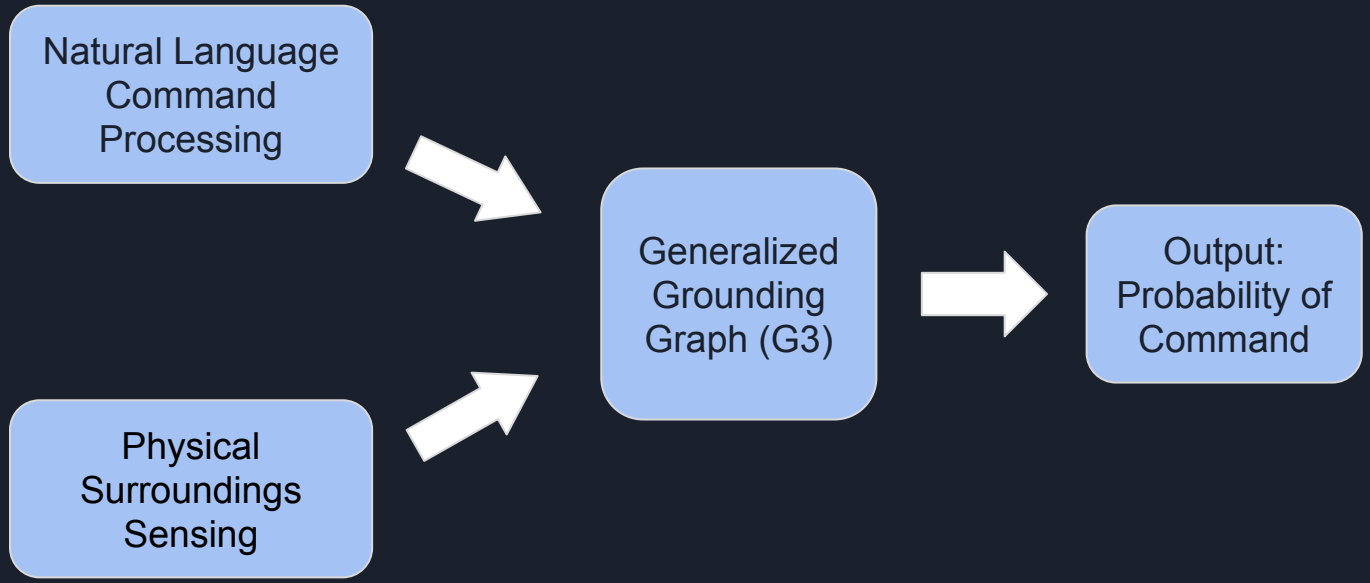
Implementation of Generalized Grounding Graphs(G3)

Presenter: Yutong Li

Motivation



Ref: Generalized Grounding Graphs: A Probabilistic Framework for Understanding Grounded Language

Solution Example

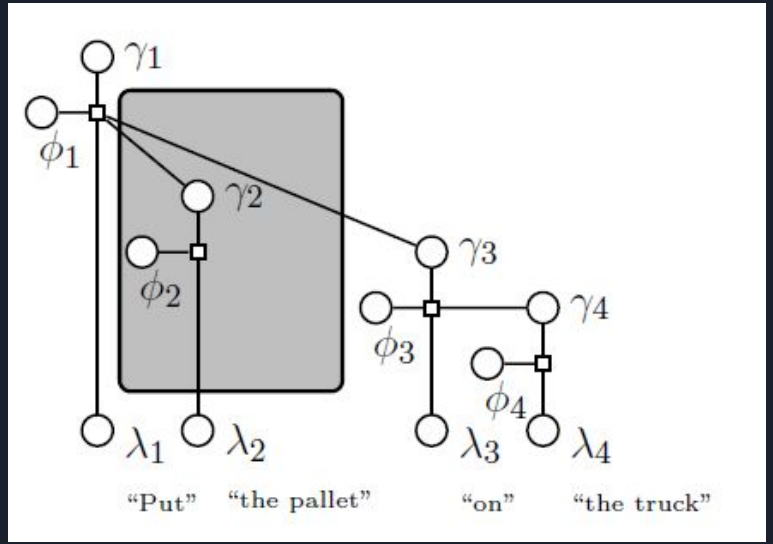


Figure1: G3 Graph of "Put the pallet on the truck"

Ref: Generalized Grounding Graphs: A Probabilistic Framework for Understanding Grounded Language

Goal:

Find γ1...γN such that maximize

p(Φ|Λ, γ1...γN)

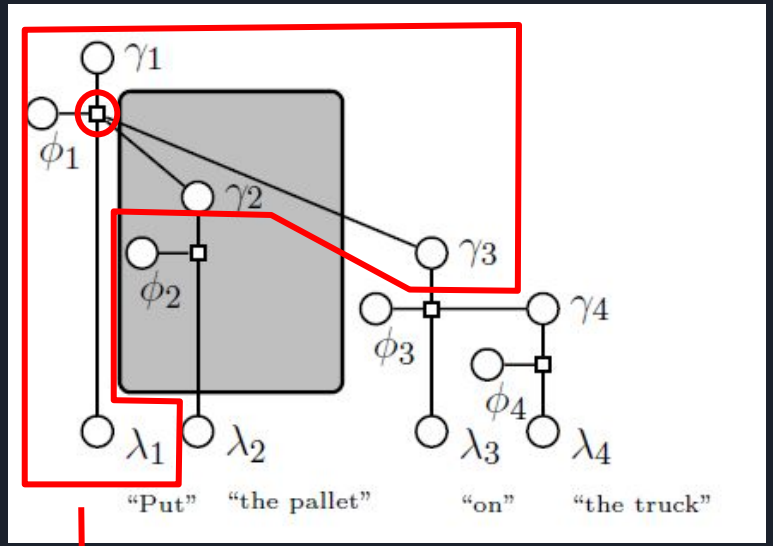
Annotations:

λ = Random Variables for the Language

γ = Groundings
(Object, Place or Path, Action)

Φ = Corresponding Variables of λ and γ(T or F)

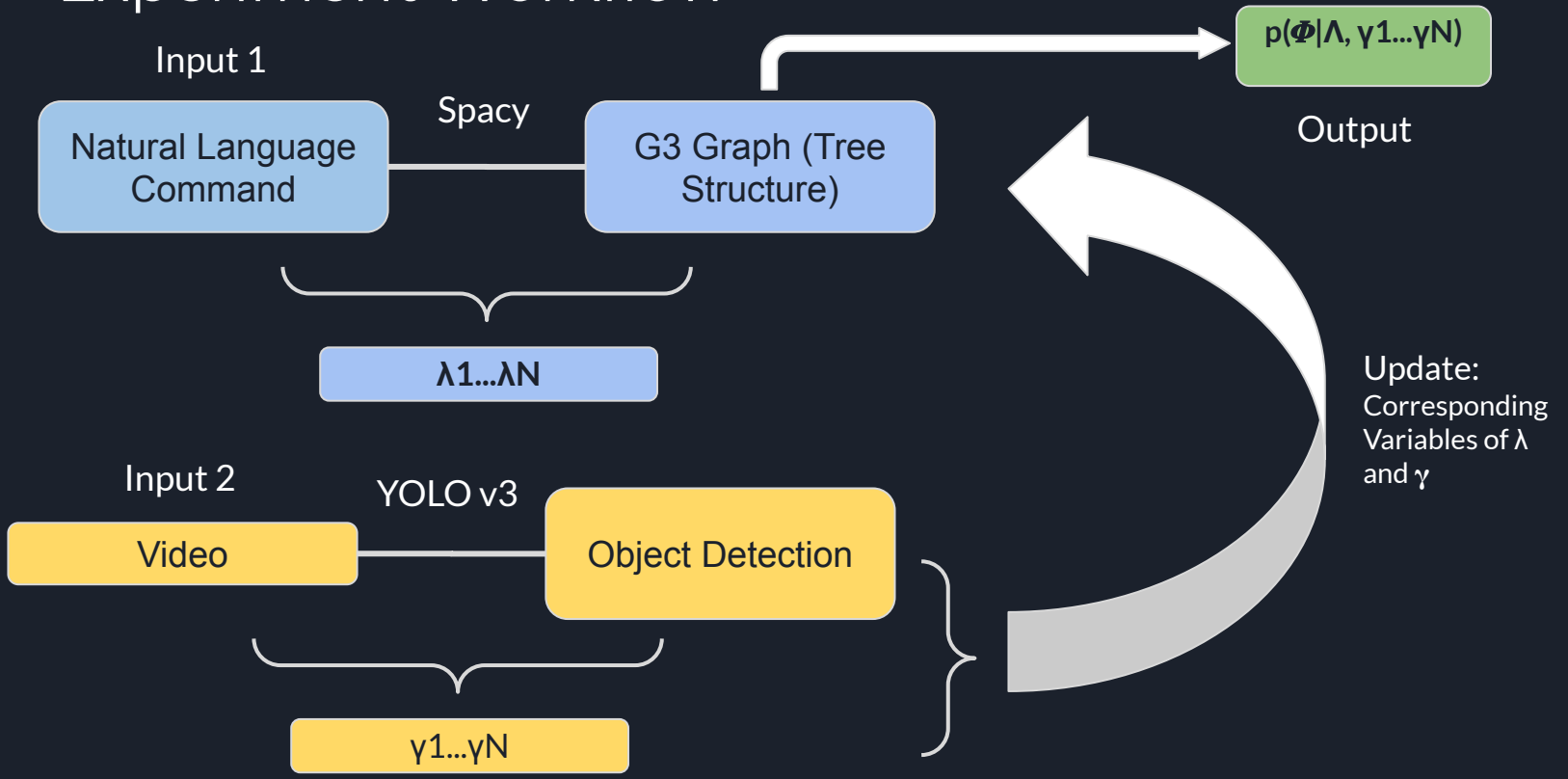
Solution Example - Calculation



$p(\Phi|\lambda1, \lambda2, \lambda3, \lambda4, \gamma1, \gamma2, \gamma3, \gamma4)$
$= p(\varphi1|\gamma1, \gamma2, \gamma3, \lambda1 = \text{Put}) \times p(\varphi2|\gamma2, \lambda2 = \text{the pallet}) \times p(\varphi3|\gamma3, \gamma4, \lambda3 = \text{on}) \times p(\varphi4|\gamma4, \lambda4 = \text{the truck})$

Ref: Generalized Grounding Graphs: A Probabilistic Framework for Understanding Grounded Language

Experiment Workflow



Experiment - More Details

- Natural Language Processing
  - Tool: Spacy
  - Target Keywords
    - Noun- Objects
    - Verb - Event
    - Preposition - place/path
- Objects Detection
  - Tool: Yolo v3
  - Confidence: 0.8
  - Trained Dataset
    - COCO
      - 330K images
      - 80 Objects

Experiment - Result

Command: Put those books on the chair

```
Semantic Language G3 Map - Before Sensing
 └── Put(EVENT) Corresponding Variance Phi:  1
    ├── those books(OBJECT) Corresponding Variance Phi:  0
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  0
Probability of language command:  0
```

Books Detected

```
 └── Put(EVENT) Corresponding Variance Phi:  1
    ├── those books(OBJECT) Corresponding Variance Phi:  1
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  0
Probability of language command:  0
```

Chair Detected

```
 └── Put(EVENT) Corresponding Variance Phi:  1
    ├── those books(OBJECT) Corresponding Variance Phi:  1
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  1
Probability of language command:  1
```

Figure 2: G3 Graph of “Put those books on the chair”

Experiment - Result Cont.

Command: Put those books on the chair

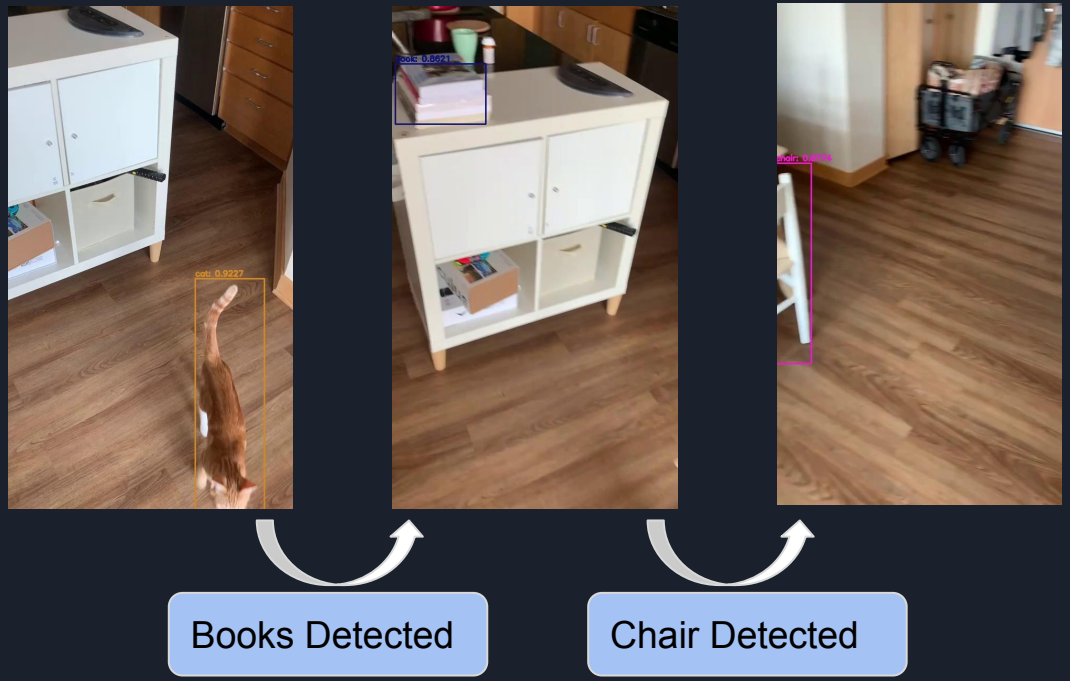


Figure 3: First frame of the video

Experiment - Result

Command: Jump on the chair

```
Semantic Language G3 Map - Before Sensing
 └── jump(EVENT) Corresponding Variance Phi:  0
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  0
Probability of language command:  0
```

Chair Detected

```
 └── jump(EVENT) Corresponding Variance Phi:  0
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  1
Probability of language command:  0
```

```
 └── jump(EVENT) Corresponding Variance Phi:  0
    └── on(PLACES OR PATHS) Corresponding Variance Phi:  1
        └── the chair(OBJECT) Corresponding Variance Phi:  1
Probability of language command:  0
```

End

Figure 4: G3 Graph of "Jump on the chair"

Experiment - Result Cont.

Command: Jump on the chair

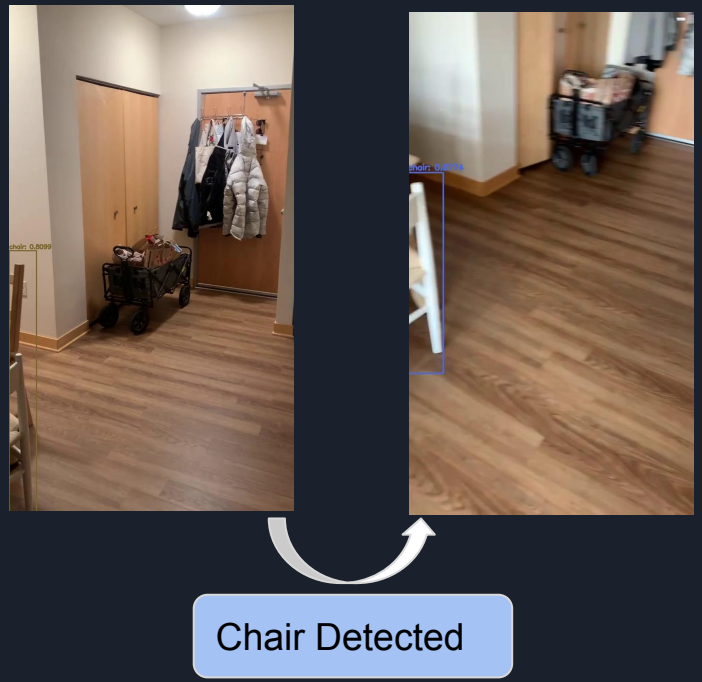


Figure 5: First frame of the video

Experiment - Result Cont.

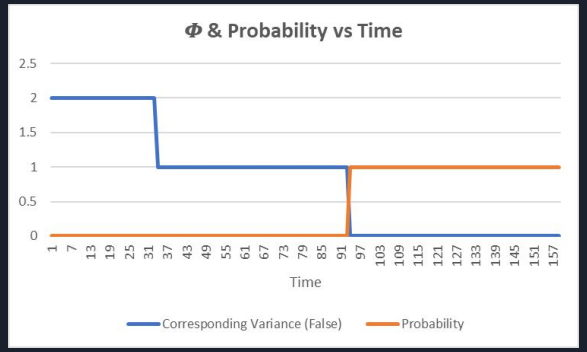


Figure 6: **Φ & Probability vs Time**
(Put those books on the chair)

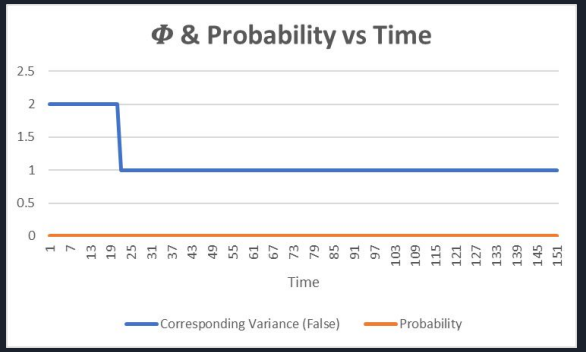


Figure 7: **Φ & Probability vs Time**
(Jump on the chair)

Challenges

- Language Processing
  - Dependency analysis
  - Event words ambiguity
- Object Detection
  - Objects detected certainty
  - More object categories
- Language Input
  - Voice

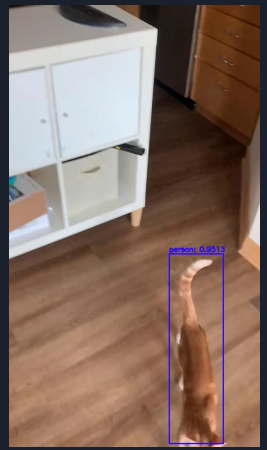


Figure 8: Cat labeled as person

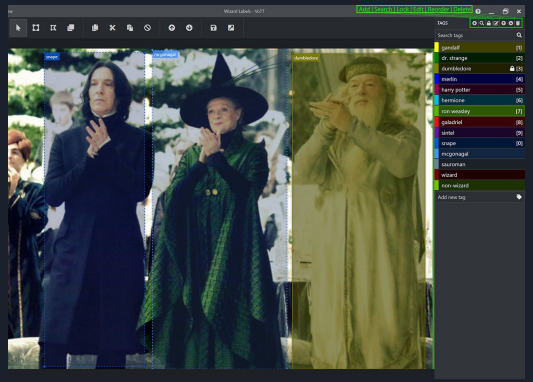


Figure 9: **VoTT Label Tool**

Questions?

Appendix - G3 Graph Generation from natural language input

Algorithm 1 Generating a grounding graph from natural language input.

Input:

Parsed natural language input, $\Lambda = \lambda_1 \dots \lambda_M$, with root $\lambda_{root} \in \Lambda$

$\Phi \leftarrow \phi_1 \dots \phi_M$
$\Gamma \leftarrow \gamma_1 \dots \gamma_M$
$F \leftarrow []$
for $\lambda_i \in \Lambda$ **do**
    $\Gamma_{children} \leftarrow []$
    for $\lambda_{child} \in children(\lambda_i)$ **do**
        Add γ_{child} to $\gamma_{children}$
    end for
    Add $(\phi_i, \lambda_i, \gamma_i, \Gamma_{children})$ to F
end for

Output: Λ, Φ, Γ, F