

High-Fidelity Lexical Axiom Construction from Verb Glosses

Gene Kim and Lenhart Schubert

Presented by: Gene Kim
August 2016



UNIVERSITY of
ROCHESTER

Understanding Language

- All language is composed of words. Understanding and inference in language requires knowledge about the words themselves.
- We build a lexical KB with inference-enabling axioms that correspond to verb entries in WordNet.

Understanding Language

- All language is composed of words. Understanding and inference in language requires knowledge about the words themselves.
- We build a lexical KB with inference-enabling axioms that correspond to verb entries in WordNet.

slam2.v

Gloss: “strike violently”

Frames: [Somebody *slam2.v* Something]

Examples: “slam the ball”



Axiom: $(\forall x,y,e: [[x \text{ slam2.v } y] ** e]$
 $\rightarrow [[[x \text{ (violently1.adv (strike1.v } y))] ** e]$
 $\text{and } [x \text{ person1.n } [y \text{ thing12.n}]]])$

Outline

- Previous Work
- High-Fidelity Lexical Axiom Construction from Verb Glosses
- Evaluation
 - EL-smatch
- Conclusions and Future Work

Why another machine-comprehensible dictionary?

- This has been done before!
 - (Hobbs, 2008)¹
 - (Allen et al. 2013)²
 - etc.

Why another machine-comprehensible dictionary?

- This has been done before!
 - (Hobbs, 2008)¹
 - (Allen et al. 2013)²
 - etc.

Semantic Representation!

Semantic Representation

- Natural language is very expressive
 - Predicates, connectives, quantifiers, equality → FOL
 - Generalized quantifiers (*e.g. most men who smoke*)
 - Intensional predicates (*e.g. believe, intend, resemble*)
 - Predicate and sentence modification (*e.g. very, gracefully, nearly, possibly*)
 - Predicate and sentence reification
(*e.g. Beauty is subjective, That exoplanets exist is now certain*)
 - Reference to events and situations
(*Many children had not been vaccinated against measles;
this situation caused sporadic outbreaks of the disease*)
- Semantic representation should be able to represent these devices!
- Semantic representation needs a formal interpretation for justified inference.

Semantic Representation

- (Hobbs, 2008)¹ - Hobbsian Logical Form (HLF)
 - Issues in the interpretation of quantifiers and conflation of events and propositions

- (Allen et al. 2013)² - Description Logic (OWL-DL)
 - Handling of predicate/sentence reification, predicate modification, self-reference, and uncertainty is unsatisfactory

Semantic Representation

Episodic Logic

- Extended FOL -- handles most natural language phenomena
- Backed by fast and comprehensive theorem prover EPILOG

Example:

“Kim believes that every galaxy harbors life”

→ [Kim.name believe.v (That (\forall x: [x galaxy.n] [x harbor.v (K life.n)])))]

Episodic Logic Basics

Notable syntax

- slam2.v → sense 2 of the verb slam.
 - (v: verb, n: noun, a: adjective, adv: adverb, p: preposition, cc: connective)
- Infix formulas in *square brackets* []
 - Predicate application - [John.name love.v Mary.name]
 - Connectives - [TRUE and.cc FALSE], [TRUE or.cc FALSE], [$\Phi \rightarrow \Psi$]
 - Episodic operators - [$\Phi ** e$], [$\Phi * e$]
- Prefixed formulas in *parentheses* ()
 - Negation - ($\neg \Phi$)
 - Modification - (loudly.adv whisper.v), (past [Alice.name message.v Bob.name])
 - Reification - (K dog.n), (That [John.name love.v Mary.name])

Episodic Logic Basics

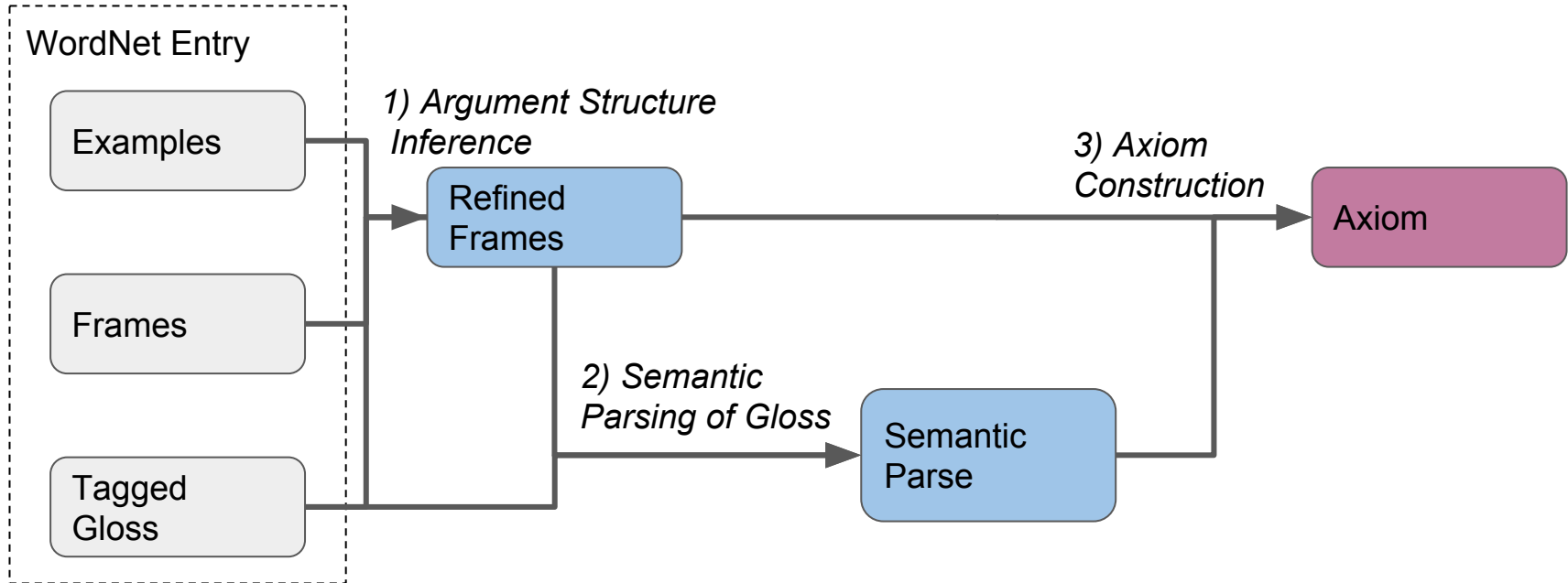
Relevant operators for this presentation

- Episodic Operators
 - $[\Phi^{**} e]$ - Formula Φ characterizes episode e .
 - $[\Phi^* e]$ - Formula Φ is true in episode e .
- Reification
 - $(\mathbf{K} \text{ man.n})$ - Predicate man.n as a kind (i.e. mankind)
 - $(\mathbf{That} [\text{John.name man.n}])$ - Sentence $[\text{John.name man.n}]$ as an object (i.e. “That John is a man”)

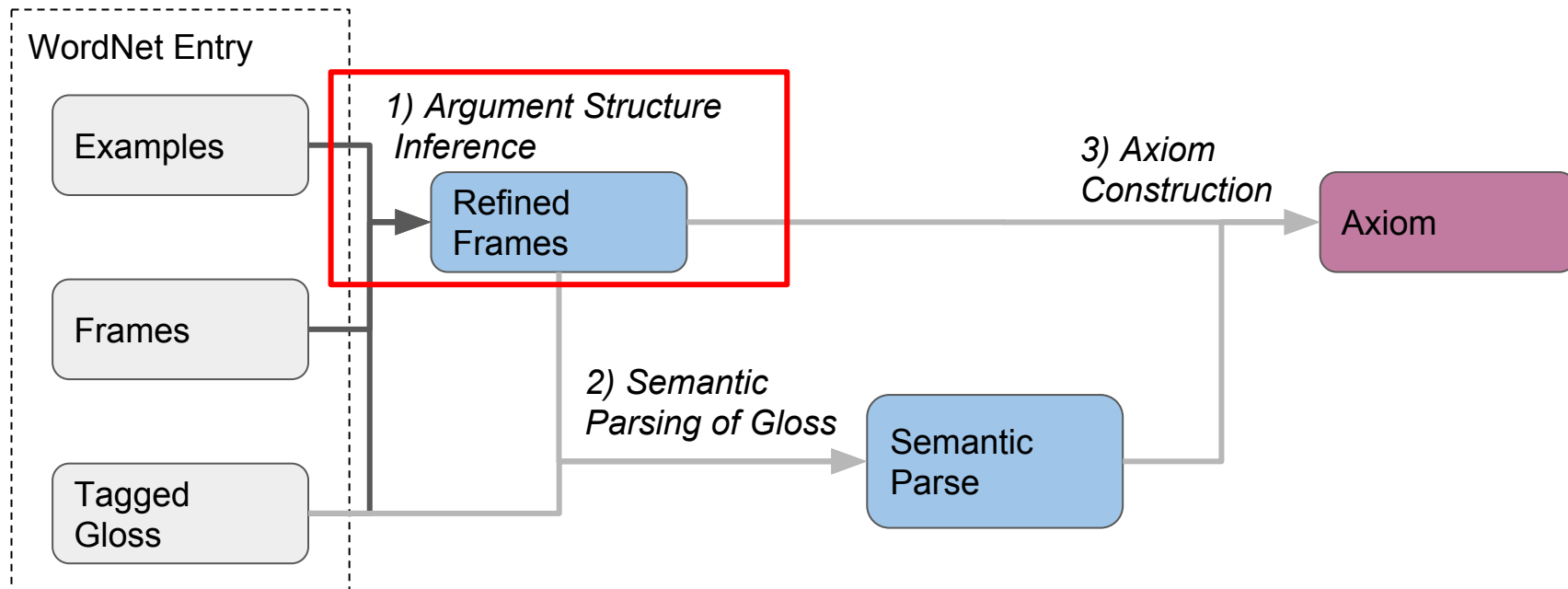
Outline

- Previous Work
- **High-Fidelity Lexical Axiom Construction from Verb Glosses**
- Evaluation
 - EL-smatch
- Conclusions and Future Work

Axiomatization - overview



Axiomatization - overview



Argument Structure Inference

1. Start with WN sentence frames

<i>quarrel1.v</i>	[Somebody quarrel1.v] [Somebody quarrel1.v PP]
<i>paint2.v</i>	[Somebody paint2.v Something]
<i>mail1.v</i>	[Somebody mail1.v Somebody Something] [Somebody mail1.v Something] [Somebody mail1.v Something to Somebody]
<i>percolate1.v</i>	[Something percolate1.v]

Argument Structure Inference

2. Refine/extend using examples and gloss(es) in synset

Refine using examples

quarrel2.v

“We quarreled over the question as to who discovered America”

“These two fellows are always scrapping over something”

→ [(plural Somebody) quarrel1.v]
[Somebody quarrel1.v PP-OVER]

Refine using gloss

paint2.v - make a painting → [Somebody paint1.v painting.n]

Argument Structure Inference

3. Remove/merge redundant frames and add dative alternations

Merge

[Somebody -s] + [Something -s]

→ [Something -s]

[Somebody -s Adjective/Noun] + [Somebody -s PP]

→ [Somebody -s Adjective/Noun/PP]

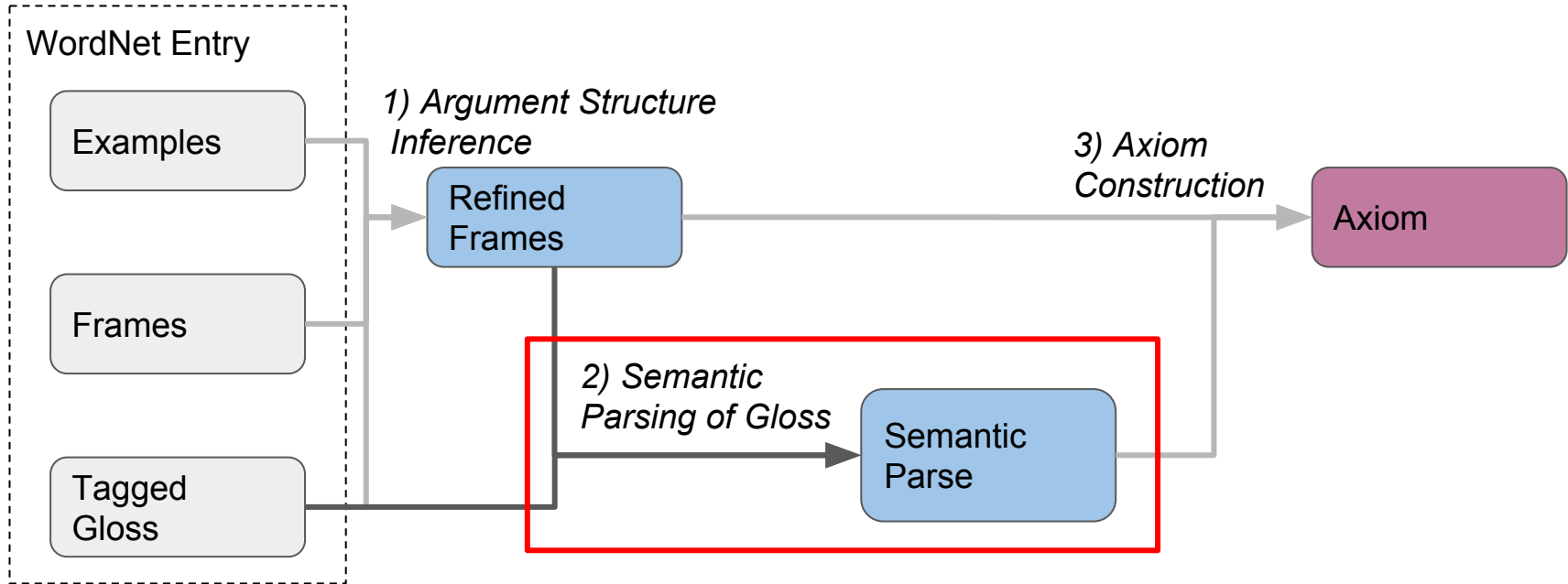
Add dative alternation

[Somebody -s Somebody Something]

→ [Somebody -s Somebody Something] +

[Somebody -s Something to Somebody]

Axiomatization - overview



Semantic Parsing of Gloss

High quality semantic parsing using preprocessing simplifications

- 1. Preprocess
 - Canonicalize arguments
 - Factor coordinated groups
- 2. Use semantic parser modeled after KNEXT system (Van Durme et al. 2009; Gordon and Schubert, 2010)
- 3. Word Sense Disambiguation (WSD)

Semantic Parsing of Gloss - arguments

- Extract argument types from gloss
- Canonicalize arguments
 - Replace existing arguments with canonical arguments
 - Insert canonical arguments if arguments are missing

Canonical Arguments

<i>Argument position</i>	<i>English text</i>	<i>EL atom</i>
subject	I/myself	Me.pro
direct object	it	It.pro
indirect object	them	They.pro

Examples of argument identification & canonicalization

slam2.v - _ strike _ violently
paint1.v - _ make a painting



slam2.v - *I'll* strike *it* violently
paint1.v - *I'll* make *it*; (*it* : a painting)

Semantic Parsing of Gloss - coordinators

- Syntactic and semantic parsers are easily thrown off by coordinated phrases
- Coordinated groups (CGs) are identified by syntactic and semantic relatedness
 - Use linguistic phrase types (NP, VP, PP, etc.) as a proxy for relatedness
 - Identified by simple POS pattern-matching
- Replace CG with first phrase in the group, save group
- Run groups through modified semantic parser for CGs and reintroduce to semantic parse of the simplified gloss

Example Extraction

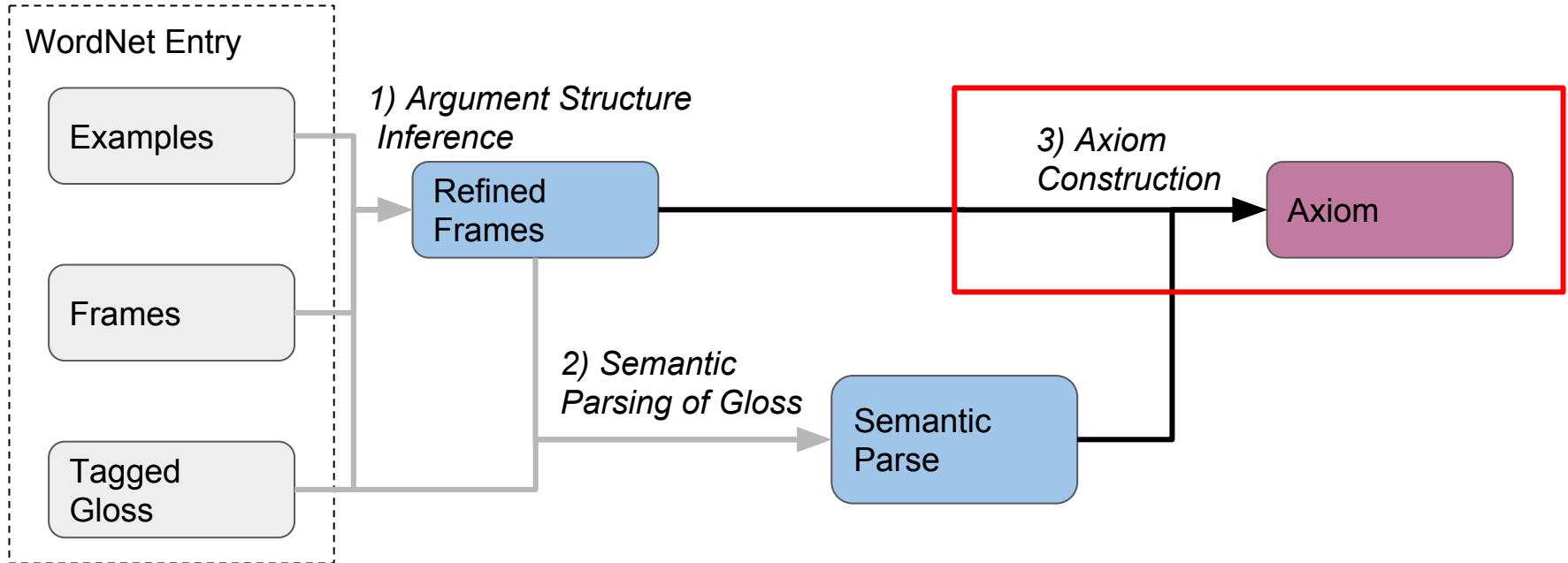
rejuvenate3.v.: (PRP I) (VB make) (PRP it) (**JJR younger**) (**CC or**) (**RBR more**) (**JJ youthful**)

→ (PRP I) (VB make) (PRP it) (**JJR younger**);
(**JJR younger**) (**CC or**) (**RBR more**) (**JJ youthful**)

Semantic Parsing of Gloss - WSD

- Princeton Annotated Gloss Corpus provides WSD for a portion of words in WordNet glosses
- Else, use POS pattern-matching to identify the context frame and select the lowest numbered sense with a matching frame.

Axiomatization - overview



Axiom Construction

- Construct an axiom asserting that an event e characterized by the predication of the entry word entails that e is also characterized by the semantic parse of the gloss with appropriate semantic types of the arguments
 - Correlate arguments between the frame and the semantic parse of the gloss
 - Replace arguments with variables
 - Constrain variable types based on frame and extracted argument types
 - Wrap entailment from frame to gloss in universal quantifiers of the variables

Argument Correlation

<i>Argument position</i>	<i>EL atom</i>
subject	Me.pro
direct object	It.pro
indirect object	They.pro

[Somebody slam2.v Something]

subject

direct object

[Me.pro (violently1.adv (strike1.v It.pro))]

Entailment Wrapping

[x slam2.v y],
[x (violently1.adv (strike1.v y))],
[x person1.n], [y thing12.n]

$(\forall x,y,e: [[x \text{ slam2.v } y] ** e]$
 $[[[x \text{ (violently1.adv (strike1.v y))}] ** e]$
 $\text{and } [x \text{ person1.n}] [y \text{ thing12.n}]])$

Outline

- Previous Work
- High-Fidelity Lexical Axiom Construction from Verb Glosses
- **Evaluation**
 - EL-smatch
- Conclusions and Future Work

Evaluation

- Gold standard set of axioms from 50 synsets
 - Hand-written
 - *EL-match* metric: allows partial credit
 - Full axiom metric

- Lexical entailment relations between verbs
 - Verb entailment dataset (Weisman et al., 2012)
 - Demonstrates inference capabilities of axioms
 - Allows comparison to previous systems

EL-smatch

- Generalization of *smatch* (Cai and Knight, 2013)
 - Maximum triple match for any variable mapping between two formulas
 - *smatch* does not allow instances that are not atoms! (e.g. (very.adv happy.a))
- Original representation (3 types)
 - *instance(variable, type)*
 - *relation(variable, variable)*
 - *attribute(variable, value)*
- Add complex instance triple
 - *instance(variable, variable)*

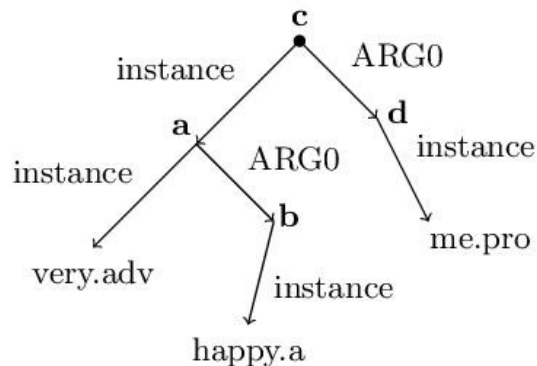
EL

(me.pro (very.adv happy.a))

EL-smatch Triple Representation

instance(a, very.adv) \wedge
instance(b, happy.a) \wedge
instance(d, me.pro) \wedge
ARG0(a, b) \wedge
instance(c, a) \wedge
ARG0(c, d)

EL-smatch Graph Representation



EL-smatch

- Generalization of *smatch* (Cai and Knight, 2013)
 - *smatch* does not allow instances that are not atoms! (e.g. (very.adv happy.a))
- Original representation (3 types)
 - *instance(variable, type)*
 - *relation(variable, variable)*
 - *attribute(variable, value)*
- Add complex instance triple
 - *instance(variable, variable)*

EL

(me.pro (very.adv happy.a))

EL-smatch Triple Representation

instance(a, very.adv) \wedge

instance(b, happy.a) \wedge

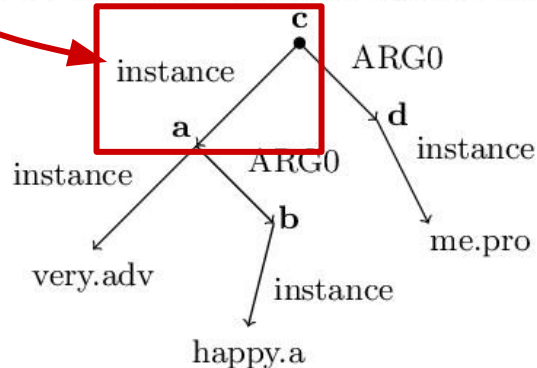
instance(d, me.pro) \wedge

ARG0(a, b) \wedge

instance(c, a) \wedge

ARG0(c, d)

EL-smatch Graph Representation



Gold Standard Set

- Manually constructed axioms for the glosses of 50 synsets
- 52 axioms
- 2,764 triples
- Evaluated using two metrics
 - *EL-smatch*
 - Full axiom matching

Results

Measure	Precision	Recall	F1
<i>EL-smatch</i>	0.85	0.82	0.83
Full Axiom	0.29	-	-

Verb Entailment

- Verb inference dataset (Weisman et al., 2012)
 - Created by randomly sampling 50 common verbs in the Reuters corpus, and is then randomly paired with 20 most similar verbs according to the Lin similarity measure (Lin, 1998)
 - 812 verb pairs - manually annotated as representing a valid entailment rule or not
 - 225 verb pairs are labeled as entailing and 587 verb pairs were labeled as non-entailing
- Comparison to previous work and demonstration of generality
- Requires simplifying our dataset
 - Remove semantic roles and word senses at start and end.
- Simple forward inference up to 3 times or until abstract word is reached

Results

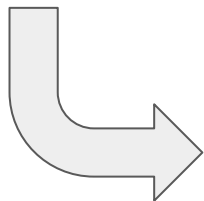
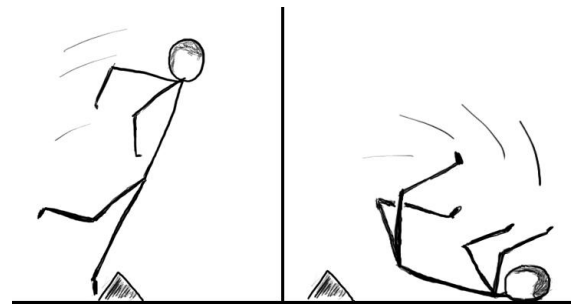
Method	Precision	Recall	F1
Our Approach	0.43	0.53	0.48
<i>TRIPS</i>	0.50	0.45	0.47
<i>Supervised</i>	0.40	0.71	0.51
<i>VerbOcean</i>	0.33	0.15	0.20
<i>Random</i>	0.28	0.29	0.28

Example Inference

Sentence “John stumbles, but does not fall”

WN entry *stumble2.v* : miss a step and fall or nearly fall

Simple Linguistic Rule (x but y) \rightarrow (x and y)



Conclusion “John misses a step and nearly falls”

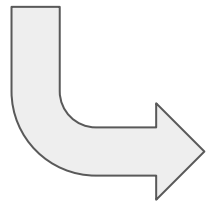
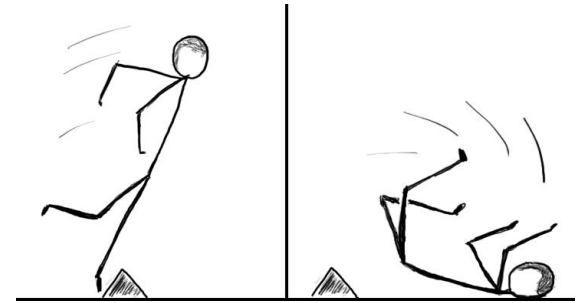
Not John

Example Inference

Sentence “John stumbles, but does not fall”

WN entry *stumble2.v* : miss a step and fall or nearly fall

Simple Linguistic Rule (x but y) \rightarrow (x and y)



Conclusion “John misses a step and nearly falls”

Not John

Intersective approaches lead to contradiction!

Conclusions

- We argued that the semantic representations used in previous approaches to extracting lexical axioms from dictionaries were insufficient.
- We presented an approach to extracting lexical axioms expressed in EL from verb definitions in WordNet
- We introduced *EL-smatch*, a generalization of *smatch* with complex instances
- Evaluated our approach using a gold standard and against an entailment task, where it is competitive to the state-of-the-art

Future Work

- Deepen gloss interpretation
 - Using the hypernym graph in WordNet
 - Use techniques from (Allen et al., 2013) → High-level ontology, generating temporary axioms, etc
 - Use techniques from (Mostafazadeh and Allen, 2015) → clustering to refine arguments
- Extend work to nouns, adjectives, and adverbs
- Incorporate more sophisticated WSD algorithms
- Concurrent information
 - Other dictionaries (e.g. Wiktionary)
 - VerbNet
 - FrameNet
 - etc.

Acknowledgements

The work was supported by a Sproull Graduate Fellowship from the University of Rochester and NSF grant IIS-1543758.



UNIVERSITY *of*
ROCHESTER

Semantic Representation Details



Semantic Representation Details

(Hobbs, 2008)¹ - Hobbsian Logical Form (HLF)

- Conflates events and propositions

John's telling of his favorite joke would make most listeners laugh; the proposition that he did so would not.

- Interpretation of quantifiers in terms of "typical elements" can lead to contradiction

“Typical elements” of sets are defined as individuals that are not members of those sets, but have all the properties *shared* by members of the sets.

Consider $S = \{0, 1\}$.

Share property of being in S .

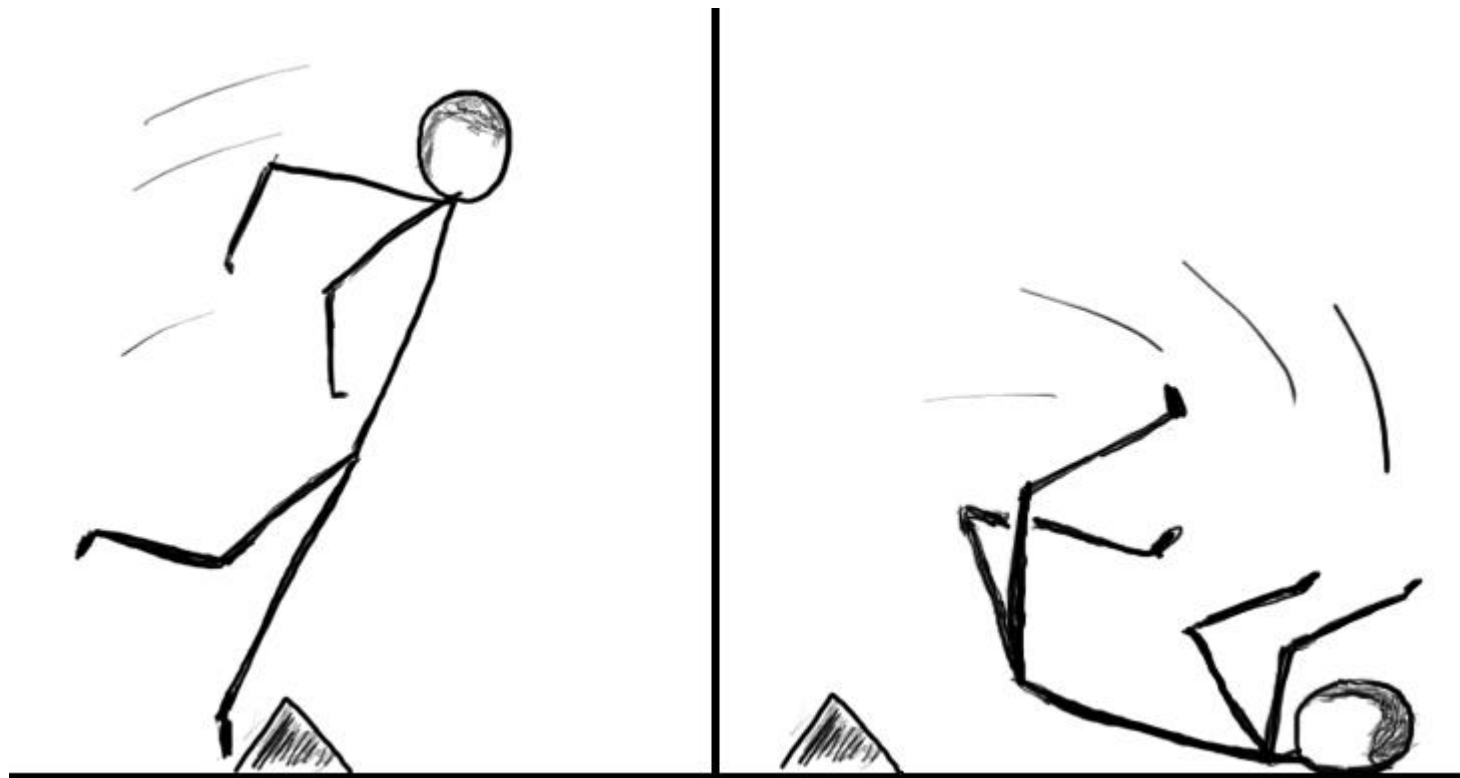
Typical element must be in S , but by definition, not in S !!!

Semantic Representation Details

(Allen et al. 2013)² - Description Logic (OWL-DL)

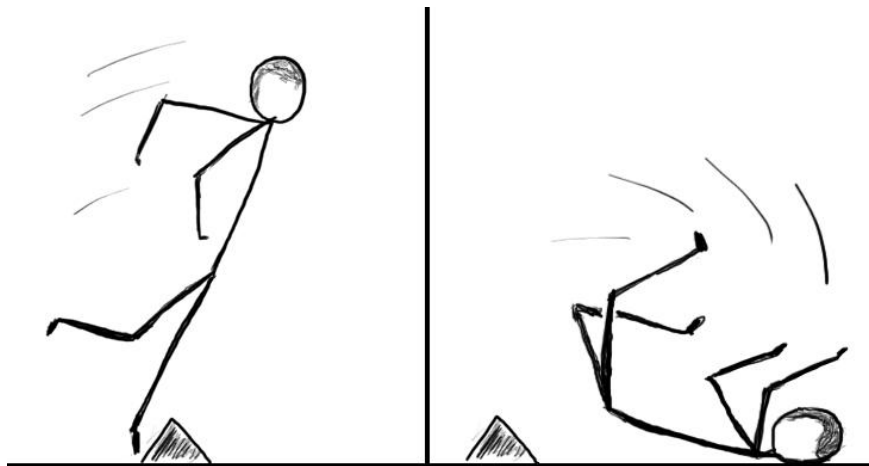
- OWL-DL: Web Ontology Language - Description Logic
 - Designed for ontologies, not full natural language
- Handling of predicate/sentence reification, predicate modification, self-reference, and uncertainty is unsatisfactory
 - Intersective predicate modification
“whisper loudly” \rightarrow **$whisper \sqcap \forall_{of}^{-1}.(loudly)$** \rightarrow **$speak \sqcap \forall_{of}^{-1}.(softly) \sqcap \forall_{of}^{-1}.(loudly)$**
 - Tree-shaped models requirement
 - *partOf* and *contains* relations in opposite directions not possible
 - review: “refresh one’s memory” - self-reference
 - Reification
 - Classes and individuals are disjoint \rightarrow can’t refer to a class as an individual

Example Inference in Detail



Example Inference in Detail

- “John stumbles, but does not fall” → “John misses a step and nearly falls” using the axiom from the WordNet entry for *stumble2.v*.
- *stumble2.v* : miss a step and fall or nearly fall
- Glossing over event handling details: [Φ ** e] simply written as Φ .



Example Inference

“John stumbles, but does not fall”

1. $((\text{John stumble2.v}) \text{ but.cc } \neg(\text{John fall123.v}))$

stumble2.v: miss a step and fall or nearly fall

2. $(\forall x, e ((x \text{ stumble2.v}) ** e) \rightarrow (((\exists z (z \text{ step2.n}) (x \text{ miss4.v } z)) \wedge ((x \text{ fall123.v}) \vee (x (\text{nearly.adv fall123.v})))) ** e)))$

If two statements are conjoined by “but”, then both statements are true

3. $(\forall wff x, y (\forall e ((x \text{ but.cc } y) ** e) \rightarrow ((x \wedge y) * e)))$

Example Inference

1. ((John stumble2.v) but.cc \neg (John fall23.v))

3. ($\forall_{wff x, y}$ ($\forall e$ (((x but.cc y) * e) \rightarrow ((x \wedge y) * e)))))

4. ((John stumble2.v) \wedge \neg (John fall23.v))

Example Inference

1. ((John stumble2.v) but.cc \neg (John fall123.v))

3. $(\forall_{wff x, y} (\forall e ((x \text{ but.cc } y) ** e) \rightarrow ((x \wedge y) * e)))$

4. $\frac{((\text{John stumble2.v}) \wedge \neg(\text{John fall123.v}))}{/}$

5. (John stumble2.v)

Example Inference

1. ((John stumble2.v) but.cc \neg (John fall23.v))

3. $(\forall_{wff x, y} (\forall e ((x \text{ but.cc } y) ** e) \rightarrow ((x \wedge y) * e)))$

4. ((John stumble2.v) \wedge \neg (John fall23.v))

5. (John stumble2.v)

6. \neg (John fall23.v)

Example Inference

5. (John stumble2.v)

6. \neg (John fall123.v)

2. $(\forall x, e ((x \text{ stumble2.v}) ** e)$
 $\rightarrow ((\exists z (z \text{ step2.n}) (x \text{ miss4.v } z)) \wedge$
 $((x \text{ fall123.v}) \vee (x \text{ (nearly.adv fall123.v)}))) ** e))$

Example Inference

5. (John stumble2.v)

6. $\neg(\text{John fall123.v})$

2. $(\forall x, e \left(((x \text{ stumble2.v}) ** e) \right.$
 $\left. \rightarrow ((\exists z (z \text{ step2.n}) (x \text{ miss4.v } z)) \wedge \right.$
 $\left. ((x \text{ fall123.v}) \vee (x (\text{nearly.adv fall123.v})))) ** e \right))$

7. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge$
 $((\text{John fall123.v}) \vee (\text{John (nearly.adv fall123.v)})))$

Example Inference

5. (John stumble2.v)

6. $\neg(\text{John fall123.v})$

2. $(\forall x, e ((x \text{ stumble2.v}) ** e) \rightarrow (((\exists z (z \text{ step2.n}) (x \text{ miss4.v } z)) \wedge ((x \text{ fall123.v}) \vee (x (\text{nearly.adv fall123.v})))) ** e))$

7. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\underline{(\text{John fall123.v})} \vee (\text{John (nearly.adv fall123.v)})))$

8. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\text{John (nearly.adv fall123.v)}))$

Example Inference

5. (John stumble2.v)

6. \neg (John fall123.v)

2. $(\forall x, e ((x\ stumble2.v) ** e) \rightarrow ((\exists z (z\ step2.n) (x\ miss4.v\ z)) \wedge ((x\ fall123.v) \vee (x\ (nearly.adv\ fall123.v)))) ** e))$

7. $((\exists z (z\ step2.n) (John\ miss4.v\ z)) \wedge ((John\ fall123.v) \vee (John\ (nearly.adv\ fall123.v))))$

8. $((\exists z (z\ step2.n) (John\ miss4.v\ z)) \wedge (John\ (nearly.adv\ fall123.v)))$

“John misses a step and nearly falls”

Example Inference

- This inference would lead to contradiction for representations using intersective modification! (e.g. OWL-DL)

8. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\text{John } (\text{nearly.adv fall123.v})))$

9. $(\text{John } (\text{nearly.adv fall123.v}))$

Example Inference

- This inference would lead to contradiction for representations using intersective modification! (e.g. OWL-DL)

8. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\text{John (nearly.adv fall23.v)}))$

9. $(\text{John (nearly.adv fall23.v)})$

10. $\forall_{\text{of}}(\text{John}) \sqcap \forall_{\text{of}}^{-1}(\text{nearly.adv}) \sqcap$
 fall23.v

Example Inference

- This inference would lead to contradiction for representations using intersective modification! (e.g. OWL-DL)

8. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\text{John (nearly.adv fall23.v)}))$

9. $(\text{John (nearly.adv fall23.v)})$

10. $\frac{\forall_{\text{of}}(\text{John}) \sqcap \forall_{\text{of}}^{-1}(\text{nearly.adv}) \sqcap \text{fall23.v}}{\quad}$

11. $\forall_{\text{of}}(\text{John}) \sqcap \text{fall23.v}$

Example Inference

- This inference would lead to contradiction for representations using intersective modification! (e.g. OWL-DL)

8. $((\exists z (z \text{ step2.n}) (\text{John miss4.v } z)) \wedge (\text{John (nearly.adv fall23.v)}))$

9. $(\text{John (nearly.adv fall23.v)})$

10. $\forall_{\text{of}}(\text{John}) \sqcap \forall_{\text{of}}^{-1}(\text{nearly.adv}) \sqcap$
 fall23.v

11. $\forall_{\text{of}}(\text{John}) \sqcap \text{fall23.v}$ - (i.e. **(John fall23.v)**) **Contradicts original parsed sentence!**
“John stumbled, **but did not fall**”