A Cognitively Motivated Approach to Spatial Information Extraction

Chao Xu
Shandong University / China
chao.xu@sdu.edu.cn

Dagmar Gromann
University of Vienna / Austria
dagmar.gromann@univie.ac.at

Emmanuelle Dietz
TU Dresden / Germany
emmanuelle.dietz@tu-dresden.de

Beihai Zhou
Peking University / China
zhoubh@pku.edu.cn

Abstract

Automatic extraction of spatial information from natural language can boost human-centered applications that rely on spatial dynamics. The field of cognitive linguistics has provided theories and cognitive models to address this task. Yet, existing solutions tend to focus on specific word classes, subject areas, or machine learning techniques that cannot provide cognitively plausible explanations for their decisions. We propose an automated spatial semantic analysis (ASSA) framework building on grammar and cognitive linguistic theories to identify spatial entities and relations, bringing together methods of spatial information extraction and cognitive frameworks on spatial language. The proposed rule-based and explainable approach contributes constructions and preposition schemas and outperforms previous solutions on the CLEF-2017 standard dataset.

1 Introduction

Human-centered technologies, such as in self-driving cars or cognitive assistance systems in aviation, face the challenge of having to provide cognitively plausible decisions, i.e., decisions that are understandable and accepted by humans. This is needed for an effortless interaction between artificial and human agents. Yet, even in spatial language, with its advantageous characteristic of following recurring, uniform patterns (Talmy, 2005), drawing automated inferences in a dynamic spatial environment remains to be an open research question. As illustration consider Example (🥛)[1]: "I poured water from the bottle into the cup until it was full"[2]. Consider the question: "What was full, the cup or the bottle?". A cognitively plausible explanation for the answer *the cup*, could be as follows: "As *water was poured from the bottle into the cup*, it is likely that the water is in the cup. Normally, if *water* is poured into *the cup* then *the cup* is *full*. Thus, there is some evidence that the cup is full." In order to produce and accept "the cup" as the plausible answer, humans do not require particular reasoning effort or expert knowledge. Yet building systems that provide such *simple* explanations remains to be a challenge.

In the past, a considerable amount of literature has been published on automated spatial semantic analysis. However, most of the solutions focus on specific spatial phenomena suffering from a lack of generalizability (e.g. Platonov and Schubert, 2018; Ulinski et al., 2019), or are generalizable machine learning approaches (e.g. Kordjamshidi et al., 2011), but lack explainability of provided decisions. Also cognitive linguistics contributed to the understanding of spatial language by proposing various cognitive models to this end (Jackendoff, 1983; Talmy, 1983).

Taking spatial cognitive models as a starting point, we propose a novel spatial representation method and investigate which type of spatial information, such as entities of spatial scenes and relations holding between them, can be extracted from natural language (NL) expressions. To this end, we rely on construction grammar and cognitive linguistic theories. For Example (🥛), the following spatial roles apply: *agent* ("I"), *figure* ("water"), *ground* ("bottle" and "cup"), and simultaneously *source* ("bottle") and *goal* ("cup").

Figure 1 illustrates the automated spatial semantic analysis (ASSA) framework proposed in this paper. ASSA provides an automated extraction method based on construction grammar and image schemas (Lakoff, 1987; Johnson, 1987), which are spatio-temporal relationships built from sensori-

[1] Instead of numbering examples, we identify them with a unique semantic symbol in brackets.

[2] Adapted from Example 24 in the Winograd Schema Challenge (WSC) dataset https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WSCollection.html (24.7.2020)

Proceedings of the Third International Workshop on Spatial Language Understanding (SpLU 2020), pages 18–28
November 19, 2020. ©2020 Association for Computational Linguistics
https://doi.org/10.18653/v1/P17

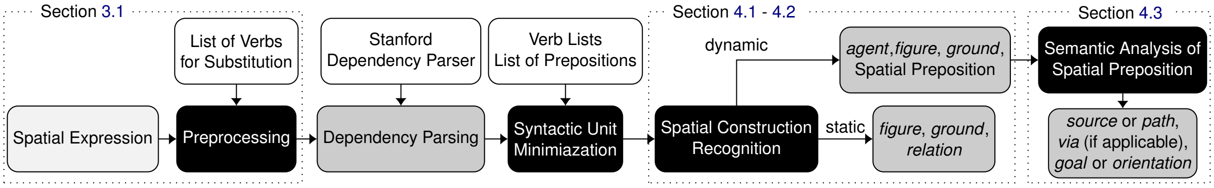

Section 3.1
List of Verbs for Substitution
Stanford Dependency Parser
Verb Lists List of Prepositions
Section 4.1 - 4.2
dynamic
agent, figure, ground, Spatial Preposition
Section 4.3
Semantic Analysis of Spatial Preposition
Spatial Expression
Preprocessing
Dependency Parsing
Syntactic Unit Minimiazation
Spatial Construction Recognition
static
figure, ground, relation
source or path, via (if applicable), goal or orientation

Figure 1: External sources are in white, input/ output are in light gray and own contributions are in black boxes.

motor experiences. Based on theories and analyses of spatial language derived from the Winograd Schema Challenge (WSC) dataset, a list of constructions to parse spatial sentences is derived (Section 3.1). We classify a list of verbs from VerbNet (Schuler, 2005) into dynamic and static, helping the classification of NL statements into dynamic and static constructions (Section 4.2). Prepositions have been found to be supreme spatial indicators in NL (Talmy, 1983; Kordjamshidi et al., 2011), so we encode spatial information of prepositions in what we call preposition schemas (Section 4.3) building on image schemas to help identifying spatial roles and their relations.

To the best of our knowledge, this is the first generalizable and explainable framework grounded in construction grammars and cognitive linguistic theories to be empirically validated. We firmly believe and show that exploiting insights gained from cognitive linguistics can help building such a framework. This framework contributes to the automated analysis of spatial semantics in NL and thereby to the analysis of the human experience of space and human-centered applications. Our contributions are: (1) A semantic analysis of spatial prepositions based on image schemas, (2) An automated extraction method that unifies constructionist theory and cognitive linguistic insights on spatial language, and (3) A prototype system of ASSA fully evaluated on a standardized dataset with competitive performance.

2 Related Work

Our work combines two strands of research, namely spatial semantic analysis and (automated) spatial role labeling. Spatial semantic analysis – the process of detecting spatial entities, their relations, and motion – has been a long-standing research endeavor in areas such as cognitive linguistics (e.g. Naidu et al., 2018; Talmy, 1983), geographical information systems (GIS) (e.g. Zhang et al., 2009; Melo and Martins, 2017), spatial language understanding in robotics (e.g. Spranger et al., 2016) as well as spatial role labeling (Kordjamshidi et al., 2011). At times, works focus on one aspect of spatial semantic analysis only, such as prepositions (Platonov and Schubert, 2018) and spatial frames and their relations (Ulinski et al., 2019).

We consider three works as particularly closely related to the proposed framework. First, Spranger et al. (2016) similarly to our work distinguish dynamic and static spatial relations, however, focus on robot-robot interactions. Second, the Embodied Construction Grammar (Bergen and Chang, 2005) brings together construction and cognitive grammar, which is, however, exemplified rather than fully evaluated. Third, Egorova et al. (2018) propose a knowledge-based approach on fictive motion, who rely on cognitive linguistic theories, rules and verbs to differentiate actual motion from fictive motion. However, they focus on fictive motion in the geographic domain, such as “The valley ran towards the sea”. Though such methods are good starting points for extracting variations of spatio-temporal information, they are less suited to reconstruct static and dynamic spatial scences from linguistic descriptions, which we propose to do with a generalizable and evaluated analysis framework.

Methods specific to spatial role labeling have mostly utilized machine learning. Kordjamshidi et al. (2011) apply a step-wise approach with two probabilistic classifiers, where the first step consists in uncovering the spatial indicator (a preposition) and the second in extracting the related trajector and landmark. While this machine learning method has a lot to offer, it disregards the influence of verbs as indicators of movement. However, verbs play a central role in identifying the relation between trajector and landmark and represent a vital element of our proposed method.

One approach testing on the same evaluation dataset, the Spatial Role Labeling shared task at CLEF-2017 (Kordjamshidi et al., 2017), is the LIP6 system (Zablocki et al., 2017). They re-implement

Saul (Kordjamshidi et al., 2015) for the two subtasks of identifying spatially-related entities (trajector, spatial indicator, landmark) and organizing them into spatial triplets. To this end, a sparse perceptron classifier is trained for each of these entities, utilizing lexical, syntactical, and contextual features. While the results are partially competitive to ours, this machine learning method requires complex feature engineering and allows for little insight into decisions taken for each entity. In contrast, ASSA is highly transparent and each decision taken can be fully explained based on the linguistic properties of the spatial expression and the assigned constructional schema and cognitive semantic roles.

3 Theoretical Foundations

Our approach is based on cognitive linguistic theories, specifically Talmy's cognitive semantic theory, Goldberg's construction grammar, and Lakoff's and Johnson's image schemas.

3.1 Talmy's Analysis of Spatial Language

Talmy (1985) introduced the following four semantic elements to characterize a motion event: *motion*, *path*, *figure*, and *ground*. A motion event consists of one object, the *figure*, moving or located w.r.t. another object, the *ground*. The *figure* refers to the entity that draws focal attention, while the *ground* refers to the entity in the periphery of attention. Consider the following sentence: "The cat was lying by the mouse hole" (🐱). The *figure* in (🐱) is "the cat", and the *ground* is "the mouse hole". Usually in English, the manner or the movement of the event is encoded by the verb, whereas the path or the spatial relation, is encoded in the preposition in a spatial expression (Talmy, 2005; Kordjamshidi et al., 2011). Prepositions as spatial indicators and verbs as movement indicators provide an excellent starting point for spatial information extraction.

Talmy lists a number of verbs that encode *path*: "enter, exit, ascend, descend, cross, pass, circle, advance, proceed, approach, arrive, depart, return, join, separate, part, rise, leave, near, follow" (Talmy, 2000b). Some of these verbs can be substituted by a verb-preposition combination with equivalent meaning. For instance, "exit" and "enter" can be substituted by "get out of" and "get into" respectively.

3.2 Goldberg's Construction Grammar

In cognitive linguistics, several construction grammars have been proposed, among others, by Fillmore et al. (1988), Goldberg (1995), and Croft (2001). For the ASSA semantic role extraction presented in Section 5, Goldberg's construction grammar was applied. Her approach focuses on the argument structure of sentence-level constructions, such as *caused-motion construction* and *ditransitive construction*. More importantly, it provides the syntactic-semantic interface between semantic roles and argument roles of an argument structure construction. One of its important principles is that the grammar of a language cannot exclusively be represented by a formal system with rules defining well-formed sequences, but rather consists of constructions. A construction is a form-meaning pair, where neither the form nor the meaning can be fully determined by its individual components, i.e., words or phrases, or other previously established constructions. For instance, "I kicked the ball into the room" (⚽), implies 'X caused Y to move Z' (*caused-motion*), where X, Y and Z, are associated with "I", "the ball", and "into the room", in (⚽), respectively. *Caused-motion* cannot be generally assigned to "kick". If it could, then "I kicked the heavy stone, but the stone did not move" would imply that "I caused the heavy stone to move, but the stone did not move", which seems unreasonable. However, we can assign the caused-motion to the construction instead.

Goldberg proposed the correspondence principle, in which correspondence describes the relations between the participant roles of the verb and the argument roles of the argument structure construction, e.g. agent, theme, location roles. For the above example, the correspondence relation between participant roles and argument roles is as follows:

Instance	I	kicked	the ball	into the room
Semantic	Agent	Action	Patient	Path
Syntactic	NP	VP	NP	PP
Grammatical	Subject	Predicate	Object	Complement

This principle provides a syntactic-semantic interface and indicates that once the construction type of a sentence is identified, we can obtain the corresponding semantic roles of syntactic elements based on their correspondence relations.

3.3 Image Schemas

Within the tradition of embodied cognition, image schemas (Johnson, 1987; Lakoff, 1987) are described as spatio-temporal relationships learned from recurring sensorimotor experiences with our environment from early infancy on that through analogical reasoning can be used to explain and predict outcomes on any current or future situation. Image schemas are believed to provide a missing link between physical experiences and high-level cognition, such as language and reasoning. For instance, if a child has learned that it can go IN and OUT of a tent – having learned the image schema of CONTAINMENT – it can transfer this information to other instances of this image schema, such as an eggshell as CONTAINER for egg yolk and white. As discussed in detail in Hedblom et al. (2018), each image schema represents a family of building blocks that can constitute different variants of the basic image schema, such as a lake can be seen as a CONTAINER with a flexible number of openings – the surface can be entered and exited at any point – whereas a tunnel has one opening to enter and one to exit. These building blocks are called spatial primitives, such as IN, OUT and CONTAINER, that compose more complex image schemas and schematic integrations (Mandler and Pagán Cánovas, 2014). For instance, variations of SOURCE_PATH_GOAL compositions could be only SOURCE_PATH or PATH_GOAL. There is no hierarchy for image schemas or the composition of spatial primitives, but in form of integrations various image schemas can come together in one spatial scene or also the lexical description of such a scene. Several spatial primitives resonate of preposition meanings, such as the image schema VERTICALITY being defined by the UP and DOWN that resonate of the prepositions "up" and "down", which represent the basis for the preposition schemas proposed in this paper.

4 Spatial Semantic Analysis Framework

The spatial semantic analysis framework covers three components depicted on the right of Figure 1: Section 4.1 specifies the semantic roles, Section 4.2 develops the spatial constructional schemas for the identification of these roles and their relations, and Section 4.3 provides the semantic analysis framework for the proposed preposition schemas.

Roles	Description
figure	Entity that needs to be located.
ground	Reference entity w.r.t. which *figure*'s location is characterized.
relation	Relative location of *figure* and *ground*.
agent	Entity that causes *figure* to move.
source	Location of *figure* before it changes location.
path	Shape of trajectory of *figure*'s movement.
goal	location of *figure* after it changes location.
via	A place on the way of *figure*'s movement.
orientation	Orientation of *figure*'s movement.

Table 1: The spatial semantic roles w.r.t. the spatial expression type.

4.1 Semantic Roles

Following Talmy's description of a motion event frame in Section 3.1, there are basic semantic roles: *figure* and *ground*. Additionally, spatial expressions can either be dynamic or static. The former describes the motion event, such as for Example (🥛), while the latter describes the state-of-affairs and does not involve movement of entities, such as for Example (🐱). For static spatial expressions, *relation* represents the spatial relations between *figure* and *ground* in Table 1. To characterize the location change in a motion event, the following semantic roles are identified as described in Table 1: *agent*, *source*, *path*, *goal*, *via*, and *orientation*. Here *path* refers to the trajectory of the *figure*'s movement, while *path* in Talmy's theory includes more information such as *source*, *via*, and *goal*, which we model separately from *path*.

4.2 Spatial Constructional Schemas

The correspondence principle introduced in Section 3.2 allows us to identify semantic roles by identifying syntactic constructions of spatial NL statements. We propose 7 constructional schemas shown in Table 2 based on a manual analysis of 78 spatial examples of the WSC dataset and 3 further ones from analyzing examples provided in literature (Sondheimer, 1978; Herskovits, 1987)[3]. Dynamic and static spatial statements are distinguished by the dynamic and stative verb in the sentence.

Each construction consists of several basic constituents. Noun phrases (NP), verbs (Verb), and prepositions (Prep) represent grammatical categories that are refined regarding their role in spa-

[3] Note that the validation of those constructions has been performed on an entirely different dataset (see Section 6).

Type	Index	Constructional Schema
	D-1	NP$_{\text{F}}$ - MV - Spatial Prep - NP$_{\text{G}}$
Dynamic	D-2	NP$_{\text{F}}$ - MV - Spatial Prep
Schemas	D-3	NP$_{\text{A}}$ - CMV - NP$_{\text{F}}$ - Spatial Prep - NP$_{\text{G}}$
	D-4	NP$_{\text{A}}$ - CMV - NP$_{\text{F}}$ - Spatial Prep
	S-1	There be - NP$_{\text{F}}$ - Spatial Prep - NP$_{\text{G}}$
	S-2	NP$_{\text{F}}$ - Spatial Prep - NP$_{\text{G}}$
Static	S-3	NP$_{\text{F}}$ - Be - Spatial Prep - NP$_{\text{G}}$
Schemas	S-4	Spatial Prep - NP$_{\text{G}}$ - Be - NP$_{\text{F}}$
	S-5	NP$_{\text{F}}$ - SV - Spatial Prep - NP$_{\text{G}}$
	S-6	Spatial Prep - NP$_{\text{G}}$ - SV - NP$_{\text{F}}$

Table 2: List of constructional schemas, where subscript “F” stands for *figure*, “G” stands for *ground*, and “A” stands for *agent*.

tial language. NPs can either represent a *figure* (NP$_{\text{F}}$), a *ground* (NP$_{\text{G}}$), or an *agent* (NP$_{\text{A}}$). We select some verbs that can appear in the constructions listed in Table 2, and classify them into three classes: *motion verbs* (MV), such as “run”, *caused motion verbs* (CMV), such as “sweep”, or *stative verbs* (SV), such as “lie”. The reason that we made such selection is that we try to exclude the non-spatial expressions in some extent by restricting the verbs, e.g., if there is no restriction on verbs, “I cut the meat into pieces” will be misrecognized as conforming to the constructional schema D-3, but it is not a motion construction. A caused motion is triggered by external force, whereas a motion exists without an external force. To further foster correct classification we treat occurrences of “to be” without any other verbs and “there is/are” as stative. The final constituent of constructional schemas are prepositions, which in the case of spatial language per definition are classified as spatial.

A verb can belong to different verb classes: *move* belongs to both, the motion verb class and the caused motion verb class, as the following examples show: *I* (NP$_{\text{F}}$) *move* (MV) *into* (Prep) *the new house* (NP$_{\text{G}}$) shows a dynamic construction where *move* functions as a motion verb. Yet, *I* (NP$_{\text{A}}$) *move* (CMV) *the box* (NP$_{\text{F}}$) *into* (Prep) *the room* (NP$_{\text{G}}$) uses the same verb as a caused motion verb. The function in the sentence determines which type of motion is being referred to. Static constructions, on the other hand, specify the static relative location of different entities. Based on these constructions we specify three general principles to identify spatial roles: (i) in a nonagentive clause, the subject and object function as *figure* and *ground* respectively; (ii) in an agentive clause, subject, direct object, and indirect object function as *agent*, *figure*, and *ground* respectively (Talmy, 2000a); (iii) a *ground* always appears after a spatial preposition. We empirically validated these principles as described below that guide our constructional schemas. The schemas in Table 2 are then utilized to recognize types and roles in linguistic expressions. These are uniquely recognized w.r.t. the syntactic structure of the spatial expression as exemplified in Section 5 (Step 4).

Relative Prepositions			
above	across	after	against
ahead of	along	alongside	amid
amidst	among	amongst	apart from
around	aside	at	atop
away from	back	back of	before
behind	below	beneath	beside
between	betwixt	beyond	but
by	close to	down	far from
in	in back of	in between	in front of
in line with	in place of	in the back of	in the front of
in the middle of	in the midst of	inside	inside of
left of	near	near to	nearby
next to	of	on	on top of
opposite	opposite of	outside	outside of
over	off	round	throughout
to the left of	to the right of	to the side of	toward
under	underneath	up	upon
within			
Dynamic Prepositions			
from	into	off	off of
on to	onto	out	out of
past	through	to	via

Table 3: List of Spatial Prepositions

4.3 Semantic Analysis of Spatial Prepositions

Spatial information is frequently encoded in spatial prepositions (Section 3.1). This section describes the contribution of prepositions to spatial role identification.

Spatial Prepositions: A spatial preposition describes the location of an entity in relation to other entities. An advantage of closed-class elements and in particular prepositions is their rather limited inventory with limited options to increase their number as opposed to other word classes. We present a fairly complete list of prepositions in Table 3 that was taken from Landau and Jackendoff (1993); Quirk et al. (2010); Dittrich et al. (2015).[4]

Spatial prepositions can be divided into two gen-

[4] There are some variations in classifications that we do not consider here: According to Landau and Jackendoff (1993) adverbs, such as *backward* and *forward*, are intransitive prepositions; Dittrich et al. (2015) classify phrases such as *west of* as prepositions.

eral types depending on the constructions in which they appear: relative prepositions and dynamic prepositions. Relative prepositions are the ones that spatially relate two or more entities, and can be used in both dynamic and static constructions. Dynamic prepositions co-occur with movement and are frequently used in dynamic constructions.

The reason we made such distinction is to distinguish Example (🥛) from: "I poured water from the bottle on the table". In this example, the corresponding *figure* of 'on the table' is 'the bottle' rather than 'water', while the *figure* of 'into the cup' is 'water' rather than 'the bottle' in Example (🥛). For the example, the first part (i.e., "I poured water from the bottle") should be recognized by dynamic constructional schema D-3, and the second part (i.e., "the bottle on the table") should be recognized by the static constructional schema S-2 in Table 2. These two examples mainly differ in their use of prepositions, namely *into* and *on*. Given a spatial expression, if its first part is recognized by a dynamic construction, and the preposition phrase (PP) after the first part describes a movement, namely using dynamic prepositions, then the *figure* of this PP is the same as the *figure* of the first part, and further roles are determined according to Table 4. Otherwise, this PP and the NP before it are recognized by another constructional schema.

Image Schemas and Semantic Analysis: Following Section 3.3, we assume that spatial primitives of image schemas can help understand and classify spatial prepositions. Figure 2 depicts a number of spatial primitives to characterize spatial relations expressed by prepositions. These schemas are utilized to identify corresponding semantic roles of spatial prepositions, and most of them come from the previous research on image schema and semantic analysis of prepositions such as Johnson (1987); Herskovits (1987); Coventry and Garrod (2004).

For instance, the Into primitive describes the dynamic spatial relation of an entity in relation to a Container. Furthermore, it describes a dynamic process of an entity moving from the exterior of a Container via an opening to the interior of the Container. This primitive characterizes the spatial relation expressed by the prepositions "into", "in", and "inside of". Movement always requires the change of location, which links it naturally to a Source_Path_Goal image schema. In Table 4, we map the spatial primitives of this schema – *source*, *via*, and *goal*, which also correspond to semantic roles – to the elements of the respective type of movement indicated by the preposition. For the Container the type of movement relates to its spatial primitives, that is, the exterior, interior, boundary, and opening. Other image schemas, such as a more relative support schema as in "the bottle on the table", can be analyzed utilizing the relative propositions presented in Table 4.

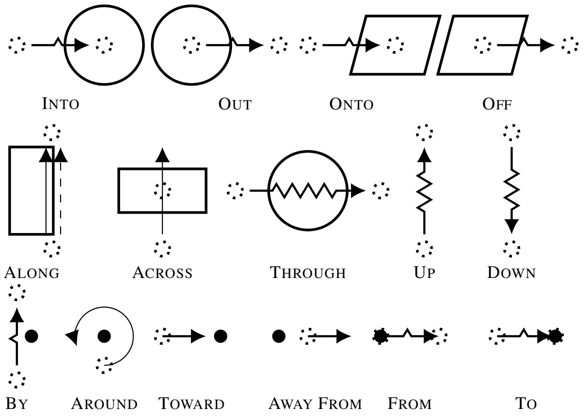


Figure 2: Diagrams for spatial primitives of spatial prepositions

Preposition	Source	Via	Goal
in, into, inside	exterior	opening	interior
out, out of	interior	opening	exterior
on, onto	-	-	surface
off, off of	surface	-	-
from	ground	-	-
to	-	-	ground
across,through	one side	points(inside)	the other side
up, upon	lower place	-	upper place
down	higher place	-	lower place
by	-	points(near)	-
Preposition	**Path**	**Via**	**Orientation**
along	line	points(inside, near)	-
around	arc	points(near)	-
toward	-	-	toward
away from	-	-	away from

Table 4: Semantic analysis of spatial prepositions

5 Procedure of ASSA

We describe the procedure of ASSA for our empirical validation of the proposed framework in Section 4. Our prototype system is written in Python and utilizes the Stanford Dependency Parser (Manning et al., 2014). Figure 1 depicts the process, where the following steps will be detailed here:

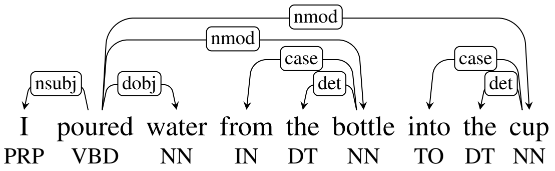
1. **Preprocessing** If the spatial expression contains a verb that encodes *path* information,

the verb is substituted by a synonymous verb+preposition. For instance, “exit” is replaced by “get out of” to enable the spatial preposition analysis. (Section 3.1).

2. **Dependency parsing** Utilizing the Stanford Dependency Parser (Chen and Manning, 2014), a parse tree is constructed.
3. **Syntactic Unit Minimization** Spliting sentence into syntactic units, or merging neighboring words into a syntactic unit.
4. **Spatial Construction Recognition** Identifying corresponding constructional schema by fuzzy pattern match, and extracting semantic roles: *agent*, *figure*, *ground*, and *relation*.
5. **Semantic Analysis of Spatial Preposition** For dynamic spatial language, indentifying spatial roles by analyzing the spatial prepositions.

As illustration, let us consider the first part of (▼): “I poured water from the bottle into the cup” (▽). After Step 1, nothing has changed for ▽, i.e., ▽' = ▽, as no verb that encodes *path* information appears in this example.

In Step 2, we generate a dependency tree and label words with their Part-of-Speech (POS) tags. The following dependency graph shows the result of Step 2:


nmod
nmod
nsubj
dobj
case
det
case
det
I poured water from the bottle into the cup
PRP VBD NN IN DT NN TO DT NN

In Step 3, we utilize dependency relations, such as `det` for determiner, to chunk words into syntactic units, e.g. “the bottle” and “the cup” in ▽'. It is worth noting that in the above dependency tree, the prepositions “from” and “to” are labeled with different POS tags, namely, “IN” and “TO” respectively. So in order to maintain consistency of the annotation of spatial prepositions, we labeled the prepositions in Table 3 with the POS tag “RP”. Additionally, each of these prepositions is treated as a single syntactic unit.

To guide the constructional schema identification, a list of verbs from VerbNet (Schuler, 2005) has been classified by two people into motion verb (MV), caused-motion verb (CMV), and stative verb (SV). For these three kinds of verbs, we use MV, CMV, or SV as their POS tags, replacing the original tags generated from Stanford dependency parser. The result of Step 3 is as follows:

I	poured	water	from	the bottle	into	the cup
PRP	CMV	NN	RP	NN	RP	NN

In Step 4, we use several fuzzy patterns to recognize the corresponding constructional schemas of spatial expressions, which are listed in Table 2. In general, each constructional schema corresponds to a fuzzy pattern[5]. For instance, the fuzzy pattern [[NN, NNS, NNP, NNPS, PRP, WP], [3], [CMV], [3], [NN, NNS, NNP, NNPS, PRP], [3], [RP], [2], [NN, NNS, NNP, NNPS, PRP]] is used to recognize the dynamic construction D-3 in Table 2,

$$\text{NP}_{\text{A}} \text{ - CMV - NP}_{\text{F}} \text{ - Spat Prep - NP}_{\text{G}}. \quad (\text{Con}(▽'))$$

The numbers in the pattern, such as “[3]”, can be replaced with less than 3 elements of any other POS tags such as “MD (modal auxiliary)”, “RB (adverb)”. It is worth mentioning that the fuzzy patterns only captures the syntactic features of the constructional schemas. Depending on the syntactic-semantic correspondence relation shown in constructional schemas, we could obtain the semantic roles *agent*, *figure*, *ground*, and *relation*.

For the example ▽', it’s syntactic feature could be represented as follows: [[PRP], [CMV], [NN], [RP], [NN], [RP], [NN]], which has been obtained in Step 3. It could be matched by the corresponding fuzzy pattern of constructional schema Con(▽'). The match algorithm could be understood as a simpler regular expression match, but it works on the word level rather than the character level. According to the syntactic-semantic correspondence relations, “I” and “water” are identified as *agent* and *figure*, respectively, and “the bottle” and “the cup” as *ground*.

In Step 5, we extract spatial roles by analyzing the semantics of spatial prepositions. Given Con(▽'), the spatial prepositions, “from” and “to” in (▽) belong respectively to the FROM and INTO schemas in Table 4. According to FROM schema, we could identify the “the bottle” as the *source* of the *figure* “water”. Similarly, we could obtain that *source*, *via*, and *goal* of the *figure* “water” are “the exterior of the cup”, “the opening of the cup”, and “the interior of the cup” respectively by analyzing INTO schema. Up to now, we have obtained all spatial

[5] A full list of constructions and the algorithm can be found here: https://github.com/chaoxu95/emnlp2020-code/blob/master/fuzzy_match.py

	Automated Spatial Semantic Analysis (ASSA)					Baseline					LIP6 System				
	Subtask 1				Subtask 2	Subtask 1				Subtask 2	Subtask 1				Subtask 2
	SP	TR	LM	Overall	Relation	SP	TR	LM	Overall	Relation	SP	TR	LM	Overall	Relation
Precision	94.43	71.01	84.52	82.77	60.26	94.76	56.72	72.97	75.36	75.18	97.59	79.29	94.05	89.55	68.33
Recall	83.15	75.40	82.90	80.06	64.43	97.74	69.56	86.21	83.81	45.47	61.13	53.43	60.73	58.03	48.03
F1	88.43	73.14	83.70	81.26	62.28	96.22	62.49	79.04	78.68	56.67	75.17	63.84	73.81	70.41	56.41
LCount	795	874	573	2242	939	-	-	-	-	-	795	874	573	2242	939
PCount	700	928	562	2190	1004	-	-	-	-	-	498	589	370	1457	660

Table 5: Overview of the performance of ASSA compared to the two state-of-the-art systems.

roles in the example (🥛). The implementation of ASSA can be found at https://github.com/chaoxu95/emnlp2020-code.

The extracted output after the application of ASSA on a spatial expression gives us insights about the expression's type of verbs, figures and grounds. We believe that these are necessary ingredients for integrating background knowledge within commonsense (spatial) reasoning systems to provide cognitively plausible explanations on the environments' change. Providing full NL explanations of our system's decisions based on presented ingredients is part of our future endeavors.

6 Evaluation

We are not aware of any specific shared task as a suitable evaluation for the novel approach proposed here. Spatial role labeling (SpRL), a subtask of SemEval (Kordjamshidi et al., 2012) seems to be the closest related task for our purpose.

We use the test data from the SpRL task of CLEF-2017 to test the coverage of our system. It has 3 subtasks: (1) label spatial indicators and their landmark(s)/trajector(s), (2) extract triplets of spatial indicator-trajector-landmark, (3) label relations with region, direction and distance. Only the first two are relevant for the evaluation of our framework[6]. Consider the following example: *About 20 kids in traditional clothing and hats waiting on stairs.* For Subtask 1 this means identifying the trajector *kids*, spatial indicator *on* and landmark as *stairs*. For Subtask 2 the system should extract *spatial_relation(kids, on, stairs)*.

6.1 Datasets and Results

The training set and test set in the SpRL dataset has 600 and 613 sentences respectively. As ASSA is based on linguistic theories and fuzzy matching, no training is required. However, as our underlying assumptions for the representation differs from the test standard in the SpRL task, the training set was utilized to adapt our prototype to this task.

The main change is representing "on the left" and "on the right". In the human-annotated corpus, these two phrases are recognized as the *spatial indicator*, whereas, in our system, "on" is recognized as the *spatial indicator*, while "the right" and "the left" are recognized as *ground*. There are 32 such phrases in the training data and 190 in the test data, which affected the final result. We added a new construction (S-7), NP_F ["on the left", "on the right"], to account for these cases.

Two systems were available for comparing the performance of ASSA, both relied on machine learning methods: (1) baseline system (Kordjamshidi et al., 2017) identified the single roles and triples by creating classifiers that used lexical, syntactical, contextual, and relational features, and (2) LIP6 system (Zablocki et al., 2017), based on a joint approach allowing for a rich feature set based on the complete relation.

Table 5 shows that our system outperforms the others on F1 score in Subtask 1 and 2, and recall in Subtask 2. SP, TR, LM refer to spatial indicator (*relation*), trajector (*figure*), landmark (*ground*) and Overall to a weighted average of these three values. LCount and PCount refer to the number of labels in the human-annotated corpus and that in the test data. Here, Overall represents their sum.

6.2 Discussion

While our approach is computationally simple yet theoretically well grounded and contributes constructional as well as preposition schemas, it still achieves a competitive performance. However, the precision of the LIP6 system is $\sim 8\%$ higher than ASSA. One reason is that the annotation of LIP6 has a higher consistency with the human-annotated

[6] The task description can be found at https://www.cs.york.ac.uk/semeval-2012/task3/index.html and CLEFLabs of the Evaluation Forum: More information can be found at http://www.cs.tulane.edu/~pkordjam/mSpRL_CLEF_lab.htm

dataset because of the training data. LIP6 uses the *indicators* constructed from training, while we take general spatial prepositions in Table 3 as indicators. An important factor that affects the precision is the treatment for the dressing descriptions. Consider the following text: "One man in a white T-shirt , grey pants and a white cap is holding a shovel"[7]. The relations (one man, in, a white t-shirt), (one man, in, grey pants), (one man, in, a white cap) are recognized by ASSA, but they do not appear in the human-annotated dataset. It seems reasonable to regard them as spatial relations. Another reason for higher precision of LIP6 is that it used the joint approach in Roberts and Harabagiu (2012), which allows for a rich feature set to reach higher precision, but it also causes lower recall. The constructions used by ASSA have been developed by us through an analysis of the Winograd Schema Challenge dataset and from the literature to recognize spatial structures. Precision and recall might be improved by adding more constructions.

ASSA has the following advantages w.r.t. the other systems: (i) for each error, a cognitively plausible explanation of the wrong result can be provided. Because of its symbolic representation, this allows us to track, understand and correct the error. (ii) ASSA can deal with dynamic expressions, and identify the roles such as *source*, *goal*. (iii) ASSA can provide a statistical distribution of types of spatial expressions, e.g., 11%, 59.3%, and 23.2% spatial expressions in the test data are respectively recognized by schemas S-1, S-2, and S-7 in Table 2.

It is worth mentioning that SpRL task is insufficient to evaluate the overall performance of ASSA, as it can not be used to evaluate how successful is ASSA at identifying the semantic roles such as *source*, *via*, and *goal*. Up to now, we have only tested it on a small number of spatial expressions extracted from WSC. The test result could be found at `https://github.com/chaoxu95/emnlp2020-code/blob/master/paper-related/wsc_output`

7 Conclusion

Although the extraction of spatial information has been widely studied, no general and human understandable approach has been proposed up to this point. Based on its interdisciplinary nature by exploiting the developments made in various areas, our contributions are two-fold: ASSA unifies methods from automated spatial information extraction with cognitive models on spatial language. The constructional schemas developed here can extract relevant spatial information, and further (in case of a dynamic construction) apply them to image schemas in order to identify corresponding semantic roles. Second, an evaluation of a prototype system outperforms machine learning systems. In terms of future work, it would be interesting to apply the prototype to a larger dataset with more dynamic spatial expressions as well as test the procedure on different languages. As discussed above, it would also be interesting to utilize the insights gained from our constructions and preposition schemas to provide natural language explanations for system decisions.

Finally, it would be interesting to investigate the interaction of ASSA with existing resources, such as FrameNet (Baker et al., 1998) to further improve on its precision and provide background knowledge to our framework. It would also be interesting to formally represent our constructional schemas, e.g. utilizing Qualitative Trajectory Calculus (Van de Weghe et al., 2005) and similar formalisms, in order to benefit from reasoning algorithms for spatial information extraction.

Acknowledgments

The first author is supported by the fundamental research funds of Shandong University.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, page 86–90, USA. Association for Computational Linguistics.

Bergen K. Bergen and Nancy Chang. 2005. Emobodied construction grammar in simulation-based language understanding. In *Construction grammars: Cognitive grounding and theoretical extensions*, pages 147–190. John Benjamins Publishing Company.

Danqi Chen and Christopher D. Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, pages 740–750. Association for Computational Linguistics.

[7] Corresponding image: images/00/116.jpg

Kenny R. Coventry and Simon C. Garrod. 2004. *Seeing, Saying and Acting: The psychological semantics of spatial prepositions*. Psychology Press.

William Croft. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

André Dittrich, Maria Vasardani, Stephan Winter, Timothy Baldwin, and Fei Liu. 2015. A classification schema for fast disambiguation of spatial prepositions. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on GeoStreaming*, pages 78–86. Association for Computing Machinery.

Ekaterina Egorova, Ludovic Moncla, Mauro Gaio, Christophe Claramunt, and Ross S Purves. 2018. Fictive motion extraction and classification. *International Journal of Geographical Information Science*, 32(11):2247–2271.

Charles J. Fillmore, Paul Kay, and Mary Catherine O'connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, pages 501–538.

Adele E. Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Maria M. Hedblom, Dagmar Gromann, and Oliver Kutz. 2018. In, out and through: formalising some dynamic aspects of the image schema containment. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 918–925. Association for Computing Machinery.

Annette Herskovits. 1987. *Language and spatial cognition: an interdisciplinary study of prepositions in English*. Cambridge University Press.

Ray Jackendoff. 1983. *Semantics and cognition*. MIT press.

Mark Johnson. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2012. Semeval-2012 task 3: Spatial role labeling. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 365–373. Association for Computational Linguistics.

Parisa Kordjamshidi, Taher Rahgooy, Marie-Francine Moens, James Pustejovsky, Umar Manzoor, and Kirk Roberts. 2017. Clef 2017: Multimodal spatial role labeling (msprl) task overview. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 367–376. Springer.

Parisa Kordjamshidi, Dan Roth, and Hao Wu. 2015. Saul: Towards declarative learning based programming. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, volume 2015, pages 1844–1851. AAAI Press.

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(3):4:1–36.

George Lakoff. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. The University of Chicago Press.

Barbara Landau and Ray Jackendoff. 1993. "what" and "where" in spatial language and spatial cognition. *Behavioral and brain sciences*, 16(2):217–238.

Jean M. Mandler and Cristóbal Pagán Cánovas. 2014. On defining image schemas. *Language and Cognition*, pages 1–23.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Fernando Melo and Bruno Martins. 2017. Automated geocoding of textual documents: A survey of current approaches. *Transactions in GIS*, 21(1):3–38.

Viswanatha Naidu, Jordan Zlatev, Vasanta Duggirala, Joost Van De Weijer, Simon Devylder, and Johan Blomberg. 2018. Holistic spatial semantics and post-talmian motion event typology: A case study of thai and telugu. *Cognitive Semiotics*, 11(2):1–27.

Georgiy Platonov and Lenhart Schubert. 2018. Computational models for spatial prepositions. In *Proceedings of the First International Workshop on Spatial Language Understanding*, pages 21–30.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 2010. *A comprehensive grammar of the English language*. Pearson Education India.

Kirk Roberts and Sanda M Harabagiu. 2012. Utd-sprl: A joint approach to spatial role labeling. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 419–424. Association for Computational Linguistics.

Karin Kipper Schuler. 2005. *Verbnet: A Broad-coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, USA. AAI3179808.

Norman K. Sondheimer. 1978. A semantic analysis of reference to spatial properties. *Linguistics and Philosophy*, 2(2):235–280.

Michael Spranger, Jakob Suchan, and Mehul Bhatt. 2016. Robust natural language processing: Combining reasoning, cognitive semantics, and construction grammar for spatial language. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 2908–2914. AAAI Press.

Leonard Talmy. 1983. How language structures space. In *Spatial orientation*, pages 225–282. Springer.

Leonard Talmy. 1985. Lexicalization patterns: Semantic structure in lexical forms. *Language typology and syntactic description*, 3(99):36–149.

Leonard Talmy. 2000a. *Toward a Cognitive Semantics: Concept Structuring Systems*, volume 1. MIT press.

Leonard Talmy. 2000b. *Toward a Cognitive Semantics: Typology and process*, volume 2. MIT press.

Leonard Talmy. 2005. The fundamental system of spatial schemas in language. In B. Hampe and J. E. Grady, editors, *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*, pages 199–234. Mouton de Gruyter.

Morgan Ulinski, Bob Coyne, and Julia Hirschberg. 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. Association for Computational Linguistics.

Nico Van de Weghe, Bart Kuijpers, Peter Bogaert, and Philippe De Maeyer. 2005. A qualitative trajectory calculus and the composition of its relations. In *International Conference on GeoSpatial Sematics*, pages 60–76. Springer.

Eloi Zablocki, Patrick Bordes, Laure Soulier, Benjamin Piwowarski, and Patrick Gallinari. 2017. Lip6@clef2017: Multi-modal spatial role labeling using word embeddings. In *Working Notes of CLEF 2017-Conference and Labs of the Evaluation Forum*.

Chunjun Zhang, Xueying Zhang, Wenming Jiang, Qijun Shen, and Shanqi Zhang. 2009. Rule-based extraction of spatial relations in natural language text. In *2009 International Conference on Computational Intelligence and Software Engineering*, pages 1–4. IEEE.