval2012] [Kolomiyets, et.al. SemEval2013] [Pustejovsky et.al, SemEval2015]

[Kordjamshidi et.al. CLEF 2017: Multimodal Spatial Role Labeling (mSpRL) Task Overview.]

Exploit ontological information and structure

Semantic representation via Ontology population



[Kordjamshidi, Moens. Global machine learning for spatial ontology population; Journal of Web Semantics, 2015]

The form of global objective

Constrained Conditional Models (CCM)

- Prediction function: assign values that maximize objective

$$h(x) = \arg\max_{y \in Y(x)} g(x, y; W)$$

- Objective is linear in features and constraints

$$g = \langle W, f(x, y) \rangle - \langle \rho, c(x, y) \rangle$$

Compile everything in an Integer Linear Program: expressive enough to support decision making in the context of any probabilistic modeling.

[Roth & Yih '04, 07; Chang, et.al.,'08,'12]

The form of the global objective

Linear Constraints and the global objective function for SpRL

$$\langle W_{sp}, f_{sp}(sp_1)\rangle + \cdots + \langle W_{spf_{sp}}(sp_{SP})\rangle + \langle W_{nsp}, f_{nsp}(nsp_1)\rangle + \cdots$$
$$+ \langle W_{nspf_{nsp}}(nsp_{SP})\rangle + \langle W_{sptr}, f_{sptr}(sp_1 tr_1)\rangle + \cdots$$
$$+ \langle W_{sptr}, f_{sptr}(sp_1 tr_{TR})\rangle + \ddots$$
$$\langle W_{sptr}, f_{sptr}(sp_{SP} tr_1)\rangle + \cdots + \langle W_{sptr}, f_{sptr}(sp_{SP} tr_{TR})\rangle$$
$$+ \langle W_{splm}, f_{splm}(sp_1 lm_1)\rangle + \cdots + \langle W_{splm}, f_{splm}(sp_1 lm_{LM})\rangle + \ddots$$
$$+ \langle W_{r_\Gamma}, f_{sptrlmr_\Gamma}(sp_{SP} tr_{TR} lm_{LM} r_\Gamma)\rangle .$$

tor of
be a

Is-a relationship

$$\sum_{\gamma \in QSR_h} r_\gamma sp_i tr_j lm_k \leq 1, \quad \forall h, \quad \forall QSR_h \subset \mathcal{H}_{leafs}$$

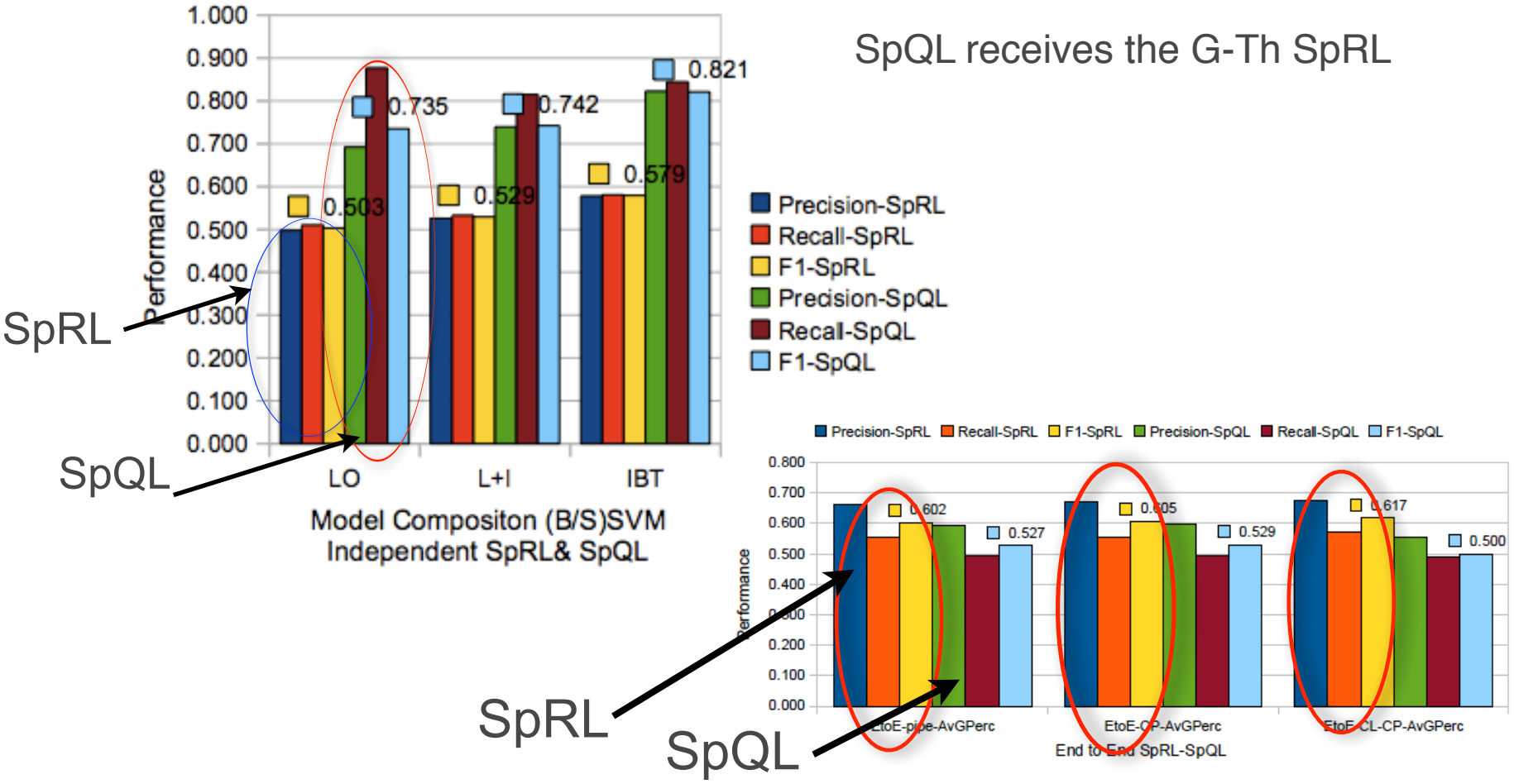
Mutual exclusivity

$$\sum_{\gamma \in \mathcal{H}_{leafs}} r_\gamma sp_i tr_j lm_k \geq r_0 sp_i tr_j lm_k.$$

[Kordjamshidi, Moens. Global machine learning for spatial ontology population; Journal of Web Semantics, 2015]

Structured Learning For Spatial Ontology Population

SpQL receives the G-Th SpRL



[Kordjamshidi, Moens. Global machine learning for spatial ontology population; Journal of Web Semantics, 2015]

Using ISO-space for Extraction

ISO Standard for Annotation of Spatial Information as expressed in Language

a. PLACES AND SPATIAL ENTITIES: natural or artificial locations in the world, as well as objects participating in spatial relations.
b. EVENTS AND MOTION EVENTS: Eventualities involving movement from one location to another.
c. SPATIAL SIGNALS AND SPATIAL MEASURES: linguistic markers that establish relations between places and spatial entities.
d. SPATIAL RELATIONSHIPS: The specific qualitative configurational, orientational, and metric relations between objects.

Spatial Relations in ISO space

Spatial Relations in ISO-Space

1. QSLINK – qualitative spatial links;
2. OLINK – orientation information;
3. MOVELINK – movement links;

DC	*the [grill] outside of the [house]*
EC	*the [cup] on the [table]*
PO	*[Russia] and [Asia]*
EQ	*[boston] and the [capital] of Massachusetts*
TPP	*the [shore] of [Delaware]*
TPPi	
NTPP	*[Austin], [Texas]*
NTPPi	
IN	*the [bookcase] in the [room]*

a. [**Boston**$_{pl1}$] is [**north of**$_{s1}$] [**New York City**$_{pl2}$].
olink(ol1, figure=pl1, ground=pl2, trigger=s1, relType="NORTH",
frame_type=ABSOLUTE, referencePt=NORTH, projective=TRUE)
b. [**The dog**$_{sne1}$] is [**in front of**$_{s2}$] [**the couch**$_{sne2}$].
olink(ol2, figure=sne1, ground=sne2, trigger=s2, relType="FRONT",
frame_type=INTRINSIC, referencePt=sne2, projective=FALSE)
c. [**The dog**$_{sne3}$] is [**next to**$_{s3}$] [**the tree**$_{sne4}$].
olink(ol3, figure=sne3, ground=sne4, trigger=s3, relType="NEXT TO",
frame_type=RELATIVE, referencePt=VIEWER, projective=FALSE)

a. [**Boston**$_{pl1}$] is [**north of**$_{s1}$] [**New York City**$_{pl2}$].
olink(ol1, figure=pl1, ground=pl2, trigger=s1, relType="NORTH",
frame_type=ABSOLUTE, referencePt=NORTH, projective=TRUE)
b. [**The dog**$_{sne1}$] is [**in front of**$_{s2}$] [**the couch**$_{sne2}$].
olink(ol2, figure=sne1, ground=sne2, trigger=s2, relType="FRONT",
frame_type=INTRINSIC, referencePt=sne2, projective=FALSE)
c. [**The dog**$_{sne3}$] is [**next to**$_{s3}$] [**the tree**$_{sne4}$].
olink(ol3, figure=sne3, ground=sne4, trigger=s3, relType="NEXT TO",
frame_type=RELATIVE, referencePt=VIEWER, projective=FALSE)

SpaceEval 2015 Tasks

Enriches SpRL (SemEval 2012)

- **SE**: Spatial Element Identification.
- **SS**: Spatial Signal Identification.
- **MS**: Motion Signal Identification.
- **MoveLink**: Motion Relation Identification.
- **QSLink**: Spatial Configuration Identification.
- **OLink**: Spatial Orientation Identification.

[James Pustejovsky, Parisa Kordjamshidi, Marie-Francine Moens, Aaron Levine, Seth Dworman, Zachary Yocum. SemEval-2015 Task 8: SpaceEval; SemEval2015 workshop.]

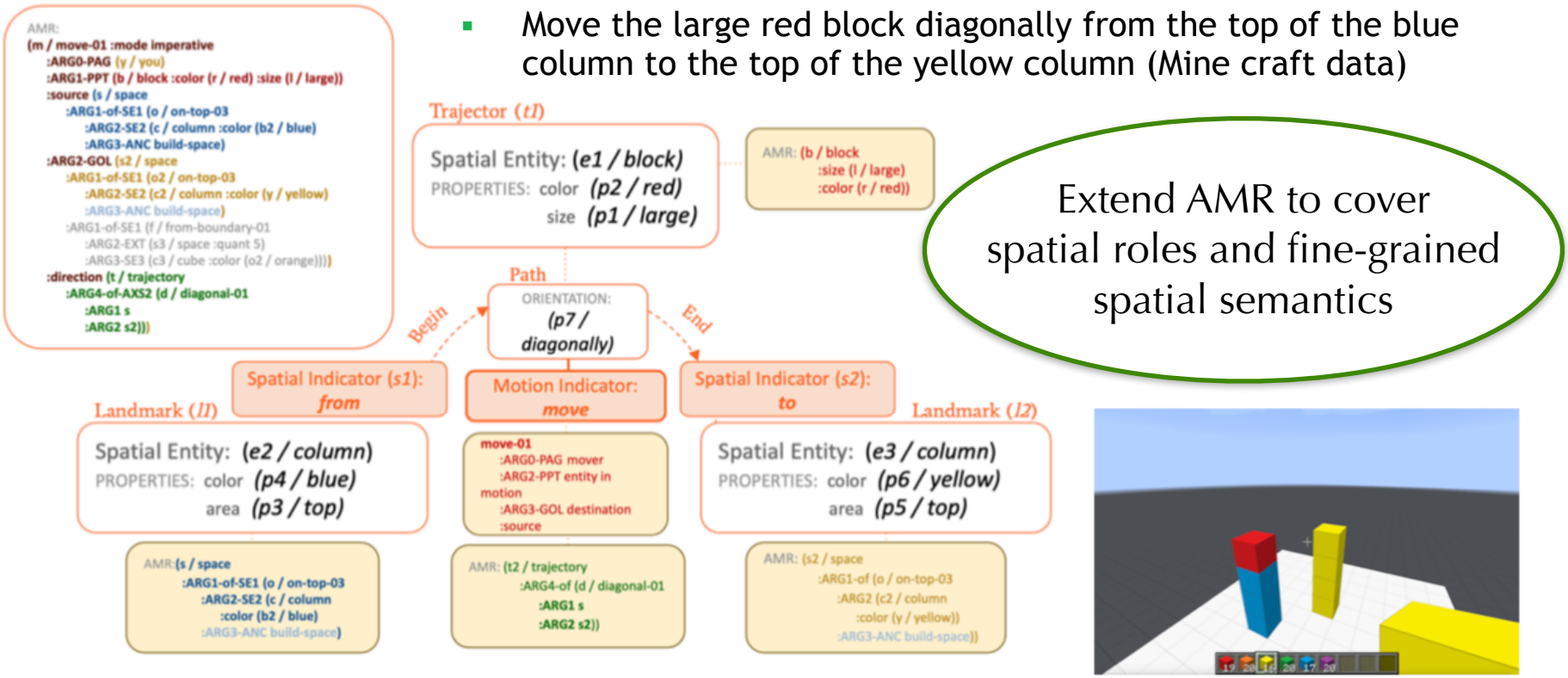
Data Resources

Annotations are applied on various datasets

- Degree Confluence Project (DCP)
- Cross Language Evaluation Forum (CLEF)
- Ride for Climate (RFC)
- Generalized Upper Model (GUM) Maptask corpus

More Resources: AMR and Spatial Roles

Abstract Meaning Representation with Spatial Roles



- Move the large red block diagonally from the top of the blue column to the top of the yellow column (Mine craft data)

Extend AMR to cover spatial roles and fine-grained spatial semantics

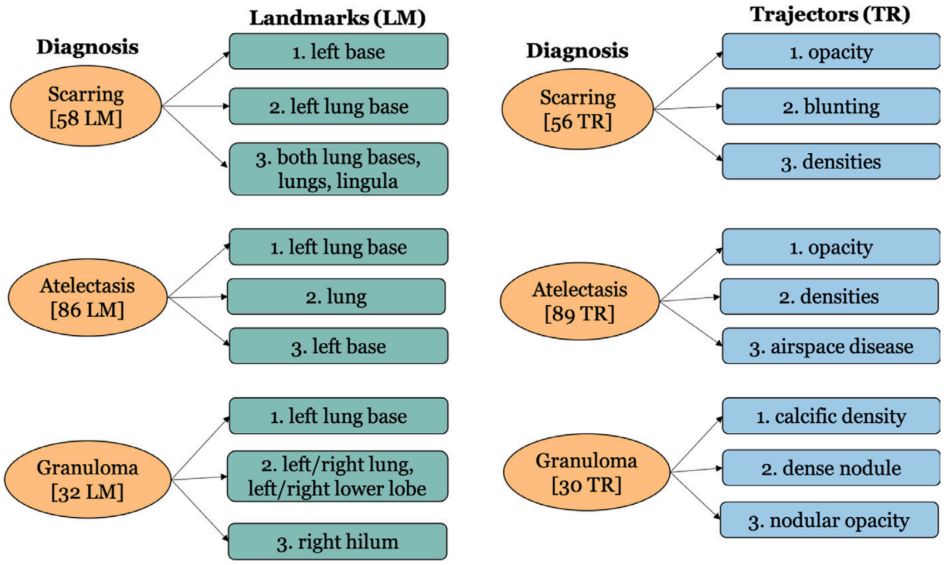
[Soham Dan, Parisa Kordjamshidi, Julia Bonn, Archna Bhatia, Zheng Cai, Martha Palmer, Dan Roth, Soham Dan, Parisa Kordjamshidi, Julia Bonn, Archna Bhatia, Zheng Cai, Martha Palmer, Dan Roth. LREC 2020.]

More Resources: Spatial Annotations in Medical Domain

Spatial roles under RadSpRLRelation

TRAJECTOR	Radiological entity (usually a radiographic finding whose position is described
LANDMARK	Anatomical location of a TRAJECTOR
DIAGNOSIS	Potential diagnosis associated with a spatial relation
HEDGE	Any uncertainty phrase used to describe a finding or diagnosis

- 2000 chest X-ray reports from a pool of 3996 de-identified reports collected from the Indiana Network for Patient Care —released by the National Library of Medicine.
- Annotations further extended and connected to spatial configurations in Rad-SpatialNet resource.



[A dataset of chest X-ray reports annotated with Spatial Role Labeling annotations, Surabhi Datta, Kirk Roberts, In J Biomed Inform, 2020]
[Rad-SpatialNet: A Frame-based Resource for Fine-Grained Spatial Relations in Radiology Reports, Surabhi Datta, Morgan Ulinski, Jordan Godfrey-Stovall, Shekhar Khanpara, Roy F. Riascos-Castaneda, Kirk Roberts, LREC 2020.]

[SpatialNet: A Declarative Resource for Spatial Relations, Morgan Ulinski, Bob Coyne, Julia Hirschberg, SpLU-RoboNLP-2019]

Using External Knowledge for Spatial Information Extraction

Combining Vision and Language for Spatial Inf. Extraction

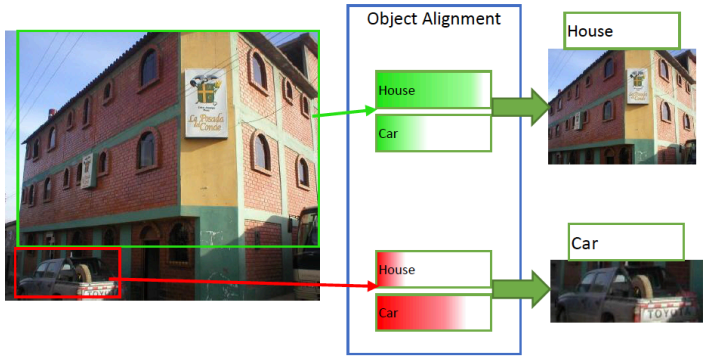
Nesting spatial constructs are a major sources of error in spatial relation extraction. *A car in front of the house on the left* can be interpreted as:
*(**A car** in front of the house) **on the left**.*

VS

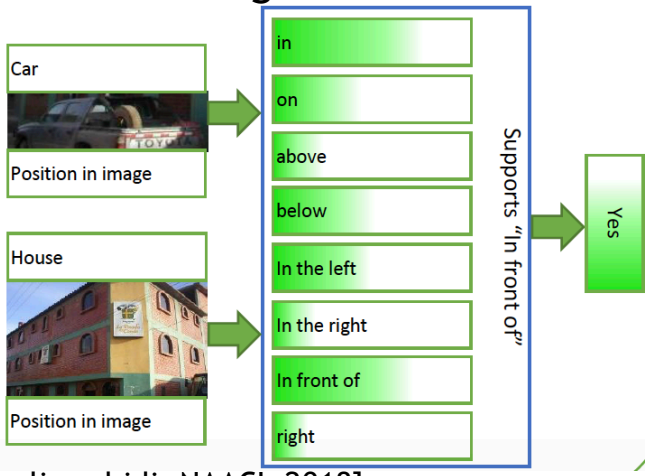
*A car in front of (**the house on the left**.)*

Image helps!

2% boost when using visual constraints!



- Align role candidates with image objects.
- For each candidate triplet check if the image supports the relation.
- Decide jointly based on image and text classifiers.



[D. Schlangen, S. Zarrieß, and C. Kennington. Resolving references to objects in photographs using the words-as-classifiers model, ACL-2016.]

[Visually Guided Spatial Relation Extraction from Text, Rahgooy, Manzoor, Kordjamshidi, NAACL-2018]

External Resources: Coreference Resolution for SpRL

"A narrow, rising street with colourful houses on both sides, among **them** a green house with balconies and a white car parked in front of it, and a blue-and-white church on the right."

Relations with pronoun landmark:

R_1 : a green house$_{tr}$, among$_{sp}$, them$_{lm}$ => "Them" is referring to "colorful houses".

R_2 : a white car$_{tr}$, in front of$_{sp}$, it$_{lm}$ => "it" is referring to "a green house".

Visual Genome Data gave another 2% boost!
Visual information can be see as a source of common sense.

[Manzoor, Kordjamshidi, Anaphora Resolution for Improving Spatial Relation Extraction from Text; NAACL, 2018, SpLU workshop]

Results

	Trajector			Landmark			Spatial indicator			Spatial triplet		
	Pr	R	F1	Pr	R	F1	Pr	R	F1	Pr	R	F1
BM	56.72	69.57	62.49	72.97	86.21	79.05	94.76	97.74	96.22	75.18	45.47	56.67
BM+C	65.56	69.91	67.66	77.74	87.78	82.46	94.83	96.86	95.83	75.21	48.46	58.94
BM+E	55.87	77.35	64.88	71.47	89.18	79.35	94.76	97.74	96.22	66.50	57.30	61.56
BM+E+C	64.40	76.77	70.04	76.99	89.35	82.71	94.85	97.48	96.15	68.34	57.93	**62.71**
BM+E+I	56.53	79.29	66.00	71.78	87.44	78.84	94.76	97.74	**96.22**	64.12	57.08	60.39
BM+E+I+C	64.49	77.92	**70.57**	77.66	89.18	**83.02**	94.87	97.61	96.22	66.46	57.61	61.72
SemEval-2012	78.2	64.6	70.7	89.4	68.0	77.2	94.0	73.2	82.3	61.0	54.0	57.3
SOP2015-10f	-	-	-	-	-	-	90.5	84	86.9	67.3	57.3	61.7

BM: Baseline, BM: Baseline, C: Constraints, E: Text Embeddings, I: Image embeddings

[Kordjamshidi, et.al., EMNLP 2017, Structured NLP workshop]

[Rahgooy, Manzoor, Kordjamshidi, NAACL-2018]

[Manzoor, Kordjamshidi, NAACL-2018, SpLU workshop]

Summary So Far

1. Spatial annotations, formal representations, connection between language and formal representations and the related resources.
2. Learning models that e**xploit ontological** and **linguistic knowledge** in learning to extract spatial information.
3. Combining with **visual information** and **join inference for spatial information extraction.**
4. **E**xternal visual resources **for pragmatics/common sense.**

Table of Content

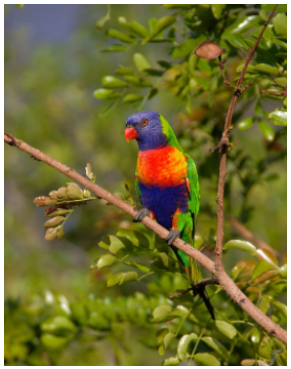
Downstream Tasks

Some tasks that involve language and vision modalities and grounding language in physical world.

- Natural Language Visual Reasoning (NLVR)
- (Visual) Question Answering (VQA)
- Navigation and Instruction Following
- …more

NLVR/VQA

The bird on the branch is looking to the left.



Natural language visual reasoning task (NLVR2)

Where is the child sitting?

fridge

arms

Visual question answering (VQA)

- Joint representations, Grounding objects and considering general relationships help.
- In these datasets, no complex spatial reasoning is needed for finding the answer.

[A Corpus for Reasoning about Natural Language Grounded in Photographs, Alane Suhr, Stephanie Zhou, Ally Zhang, Iris Zhang, Huajun Bai, Yoav Artzi, ACL-2019.]

[Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering, Yash Goyal and Tejas Khot and Douglas Summers-Stay and Dhruv Batra and Devi Parikh, CVPR-2017.]

[GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering, Hudson, Manning. CVPR-2019.]

[Chen Zheng, Quan Guo, Parisa Kordjamshidi. Cross-Modality Relevance for Reasoning on Language and Vision, Annual Conference of the Association for Computational Linguistics, ACL-2020.]

Textual Question Answering

Do we have relevant corpora to evaluate spatial meaning representations in helping downstream tasks?

- SQuAD, Hotpot QA, WiQA
- bAbi (task 17 on spatial reasoning), BoolQA

We realized these do not include complex spatial descriptions and spatial reasoning does not seem to be a key issue for solving these tasks when looking at samples of these datasets.

Spatial Question Answering

A new Benchmark: SpartQA

Formal Representations:

- Topological relations. (contains, part-of, overlap,…)
- Relative directions. (Left, Right, under, above)
- Qualitative distance. (near to, close to, far from)

The girl is on the left of the bookcase. She holds a box with a cat in it.
What is to the right of the cat? The girl or the bookcase?

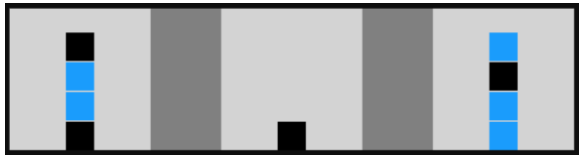
Rules of Reasoning:

- Symmetry : near to (girl, cat) -> near to (cat, girl)
- Transitivity : left (girl, bookcase) & left (cat, girl) -> left (cat, bookcase)
- Reverse : left (girl, bookcase) -> right (bookcase, girl)

[SpaRTQA: A Textual Question Answering Benchmark for Spatial Reasoning. Roshanak Mirzaee, Hossein R. Faghihi, Qiang Ning and Parisa Kordjamshidi, EMNLP-2020, Spatial Language Understanding workshop, non-arxival.]

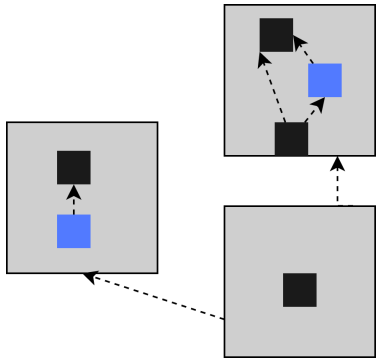
Spatial Reasoning QA dataset

Generate Dataset (SPARTQA) use Visual info and Rules of reasoning as a distant source of supervision



NLVR1 image

Random Sampling




```
[[{"y_loc":80,"size":20,"type":"square","x_loc":40,"color":"Black"},
  {"y_loc":59,"size":20,"type":"square","x_loc":40,"color":"#0099ff"},
  {"y_loc":38,"size":20,"type":"square","x_loc":40,"color":"#0099ff"},
  {"y_loc":17,"size":20,"type":"square","x_loc":40,"color":"Black"}],
```


NLVR1 scene graph (image data)

Story

We have three blocks, A, B and C. Block B is to the right of block C and it is below block A. Block A has two black medium squares. Medium black square number one is below medium black square number two and a medium blue square. It is touching the bottom edge of this block. The medium blue square is below medium black square number two. Block B contains one medium black square. Block C contains one medium blue square and one medium black square. The medium blue square is below the medium black square.

Questions

YN: Is there a square that is below medium square number two above all medium black squares that are touching the bottom edge of a block? Yes

CO: Which object is above a medium black square? the medium black square which is in block C or medium black square number two? medium black square number two

FR: What is the relation between the medium black square which is in block C and the medium square that is below a medium black square that is touching the bottom edge of a block? Left

FB: Which block(s) has a medium thing that is below a black square? A, B, C

FB: Which block(s) doesn't have any blue square that is to the left of a medium square? A, B

Improve Language Models for Spatial Reasoning

EXP1: Evaluating BERT on spatial Understanding and Reasoning.

EXP2: Fine-tune BERT on MLM task (using auto-SpartQA stories).

EXP3: Fine-tune BERT on auto-SpartQA's training set.

Model	Eval Set	FB	FR	CO	YN	Avg
EXP 1	Seen test	27.35	37.73	40.38	67.85	43.32
	Unseen test	20.5	32.64	39.71	67.74	40.14
	Human	20.21	14.67	11.57	**48.67**	23.78
EXP 2	Seen test	26.79	34.13	41.33	67.91	42.54
	Unseen test	23.2	29.94	40.72	65.82	39.92
	Human	22.12	6.4	16.81	47.57	23.22
EXP 3	Seen test	**86.85**	**85.86**	**71.47**	**78.29**	**80.61**
	Unseen test	**69.77**	**74.61**	**61.18**	**78.06**	**70.9**
	Human	**39.82**	**44.95**	**36.28**	43.2	**41.06**

Models	FB			FR			CO			YN		
	Test	Unseen	Human	Test	Unseen	Human	Test	Unseen	Human	Test	Unseen	Human
Majority	48.70	48.70	27.43	40.81	40.81	31.81	20.59	20.38	35.39	49.94	49.91	**51.94**
BERT	87.13	69.38	44.24	85.68	73.71	42.2	71.44	61.09	31.85	78.29	76.81	42.23
ALBERT	97.66	83.53	24.77	91.61	83.70	49.54	95.20	84.55	47.78	79.38	75.05	37.37
XLNet	**98.00**	**84.85**	**47.78**	**94.60**	**91.63**	**55.96**	**97.11**	**90.88**	**49.55**	**79.91**	**78.54**	36.40
Human	90.11			92.10			86.66			98.12		

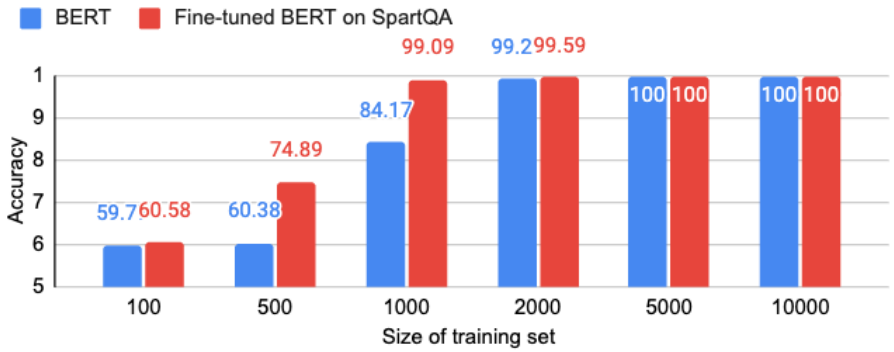
Fine-tuned LM with SpartQA

bAbI dataset (task 17)

The **pink rectangle** is to the *left* of the **red square.**

The **blue square** is to the *right* of the **red square.**

Is the **blue square** to the *left* of the **pink rectangle**? No
Is the **red square** to the *left* of the **blue square**? Yes



Model	Accuracy
Majority baseline	62.2
Recurrent model (ReM)	62.2
ReM fine-tuned on SQuAD	69.8
BERT (our setup)	71.89
ReM fine-tuned on QNLI	71.4
ReM fine-tuned on NQ	72.8
BERT fine-tuned on auto-SPARTQA	**74.18**

boolQ dataset

Q: Has the UK been hit by a hurricane?
P: The Great Storm of 1987 was a violent extratropical cyclone which caused casualties in England, France and the Channel Islands . . .
A: Yes. [An example event is given.]

Following Navigation Instructions

- Spoken dialogue describing a path through a map
- No linguistic annotations
- No alignment between text and route
- Using reinforcement learning
- State space combines linguistic features and the current location in the map, the reward is computed using the reference path

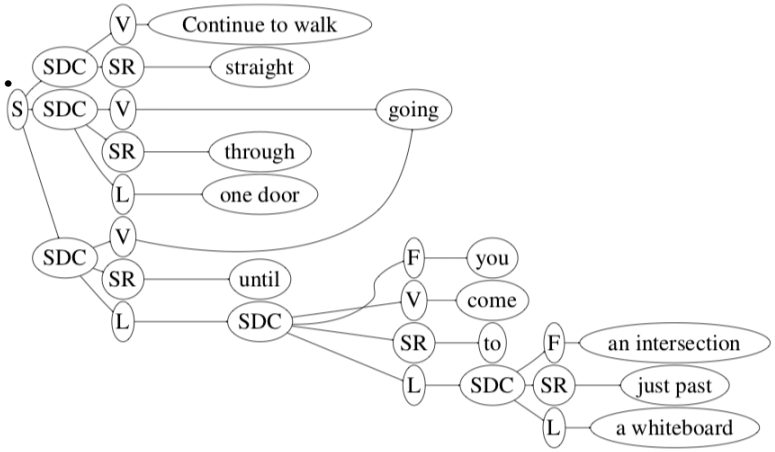
1. go vertically down until you're underneath eh diamond mine
2. then eh go right until you're
3. you're between springbok and highest viewpoint

HCRC Map task corpus

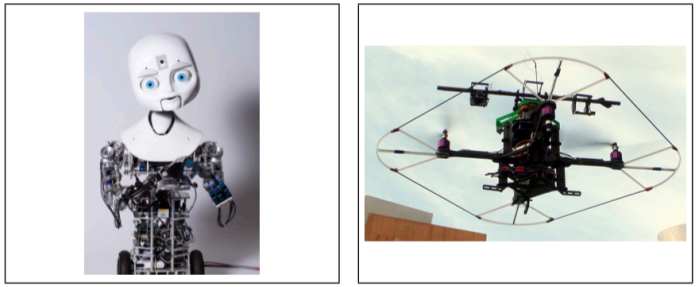
[Learning to Follow Navigational Directions , Adam Vogel and Dan Jurafsky, ACL-2010.]

Following Navigation Instructions

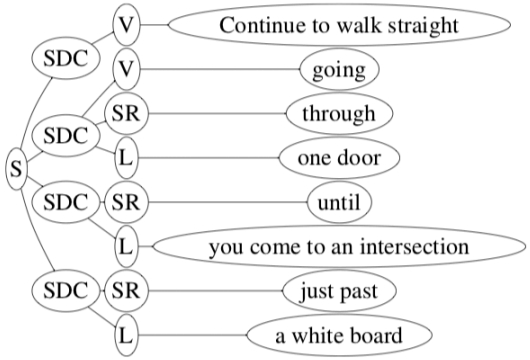
- Spatial language is represented as a hierarchy of spatial description clauses (SDC).
- SDC are hand annotated for a set of instructions.
- A discriminative probabilistic graphical models finds the most probable path by extraction of the SDCs and using the detected visual landmarks.



(a) Ground Truth



(a) humanoid (b) helicopter



(b) Automatic

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, …

[Toward Understanding Natural Language Directions, 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Thomas Kollar, Stefanie Tellex, Deb Roy, Nicholas Roy, 2010]

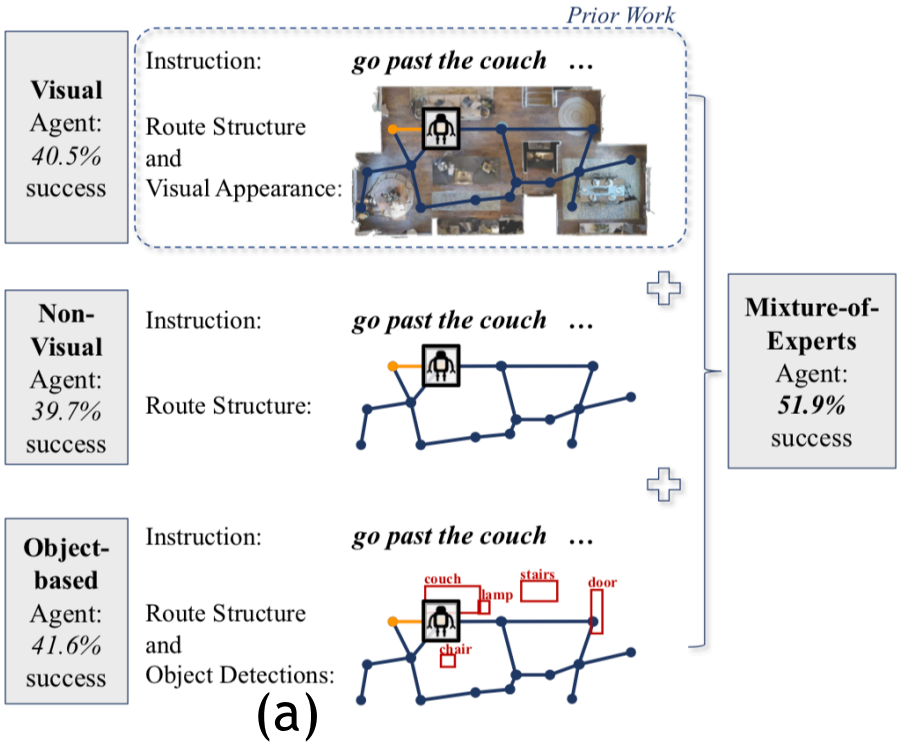
Spatial Semantics in Navigation

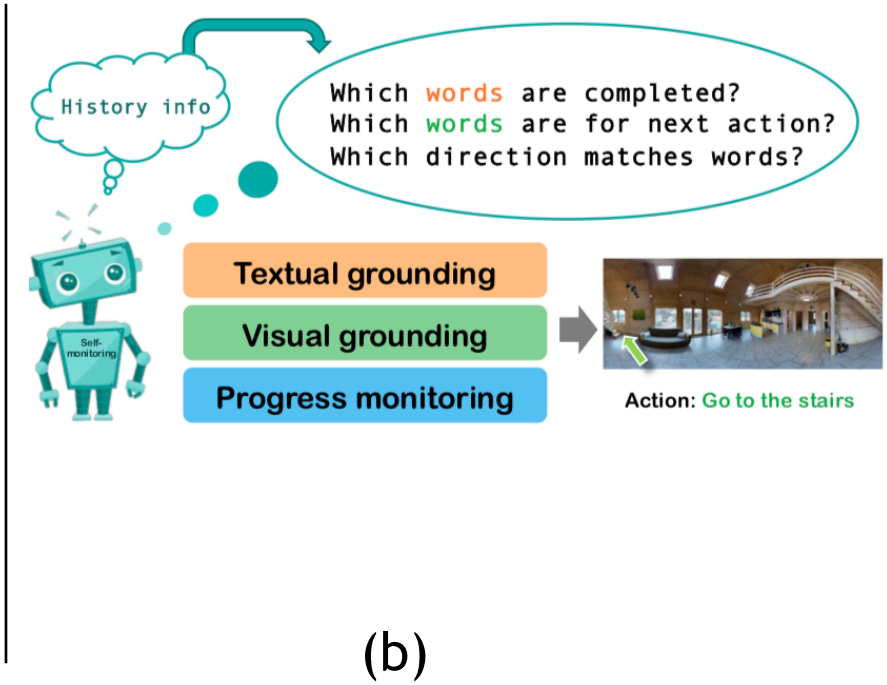
Instruction: Head upstairs and walk past the piano through an archway directly in front. Turn right when the hallway ends at pictures and table. Wait by the moose antlers hanging on the wall.

Room2Room dataset

Anderson P, Wu Q, Teney D, et al. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3674-3683.

Following Navigation Instructions





a) Are You Looking? Grounding to Multiple Modalities in Vision-and-Language Navigation, Ronghang Hu, Daniel Fried, Anna Rohrbach, Dan Klein, Trevor Darrell, Kate Saenko. ACL-2019.

b) Self-Monitoring Navigation Agent via Auxiliary Progress Estimation, Chih-Yao Ma, Jiasen Lu, Zuxuan Wu, Ghassan AlRegib, Zsolt Kira, Richard Socher, Caiming Xiong. ICLR-2019.

c) Learning to Navigate Unseen Environments: Back Translation with Environmental Dropout, Hao Tan, Licheng Yu, Mohit Bansal. NAACL-2019.

Spatial Semantics in Navigation

- Formal spatial representation (Spatial Configuration)
- Formal Representation could support reasoning capabilities and have a critical role in improving both interpretability and generalizability of deep learning models.

Go straight and pass the bar with the chair/stools then pass the clear glass table with the white chairs and turn right.

Trajector *(tr)*	The entity whose location or trans-location is described in a spatial configuration.
Landmark *(lm)*	The reference object that describes the location of the *tr* or is a part of its path of motion.
Motion Indicator*(m)*	Spatial movement usually described by a motion verb.
Frame-of-Ref.*(FoR)*	A coordinate system to identify location of an object, can be intrinsic, relative or absolute.
Path *(path)*	The *tr*'s location can be described via a path of motion instead of a basic *lm*.
Viewer *(v)*	When FoR is relative, this indicates the viewer as first, second or third person.
Spatial Indicator *(sp)*	The lexical form of the relation between the trajector and landmark.
Qualitative type *(QT)*	The qualitative/formal type of the relation between the *tr* and *lm*.

Table 2: The components of a generic spatial configuration

[Dan S, Kordjamshidi P, Bonn J, et al. From Spatial Relations to Spatial Configurations, Proceedings of The 12th Language Resources and Evaluation Conference. 2020: 5855-5864.]

Spatial Semantics in Navigation

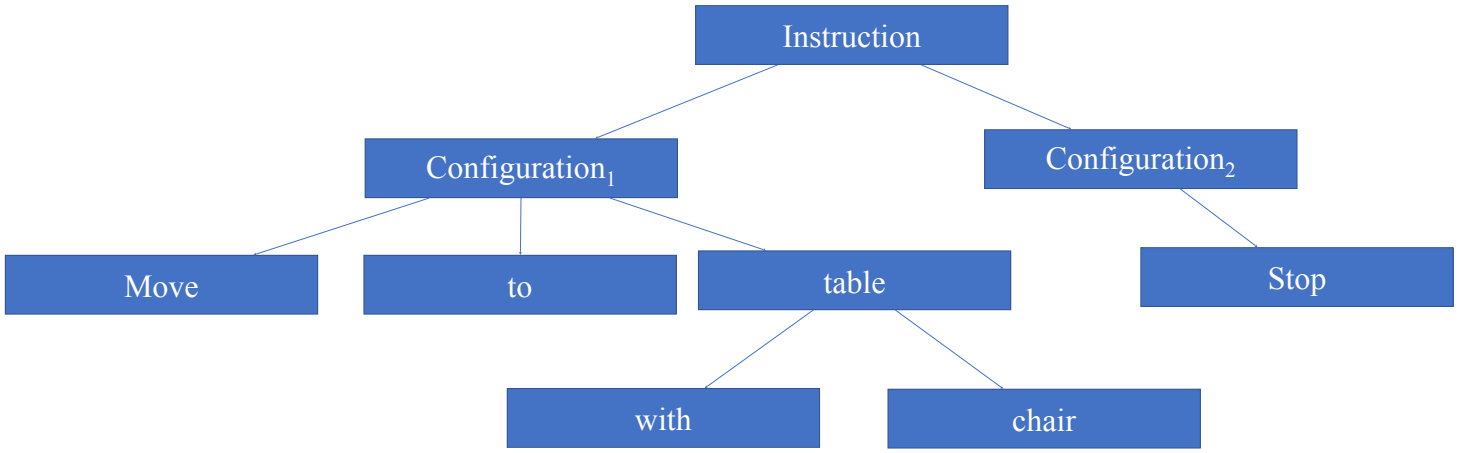
-Utilizing the spatial semantics in modeling instructions.

-Using annotation scheme based on spatial configuration.

- Automatically obtaining spatial configuration in navigation instruction.

- Automatically extracting spatial semantic components in each spatial configuration (motion indicator, spatial indicator and landmarks).

Example: Move to the table with chair.



[Vision-and-Language Navigation by Reasoning over Spatial Configurations. Yue Zhang, Quan Guo and Parisa Kordjamshidi, SpLU-2020 workshop at EMNLP, nonArxival.]

Spatial Semantics in Navigation

Design a neural network model that incorporates spatial semantic knowledge in Vision and Language Task (NLV).

- Getting representation of spatial configuration, motion indicator, spatial indicator and landmark.

- Designing a state attention (controller mechanism) that guarantees configurations are executed sequentially.

- Using the extracted landmarks to ground the objects in the image to control the state attention to decide the time of executing the next configuration (grounded configuration).

- Using the grounded configuration to attend the objects, and finally help to select the next viewpoints.

59

Experimental Results

	Validation seen		Validation unseen	
	success rate	spl	success rate	spl
Base model	0.62	0.53	0.39	0.29
Motion Indicator	0.60	0.52	0.40	0.30
Landmark	0.62	0.54	0.39	0.29
Motion indicator + Landmark + similarity	**0.65**	**0.59**	0.39	**0.32**
Self-Monitor	0.63	0.56	0.44	0.30

[Vision-and-Language Navigation by Reasoning over Spatial Configurations. Yue Zhang, Quan Guo and Parisa Kordjamshidi, SpLU-2020 workshop at EMNLP, nonArxival.]

Related Venues

Spatial Language Understanding (SpLU) workshop at EMNLP-2020:

* https://spatial-language.github.io/

A combined version of SpLU with RoboNLP will be held at ACL-2021.