

Inferring Temporally-Anchored Spatial Knowledge from Semantic Roles

Eduardo Blanco and Alakananda Vempala

Human Intelligence and Language Technologies Lab

University of North Texas

Denton, TX, 76203

eduardo.blanco@unt.edu, AlakanandaVempala@my.unt.edu

Abstract

This paper presents a framework to infer spatial knowledge from verbal semantic role representations. First, we generate potential spatial knowledge deterministically. Second, we determine whether it can be inferred and a degree of certainty. Inferences capture that something is located or is not located somewhere, and temporally anchor this information. An annotation effort shows that inferences are ubiquitous and intuitive to humans.

1 Introduction

Extracting semantic relations from text is at the core of text understanding. Semantic relations encode semantic connections between words. For example, from (1) *Bill couldn't handle the pressure and quit yesterday*, one could extract that the CAUSE of *quit* was *the pressure*. Doing so would help answering question *Why did Bill quit?* and determining that *the pressure* started before *Bill quit*.

In the past years, computational semantics has received a significant boost. But extracting all semantic relations in text—even in single sentences—is still an elusive goal. Most existing approaches target either a single relation, e.g., PART-WHOLE (Girju et al., 2006), or relations that hold between arguments following some syntactic construction, e.g., possessives (Tratz and Hovy, 2013). Among the latter kind, the task of verbal semantic role labeling focuses on extracting semantic links exclusively between verbs and their arguments. PropBank (Palmer et al., 2005) is a popular corpus for this task, and tools to extract verbal semantic roles have been proposed for years (Carreras and Màrquez, 2005).

Some semantic relations hold forever, e.g., the CAUSE of event *quit* in example (1) above is *pressure*. Discussing when this CAUSE holds is somewhat artificial: at some point *Bill quit*, and he did so

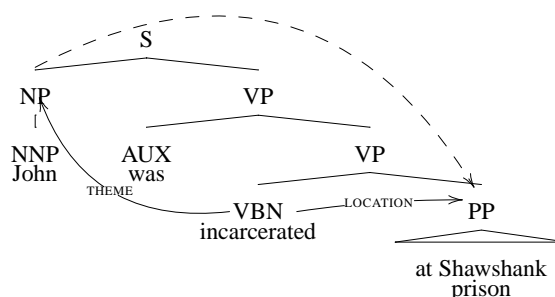


Figure 1: Semantic roles (solid arrows) and additional spatial knowledge (discontinuous arrow).

because of *the pressure*. But LOCATION and other semantic relations often do not hold forever. For example, while buildings typically have one location during their existence, people and objects such as cars and books do not: they participate in events and as a result their locations change.

This paper presents a framework to infer temporally-anchored spatial knowledge from verbal semantic roles. Specifically, our goal is to infer whether something is located somewhere or not located somewhere, and temporally anchor this spatial information. Consider sentence (2) *John was incarcerated at Shawshank prison* and its semantic roles (Figure 1, solid arrows). Given these roles, we aim at inferring that *John* had LOCATION *Shawshank prison* during event *incarcerated*, and that he (probably) did not have this LOCATION before and after (discontinuous arrow). Our intuition is that knowing that *incarcerated* has THEME *John* and LOCATION *Shawshank prison* will help making these inferences. As we shall discuss, sometimes we have evidence that something is (or is not) located somewhere, but cannot completely commit.

We target temporally-anchored spatial knowledge between intra-sentential arguments of verbs, not only between arguments of the same verb as exemplified in Figure 1. The main contributions are:

(1) analysis of spatial knowledge inferable from PropBank-style semantic roles; (2) annotations of temporally-anchored LOCATION relations on top of OntoNotes;¹ (3) supervised models to infer the additional spatial knowledge; and (4) experiments detailing results using lexical, syntactic and semantic features. The framework presented here infers over 44% spatial knowledge on top of the PropBank-style semantic roles annotated in OntoNotes (`certYES` and `certNO` labels, Section 3.3).

2 Semantic Roles and Additional Spatial Knowledge

We denote a semantic relation R between x and y as $R(x, y)$. $R(x, y)$ could be read “ x has R y ”, e.g., $\text{AGENT}(\text{moved}, \text{John})$ could be read “moved has AGENT John”. Semantic roles² are semantic relations $R(x, y)$ such that x is a verb and y is an argument of x . We refer to any spatial relation $\text{LOCATION}(x, y)$ where (1) x is not a verb, or (2) x is a verb but y is not an argument of x , as additional spatial knowledge. As we shall see, we target additional spatial knowledge beyond plain $\text{LOCATION}(x, y)$ relations, which only specify the location y of x . Namely, we consider polarity, i.e., whether something is or is not located somewhere, and temporally anchor this information.

This paper complements semantic role representations with additional spatial knowledge. We follow a practical approach by inferring spatial knowledge from PropBank-style semantic roles. We believe this is an advantage since PropBank is well-known in the field and several tools to predict PropBank roles are documented and publicly available.³ The work presented here could be incorporated into any NLP pipeline after role labeling without modifications to other components.

2.1 PropBank and OntoNotes

PropBank (Palmer et al., 2005) adds semantic role annotations on top of the parse trees of the Penn

¹Available at <http://hilt.cse.unt.edu/>

²We use *semantic role* to refer to PropBank-style (verbal) semantic roles. NomBank (Meyers et al., 2004) and FrameNet (Baker et al., 1998) also annotate semantic roles.

³E.g., <http://cogcomp.cs.illinois.edu/page/software>, <http://ml.nec-labs.com/senna/>;

[Mr. Cray] _{ARG0} [will] _{ARGM-MOD} [work] _{verb} [for the Colorado Springs CO company] _{ARG2} [as an independent contractor] _{ARG1} .
[I] _{ARG0} 'd [slept] _{verb} [through my only previous brush with natural disaster] _{ARG2} , [...]

Table 1: Examples of PropBank annotations.

ARGM-LOC: location	ARGM-CAU: cause
ARGM-EXT: extent	ARGM-TMP: time
ARGM-DIS: discourse connective	ARGM-PNC: purpose
ARGM-ADV: general-purpose	ARGM-MNR: manner
ARGM-NEG: negation marker	ARGM-DIR: direction
ARGM-MOD: modal verb	

Table 2: Argument modifiers in PropBank.

Treebank. It uses a set of numbered arguments⁴ (ARG_0 , ARG_1 , etc.) and modifiers (ARGM-TMP , ARGM-MNR , etc.). Numbered arguments do not share a common meaning across verbs, they are defined on verb-specific framesets. For example, ARG_2 is used to indicate “*employer*” with verb *work.01* and “*expected terminus of sleep*” with verb *sleep.01* (Table 1). Unlike numbered arguments, modifiers have the same meaning across verbs (Table 2).

The original PropBank corpus consists of (1) 3,327 framesets, each frameset defines the numbered roles for a verb, and (2) actual semantic role annotations (numbered arguments and modifiers) for 112,917 verbs. On average, each verb has 1.93 numbered arguments and 0.66 modifiers annotated. Only 7,198 verbs have an ARGM-LOC annotated, i.e., location information is present in 6.37% of verbs. For more information about PropBank and examples, refer to the annotation guidelines.⁵

OntoNotes (Hovy et al., 2006) is a more recent corpus that includes POS tags, word senses, parse trees, speaker information, named entities, PropBank-style semantic roles and coreference. While the original PropBank annotations were done exclusively in the news domain, OntoNotes includes other genres as well: broadcast and telephone conversations, weblogs, etc. Because of the additional annotation layers and genres, we work with OntoNotes instead of PropBank.

⁴Numbered arguments are also referred to as *core*.

⁵<http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>

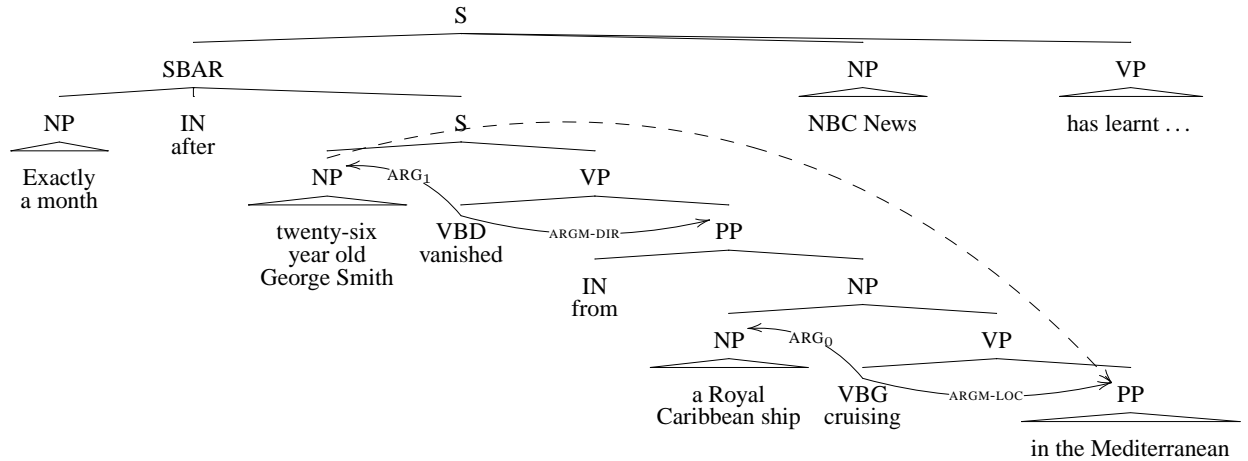


Figure 2: Semantic roles (solid arrows) and additional spatial knowledge (discontinuous arrow) of type (1b). The additional LOCATION(*a Royal Caribbean ship, in the Mediterranean*) of type (1a) is not shown.

2.2 Additional Spatial Knowledge

Sentences contain spatial information beyond ARGM-LOC semantic role, i.e., beyond links between verbs and their arguments. There are two main types of additional LOCATION(x, y) relations:⁶ (1) those whose arguments x and y are semantic roles of a verb, and (2) those whose arguments x and y are not semantic roles of a verb.

The first kind can be further divided into (1a) those whose arguments are semantic roles of the same verb (Figure 1), and (1b) those whose arguments are semantic roles of different verbs. Figure 2 illustrates type (1b). Semantic roles indicate ARG₁ and ARGM-DIR of *vanished*, and ARG₀ and ARGM-LOC of *cruising*. In this example, one can infer that *twenty-six year old George Smith* (ARG₁ of *vanished*) has LOCATION *in the Mediterranean* (ARGM-LOC of *cruising*) during the *cruising* event.

The second kind of additional LOCATION(x, y) is exemplified in the following sentence: [*Residents of Biddeford apartments*]_{ARG₀} can [*enjoy*]_{verb} [*the recreational center*]_{ARG₁} [*free of charge*]_{MANNER}. LOCATION(*recreational center, Biddeford apartments*) could be inferred yet *Biddeford apartments* is not a semantic role of a verb.⁷ Inferring this kind of relations would require splitting semantic roles;

⁶Both ARGM-LOC(x, y) and LOCATION(x, y) encode the same meaning, but we use ARGM-LOC for the PropBank semantic role and LOCATION for additional spatial knowledge.

⁷Note that the head of ARG₀ is *residents*, not the apartments.

one could also extract that the *residents* have LOCATION *Biddeford apartments*.

In this paper, we focus on extracting additional spatial knowledge of type (1), and reserve type (2) for future work. More specifically, we infer spatial knowledge between x and y , where the following semantic roles exist: ARG _{i} (x_{pred}, x) and ARGM-LOC(y_{pred}, y). ARG _{i} indicates any numbered argument (ARG₀, ARG₁, ARG₂, etc.) and x_{pred} (y_{pred}) indicates the verbal predicate to which x (y) attaches. Targeting additional spatial knowledge exclusively for numbered arguments is not a significant limitation: most semantic roles annotated in OntoNotes (75%) are numbered arguments, and it is pointless to infer spatial knowledge for most modifiers, e.g., ARGM-EXT, ARGM-DIS, ARGM-ADV, ARGM-MOD, ARGM-NEG, ARGM-DIR.

3 Annotating Spatial Knowledge

Annotating all additional spatial knowledge in OntoNotes inferable from semantic roles is a daunting task. OntoNotes is a large corpus with 63,918 sentences and 9,924 ARGM-LOC semantic roles annotated. Our goal is not to present an extensive annotation effort, but rather show that additional temporally-anchored spatial knowledge can be (1) annotated reliably by non-experts following simple guidelines, and (2) inferred automatically using supervised machine learning. Thus, we focus on 200 sentences from OntoNotes that have at least one ARGM-LOC role annotated.

```

foreach sentence  $s$  do
  foreach  $sem. role$  ARGM-LOC( $y_{pred}, y$ )  $\in s$  do
    foreach  $sem. role$  ARG $_i$ ( $x_{pred}, x$ )  $\in s$  do
      if  $is\_invalid(x, y)$  then
        Is  $x$  located at  $y$  before  $y_{pred}$ ?
        Is  $x$  located at  $y$  during  $y_{pred}$ ?
        Is  $x$  located at  $y$  after  $y_{pred}$ ?

```

Algorithm 1: Procedure to generate potential additional spatial knowledge of type (1) (Section 2.2).

Obviously, [the pilot] _{ARG0, v1} did[n't] _{ARGM-NEG, v1} [think] _{v1} [too much] _{ARGM-EXT, v1} [about [what] _{ARG1, v2} was [happening] _{v2} [on the ground] _{ARGM-LOC, v2, or ...]_{ARG1, v1}}

Figure 3: Sample sentence and semantic roles. Pair (x : *about what was happening on the ground*, y : *on the ground*) is invalid because x contains y .

All potential additional spatial knowledge is generated with Algorithm 1, and a manual annotation effort determines whether spatial knowledge should be inferred. Algorithm 1 loops over all ARGM-LOC roles, and generates questions regarding whether spatial knowledge can be inferred for any numbered argument within the same sentence. $is_valid(x, y)$ returns True if (1) x is not contained in y and (2) y is not contained in x . Considering invalid pairs would be trivial or nonsensical, e.g., pair (x : *about what was happening on the ground*, y : *on the ground*) is invalid in the sentence depicted in Figure 3.

3.1 Annotation Process and Guidelines

In a first batch of annotations, two annotators were asked questions generated by Algorithm 1 and required to answer YES or NO. The only information they had available was the source sentence without semantic role information. Feedback from this first attempt revealed that (1) because of the nature of x or y , sometimes questions are pointless, and (2) because of uncertainty, sometimes it is not correct to answer YES or NO, even though there is some evidence that makes either answer likely.

Based on this feedback, and inspired by previous annotation guidelines (Sauri and Pustejovsky, 2012), in a second batch we allowed five answers:

- **certYES**: I am certain that the answer is yes.
- **probYES**: It is probable that the answer is yes, but it is not guaranteed.
- **certNO**: I am certain that the answer is no.

- **probNO**: It is probable that the answer is no, but it is not guaranteed.
- **UNK**: There is not enough information to answer, I can't tell the location of x .

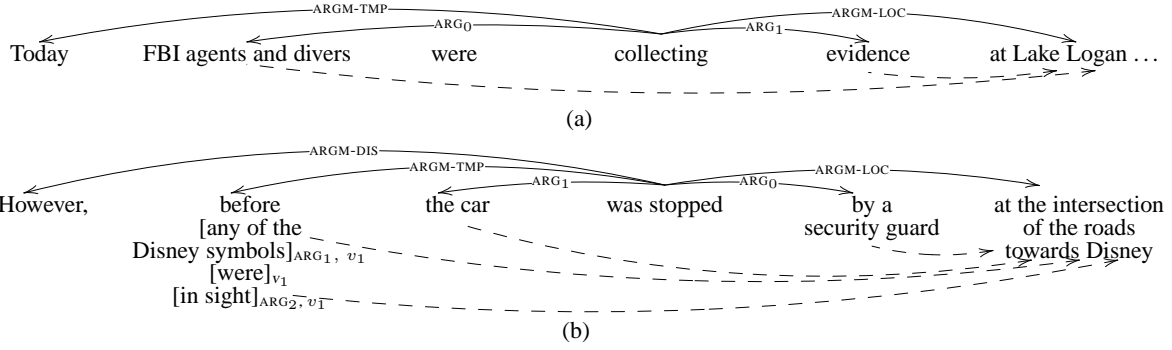
The goal is to infer spatial knowledge as gathered by humans when reading text. Thus, annotators were encouraged to use commonsense and world knowledge. While simple and somewhat open to interpretation, these guidelines allowed as to gather annotations with “good reliability” (Section 3.3.1).

3.2 Annotation Examples

In this section, we present annotation examples after resolving conflicts (Figure 4). These examples show that ambiguity is common and sentences must be fully interpreted before annotating.

Sentence 4(a) has four semantic roles for verb *collecting* (solid arrows), and annotators are asked to decide whether ARG₀ and ARG₁ of *collecting* are located at the ARGM-LOC before, during or after *collecting* (discontinuous arrows). Annotators interpreted that the *FBI agents and divers* (ARG₀) and *evidence* (ARG₁) were located *at Lake Logan* (ARGM-LOC) during *collecting* (**certYES**). They also annotated that the *FBI agents and divers* were likely to be located at *Lake Logan* before and after (**probYES**). Finally, they determined that the *evidence* was located *at Lake Logan* before the *collecting* (**certYES**), but probably not after (**probNO**). These annotations reflect the natural reading of sentence 4(a): (1) people and whatever they collect are located where the collecting takes place during the event, (2) people collecting are likely to be at that location before and after (i.e., presumably they do not arrive immediately before and leave immediately after), and (3) the objects being collected are located at that location before collecting, but probably not after.

Sentence 4(b) is more complex. First, potential relation **LOCATION**(*in sight, at the intersection*) is annotated **UNK**: it is nonsensical to ask for the location of *sight*. Second, the *Disney symbols* are never located *at the intersection* (**certNO**). Third, both *the car* and *security guard* were located *at the intersection* during the stop for sure (**certYES**). Fourth, annotators interpreted that *the car* was not *at the intersection* before (**certNO**), but they were not sure about after (**probNO**). Fifth, they considered that the *security guard* was probably located *at the intersec-*



x	y	y_{pred}	Before	During	After
FBI agents and divers	at Lake Logan	collecting	probYES	certYES	probYES
evidence	at Lake Logan	collecting	certYES	certYES	probNO
any of the Disney symbols	at the intersection of the roads ...	stopped	certNO	certNO	certNO
in sight	at the intersection of the roads ...	stopped	UNK	UNK	UNK
the car	at the intersection of the roads ...	stopped	certNO	certYES	probNO
by a security guard	at the intersection of the roads ...	stopped	probYES	certYES	probYES

Figure 4: Examples of semantic role representations (solid arrows), potential additional spatial knowledge (discontinuous arrows) and annotations with respect to the verb to which y attaches (*collecting* or *stopped*).

	Label									
	certYES		probYES		certNO		probNO		UNK	
	#	%	#	%	#	%	#	%	#	%
Before	100	15.04	225	33.83	57	8.57	248	37.29	35	5.26
During	477	71.51	36	5.40	60	9.00	59	8.85	35	5.25
After	140	21.12	344	51.89	57	8.60	87	13.12	35	5.28
All	717	35.94	605	30.33	174	8.72	394	19.75	105	5.26

Table 3: Annotation counts. Over 44% of potential spatial knowledge can be inferred (*certYES* and *certNO*).

tion before and after. In other words, annotators understood that (1) the car was moving down a road and arrived at the intersection; (2) then, it was pulled over by a security guard who is probably stationed at the intersection; and (3) after the stop, the car probably continued with its route but the guard probably stayed at the intersection.

3.3 Annotation Analysis

Each annotator answered 1,995 questions generated with Algorithm 1. Basic label counts after resolving conflicts are shown in Table 3. First, it is worth noting that annotators used *UNK* to answer only 5.26% of questions. Thus, over 94% of times *ARGM-LOC* semantic role is found, additional spatial knowledge can be inferred with some degree of certainty. Second, annotators were certain about the additional spatial knowledge, i.e., labels *certYES* and *certNO*, 35.94% and 8.72% of times respectively. Thus, 44% of times one encounters *ARGM-LOC* seman-

	Observed	Cohen Kappa
Before	89.0%	0.845
During	91.2%	0.848
After	87.8%	0.814
All	89.8%	0.862

Table 4: Inter-annotation agreements. Kappa scores indicate “good reliability”.

tic role, additional spatial knowledge can be inferred with certainty. Finally, annotators answered around 50% of questions with *probYES* or *probNO*. In other words, they found it likely that spatial information can be inferred, but were not completely certain.

3.3.1 Inter-Annotator Agreements

Table 4 presents observed agreements, i.e., raw percentage of equal annotations, and Cohen Kappa scores (Cohen, 1960) per temporal anchor and for all questions. Kappa scores are above 0.80, indicating “good reliability” (Artstein and Poesio, 2008).

	No.	Name	Description
	0	temporal anchor	are we predicting LOCATION(x, y) before, during or after y_{pred} ?
lexical	1–4	first word, POS tag	first word and POS tag in x and y
	5–8	last word, POS tag	last word and POS tag in x and y
	9,10	num tokens	number of tokens in x and y
	11,12	subcategory	concatenation of (1) x 's children and (2) y 's children
	13	direction	whether x occurs before or after y
syntactic	14,15	syntactic node	syntactic node of x and y
	16–19	head word, POS tag	head word and POS tag of x and y
	20–23	left and right sibling	syntactic nodes of the left and right siblings of x and y
	24–27	parent node and index	syntactic nodes and child indices of parents of x and y
	28	common subsumer	syntactic node subsuming x and y
	29	syntactic path	syntactic path between x and y
semantic	30–33	word, POS tag	predicate and POS tag of x_{pred} and y_{pred}
	34	isRole	semantic role label between x_{pred} and x
	35	same predicate	whether x_{pred} and y_{pred} are the same token
	36–39	firstRole, lastRole	the first and last semantic roles of x_{pred} and y_{pred}
	40–59	hasRole	flags indicating whether x_{pred} and y_{pred} have each semantic role
	60–99	role index and node	for each semantic role, the order of appearance and syntactic node
	100	$x_containedIn_y_role$	semantic role of y_{pred} that fully contains x
	101	$y_containedIn_x_role$	semantic role of x_{pred} that fully contains y

Table 5: Feature set to infer temporally-anchored spatial knowledge from semantic role representations.

We believe the high Kappa scores are due to the fact that we start from PropBank-style roles instead of plain text, and questions asked are intuitive. Note that not all disagreements are equal, e.g., the difference between `certYES` and `certNO` is much larger than the difference between `certYES` and `probYES`.

4 Inferring Spatial Knowledge

We follow a standard supervised machine learning approach. The 200 sentences were divided into train (80%) and test (20%), and the corresponding instances assigned to the train and test sets.⁸ We trained an SVM with RBF kernel using scikit-learn (Pedregosa et al., 2011). Parameters C and γ were tuned using 10-fold cross-validation with the training set, and results are calculated with test instances.

4.1 Feature selection

Selected features (Table 5) are a mix of lexical, syntactic and semantic features, and are extracted from tokens (words and POS tags), full parse trees and semantic roles. Lexical and syntactic features are standard in semantic role labeling (Gildea and Jurafsky, 2002) and we do not elaborate on them. Hereafter

⁸Splitting instances randomly would be unfair, as instances from the same sentence would be assigned to the train and test sets. Thank you to an anonymous reviewer for pointing this out.

Sentence: [In this laboratory] _{ARGM-LOC, v1} [I] _{ARG0, v1} 'm [surrounded] _{v1} [by the remains of [20 service members who] _{ARG1, v2} are in the process of being [identified] _{v2}] _{ARG1, v1}
Potential additional spatial knowledge: x : <i>20 service members who</i> , y : <i>In this laboratory</i> ; $x_containedIn_y_role = ARG1$
Sentence: [Children] _{ARG0, v1} can get to [know] _{v1} [different animals and plants, and [even some crops that] _{ARG1, v2} are [rarely] _{ARGM-ADV, v2} [seen] _{v2} [in our daily life] _{ARGM-LOC, v2}] _{ARG1, v1} .
Potential additional spatial knowledge: x : <i>Children</i> , y : <i>in our daily life</i> ; $y_containedIn_x_role = ARG1$

Figure 5: Pairs (x, y) for which $x_containedIn_y_role$ and $y_containedIn_x_role$ features have a value.

we describe semantic features, which include any feature derived from semantic role representations.

Features 30–33 correspond to the surface form and POS tag of the verbs to which x and y attach to. Feature 34 indicates the semantic role between x_{pred} and x ; note that the semantic role between y_{pred} and y is always ARGM-LOC (Algorithm 1). Feature 35 distinguishes inferences of type (1a) from (1b) (Section 2.2): it indicates whether both x and y attach to the same verb, as in Figure 1, or not, as in Figure 2. Features 36–39 encode the first and last semantic role of x_{pred} and y_{pred} by order of appearance. Features 40–59 are binary flags signalling which se-

		Before			During			After			All		
		P	R	F	P	R	F	P	R	F	P	R	F
most frequent baseline	certYES	0.11	1.00	0.20	0.74	1.00	0.85	0.26	1.00	0.42	0.37	1.00	0.54
	other labels	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	weighted avg.	0.01	0.11	0.02	0.54	0.74	0.63	0.07	0.26	0.11	0.14	0.37	0.20
most frequent per temporal anchor baseline	certYES	0.00	0.00	0.00	0.75	1.00	0.86	0.00	0.00	0.00	0.75	0.62	0.68
	probYES	0.00	0.00	0.00	0.00	0.00	0.00	0.45	1.00	0.62	0.45	0.56	0.50
	probNO	0.38	1.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.62	0.47
	other labels	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	weighted avg.	0.14	0.38	0.21	0.57	0.75	0.65	0.20	0.45	0.28	0.50	0.53	0.50
lexical features	certYES	0.13	0.20	0.16	0.74	1.00	0.85	0.53	0.29	0.37	0.63	0.75	0.69
	probYES	0.39	0.34	0.36	0.00	0.00	0.00	0.56	0.90	0.69	0.51	0.63	0.56
	certNO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	probNO	0.39	0.53	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.37	0.38
	UNK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	weighted avg.	0.31	0.35	0.32	0.54	0.74	0.63	0.44	0.56	0.47	0.47	0.55	0.50
lexical + syntactic features	certYES	0.41	0.47	0.44	0.74	0.99	0.85	0.27	0.09	0.13	0.67	0.72	0.70
	probYES	0.53	0.34	0.41	0.00	0.00	0.00	0.54	0.90	0.67	0.54	0.63	0.58
	certNO	0.33	0.10	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.04	0.06
	probNO	0.38	0.64	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.44	0.41
	UNK	1.00	0.12	0.22	1.00	0.12	0.22	1.00	0.12	0.22	1.00	0.12	0.22
	weighted avg.	0.48	0.43	0.41	0.61	0.74	0.64	0.42	0.51	0.41	0.57	0.56	0.53
lexical + semantic features	certYES	0.18	0.20	0.19	0.74	1.00	0.85	0.65	0.31	0.42	0.67	0.76	0.71
	probYES	0.48	0.42	0.44	0.00	0.00	0.00	0.57	0.92	0.70	0.54	0.66	0.60
	certNO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	probNO	0.35	0.51	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.35	0.35
	UNK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	weighted avg.	0.33	0.37	0.34	0.54	0.74	0.63	0.47	0.57	0.49	0.49	0.56	0.52
all features	certYES	0.50	0.20	0.29	0.76	0.97	0.85	0.50	0.14	0.22	0.73	0.70	0.71
	probYES	0.51	0.36	0.42	0.50	0.14	0.22	0.56	0.93	0.70	0.55	0.66	0.60
	certNO	0.33	0.10	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.04	0.05
	probNO	0.40	0.72	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.50	0.44
	UNK	1.00	0.12	0.22	0.33	0.12	0.18	0.50	0.12	0.20	0.50	0.12	0.20
	weighted avg.	0.49	0.44	0.41	0.61	0.73	0.65	0.46	0.54	0.45	0.56	0.57	0.55

Table 6: Results obtained with two baselines, and training with several feature combinations. Models are trained with all instances (before, during and after).

semantic roles x_{pred} and y_{pred} have, and features 60–99 capture the index of each role (first, second, third, etc.) and its syntactic node (NP, PP, SBAR, etc.).

Finally, features 100 and 101 capture the semantic role of x_{pred} and y_{pred} which fully contain y and x respectively, if such roles exist. These features are especially designed for our inference task and are exemplified in Figure 5.

5 Experiments and Results

Results obtained with the test set using two baselines and models trained with several feature combinations are presented in Table 6. The *most frequent* baseline always predicts certYES, and the *most frequent per temporal anchor* baseline pre-

dicts probNO, certYES and probYES for instances with temporal anchor *before*, *during* and *after* respectively. The *most frequent* baseline obtains a weighted F-measure of 0.20, and *most frequent per temporal anchor* baseline 0.50. Results with supervised models are better, but we note that always predicting certYES for *during* instances obtains the same F-measure than using all features (0.65).

The bottom block of Table 6 presents results using all features. The weighted F-measure is 0.55, and the highest F-measures are obtained with labels certYES (0.71) and probYES (0.60). Results with certNO and probNO are lower (0.05 and 0.44), we believe this is due to the fact that few instances are annotated with these labels (8.72% and 19.75%, Ta-

ble 3). Results are higher (0.65) with *during* instances than with *before* and *after* instances (0.41 and 0.45). These results are intuitive: certain events such as *press* and *write* require participants to be located where the event occurs only during the event.

5.1 Feature Ablation and Detailed Results

The weighted F-measure using lexical features is the same than with the *most frequent per temporal anchor* baseline (0.50). F-measures go up with *before* (0.21 vs. 0.32, 52.38%) and *after* (0.28 vs. 0.47, 67.85%) instances, but slightly down with *during* instances (0.65 vs. 0.63, -3.08%).

Complementing lexical features with syntactic and semantic features brings the overall weighted F-measure slightly up: 0.53 with syntactic and 0.52 with semantic features (+0.03 and +0.02, 6% and 4%). *Before* instances benefit the most from syntactic features (0.32 vs. 0.41, 28.13%), and *after* instances benefit from semantic features (0.47 vs. 0.49, 4.26%). *During* instances do not benefit from semantic features, and only gain 0.01 F-measure (1.59%) with syntactic features.

Finally, combining lexical, syntactic and semantic features obtains the best overall results (weighted F-measure: 0.55 vs. 0.53 and 0.52, 3.77% and 5.77%). We note, however, that *before* instances do not benefit from including semantic features (same F-measure, 0.41), and the best results for *after* instances are obtained with lexical and semantic features (0.49 vs. 0.45, 8.16%),

6 Related Work

Tools to extract the PropBank semantic roles we infer from have been studied for years (Carreras and Màrquez, 2005; Hajič et al., 2009; Lang and Lapata, 2010). These systems only extract semantic links between predicates and their arguments, not between arguments of predicates. In contrast, this paper complements semantic role representations with spatial knowledge for numbered arguments.

There have been several proposals to extract semantic links not annotated in well-known corpora such as PropBank (Palmer et al., 2005), FrameNet (Baker et al., 1998) or NomBank (Meyers et al., 2004). Gerber and Chai (2010) augment NomBank annotations with additional numbered argu-

ments appearing in the same or previous sentences; posterior work obtained better results for the same task (Gerber and Chai, 2012; Laparra and Rigau, 2013). The SemEval-2010 Task 10: Linking Events and their Participants in Discourse (Ruppenhofer et al., 2009) targeted cross-sentence missing numbered arguments in PropBank and FrameNet. We have previously proposed an unsupervised framework to compose semantic relations out of previously extracted relations (Blanco and Moldovan, 2011; Blanco and Moldovan, 2014a), and a supervised approach to infer additional argument modifiers (ARGM) for verbs in PropBank (Blanco and Moldovan, 2014b). Unlike the current work, these previous efforts (1) improve the semantic representation of verbal and nominal predicates, or (2) infer relations between arguments of the same predicate. None of them target temporally-anchored spatial knowledge or account for uncertainty.

Attaching temporal information to semantic relations is uncommon. In the context of the TAC KBP temporal slot filling track (Garrido et al., 2012; Surdeanu, 2013), relations common in information extraction (e.g., SPOUSE, COUNTRY_OF_RESIDENCY) are assigned a temporal interval indicating when they hold. The task proved very difficult, and the best system achieved 48% of human performance. Unlike this line of work, the approach presented in this paper starts from semantic role representations, targets temporally-anchored LOCATION relations, and accounts for degrees of uncertainty (*certYES / certNO* vs. *probYES / probNO*).

The task of spatial role labeling (Hajič et al., 2009; Kolomiyets et al., 2013) aims at thoroughly representing spatial information with so-called spatial roles, i.e., trajectory, landmark, spatial and motion indicators, path, direction, distance, and spatial relations. Unlike us, the task does not consider temporal spans nor certainty. But as the examples throughout this paper show, doing so is useful because (1) spatial information for most objects changes over time, and (2) humans sometimes can only state that an object is *probably* located somewhere. In contrast to this task, we infer temporally-anchored spatial knowledge as humans intuitively understand it, and purposely avoid following any formalism.

7 Conclusions

Semantic roles encode semantic links between a verb and its arguments. Among other role labels, PropBank uses numbered arguments (ARG_0 , ARG_1 , etc.) to encode the core arguments of a verb, and ARGM-LOC to encode the location. This paper exploits these numbered arguments and ARGM-LOC in order to infer temporally-anchored spatial knowledge. This knowledge encodes whether a numbered argument x is or is not located in a location y , and temporally anchors this information with respect to the verb to which y attaches.

An annotation effort with 200 sentences from OntoNotes has been presented. First, potential additional spatial knowledge is generated automatically (Algorithm 1). Then, annotators following straightforward guidelines answer questions asking for intuitive spatial information, including uncertainty. The result is annotations with high inter-annotator agreements that encode spatial knowledge as understood by humans when reading text.

Experimental results show that inferring additional spatial knowledge can be done with a modest weighted F-measure of 0.55. Results are higher for `certYES` and `probYES` (0.71 and 0.60), the labels that indicate that something is certainly or probably located somewhere. Simple majority baselines provide strong results, but combining lexical, syntactic and semantic features yields the best results (0.50 vs. 0.55). Inferring spatial knowledge for numeric arguments before and after an event occurs is harder than during the event (0.41 and 0.45 vs. 0.65).

The most important conclusion of this work is the fact that given an ARGM-LOC semantic role, temporally-anchored spatial knowledge can be inferred for numbered arguments in the same sentence. Indeed, annotators answered 44% of questions with `certYES` or `certNO`, and 50% of questions with `probYES` or `probNO`. Another important observation is that spatial knowledge can be inferred from most verbs, not only motion verbs. While it is fairly obvious to infer from *John went to Paris* that he had LOCATION *Paris* after *went* but not before or during, we have shown that verbs such as *incarcerated* (Figure 1) also grant spatial inferences.

References

- Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, December.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 17th international conference on Computational Linguistics*, Montreal, Canada.
- Eduardo Blanco and Dan Moldovan. 2011. Unsupervised learning of semantic relation composition. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, pages 1456–1465, Portland, Oregon.
- Eduardo Blanco and Dan Moldovan. 2014a. Composition of semantic relations: Theoretical framework and case study. *ACM Trans. Speech Lang. Process.*, 10(4):17:1–17:36, January.
- Eduardo Blanco and Dan Moldovan. 2014b. Leveraging verb-argument structures to infer semantic relations. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, pages 145–154, Gothenburg, Sweden.
- Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: semantic role labeling. In *CONLL '05: Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 152–164.
- J. Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37.
- Guillermo Garrido, Anselmo Peñas, Bernardo Cabaleiro, and Álvaro Rodrigo. 2012. Temporally anchored relation extraction. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1, ACL '12*, pages 107–116.
- Matthew Gerber and Joyce Chai. 2010. Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1583–1592, Uppsala, Sweden, July.
- Matthew Gerber and Joyce Chai. 2012. Semantic role labeling of implicit arguments for nominal predicates. *Computational Linguistics*, 38:755–798, 2012.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, September.
- Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135, March.
- Jan Hajič, Massimiliano Ciaramita, Richard Johanson, Daisuke Kawahara, Maria Antònia Martí, Lluís

- Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL '09*, pages 1–18.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% Solution. In *NAACL'06: Proceedings of the Human Language Technology Conference of the NAACL*, pages 57–60, Morristown, NJ, USA.
- Oleksandr Kolomiyets, Parisa Kordjamshidi, Marie-Francine Moens, and Steven Bethard. 2013. Semeval-2013 task 3: Spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 255–262.
- Joel Lang and Mirella Lapata. 2010. Unsupervised induction of semantic roles. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*, pages 939–947.
- Egoitz Laparra and German Rigau. 2013. Impar: A deterministic algorithm for implicit semantic role labelling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1180–1189, Sofia, Bulgaria, August.
- A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The Nom-Bank Project: An Interim Report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Josef Ruppenhofer, Caroline Sporleder, Roser Morante, Collin Baker, and Martha Palmer. 2009. SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, pages 106–111, Boulder, Colorado, June.
- Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299, June.
- Mihai Surdeanu. 2013. Overview of the tac2013 knowledge base population evaluation: English slot filling and temporal slot filling. In *Proceedings of the TAC-KBP 2013 Workshop*.
- Stephen Tratz and Eduard Hovy. 2013. Automatic interpretation of the english possessive. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 372–381. Association for Computational Linguistics.