

# Context Model Acquisition from Spoken Utterances

Sebastian Weigelt, Tobias Hey, and Walter F. Tichy

KIT – Department of Informatics – Institute for Program Structures and Data Organization (IPD).



Intelligent Assistants seem to be  
quite smart today!

*How far away is the sun?*

*When is my next meeting?*

*Is my daughter at home?*

*Set a timer for five minutes*

*Call my brother at work*

*How many dollars is 45 euro*

*Remind me to call mom*

*Google the war of 1812*

*Who is near me?*

*Give me directions home*

*Who is near me?*

*Text Brian I'm on my way*

*Find Disney movies*

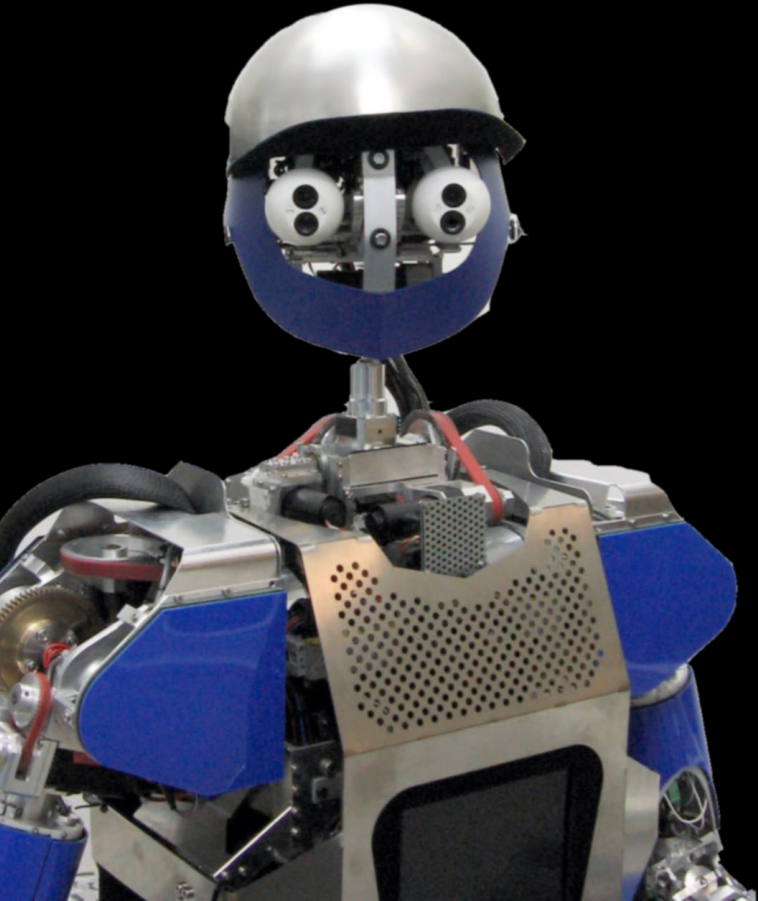
*Play iTunes Radio*

*Should I bring an umbrella?*

*What's trending on twitter?*

But do they really understand?

“Shut all kitchen windows”  
“Leave the one over the sink open”



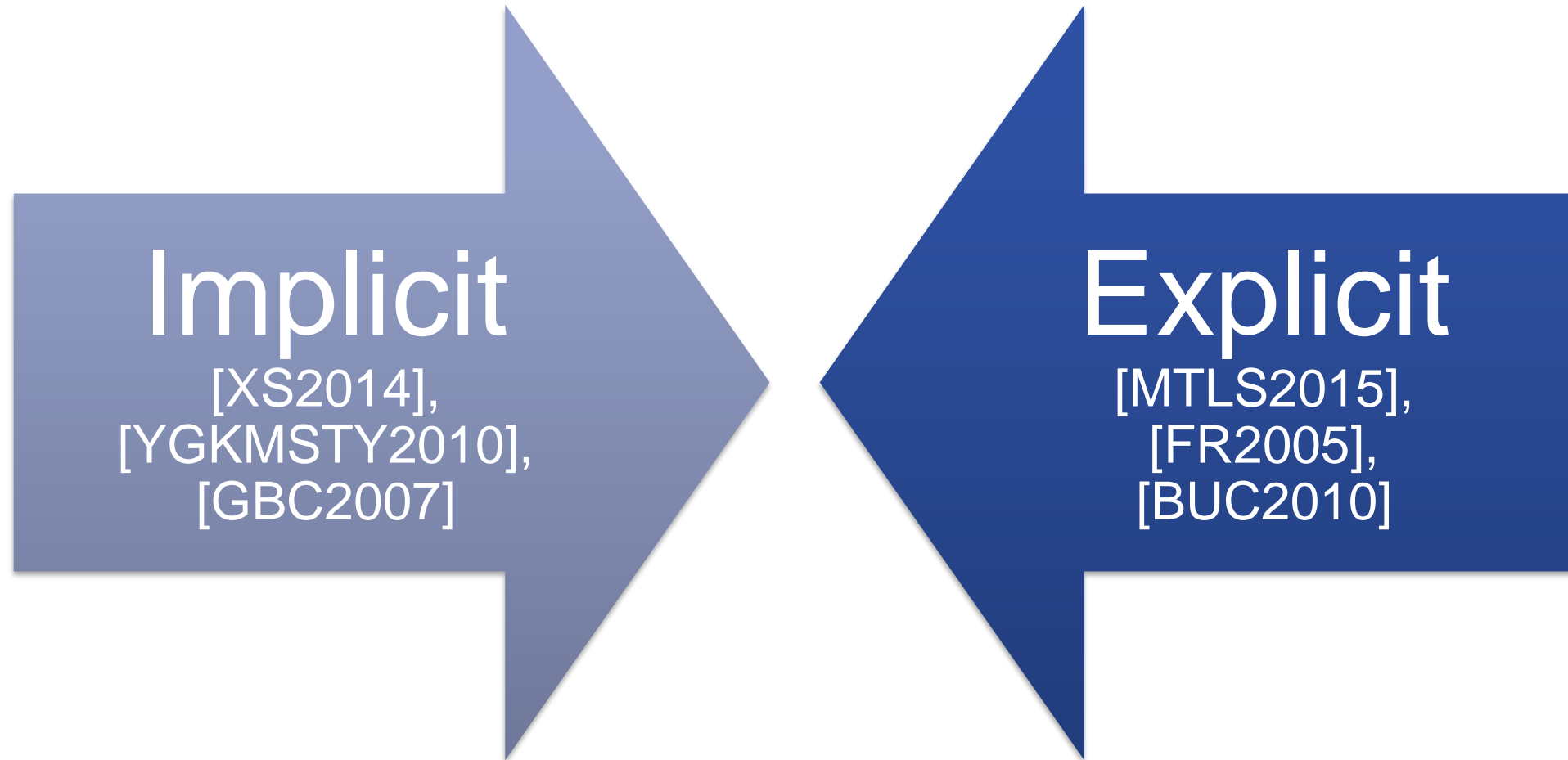
“Close the fridge”  
“Open the dishwasher”  
“Then close all open appliances”

# The Idea:

## A comprehensive context model

# What is Context?

- **Context** describes information that is used to understand the meaning of an **artifact** [D1992], [F2004], [S1999]
- Context relations (include but are not limited to)
  - Situational (place, time, ...)
  - Social (relations between communication partners, mutual knowledge assumptions, ...)
  - Discoursal (grammatical, semantic, ...)
- Artifacts are parts of a communication situation (e.g. words or phrases)





# Context in spoken utterances

## Hierarchical

- Super Concept
- Part-Of

- REFRIGERATOR → WHITE\_GOODS → APPLIANCE
- REFRIGERATOR → REFRIGERATOR\_DOOR

## Conceptual

- Concept
- State

- *the fridge* → REFRIGERATOR
- *close it* → CLOSED

## Individual

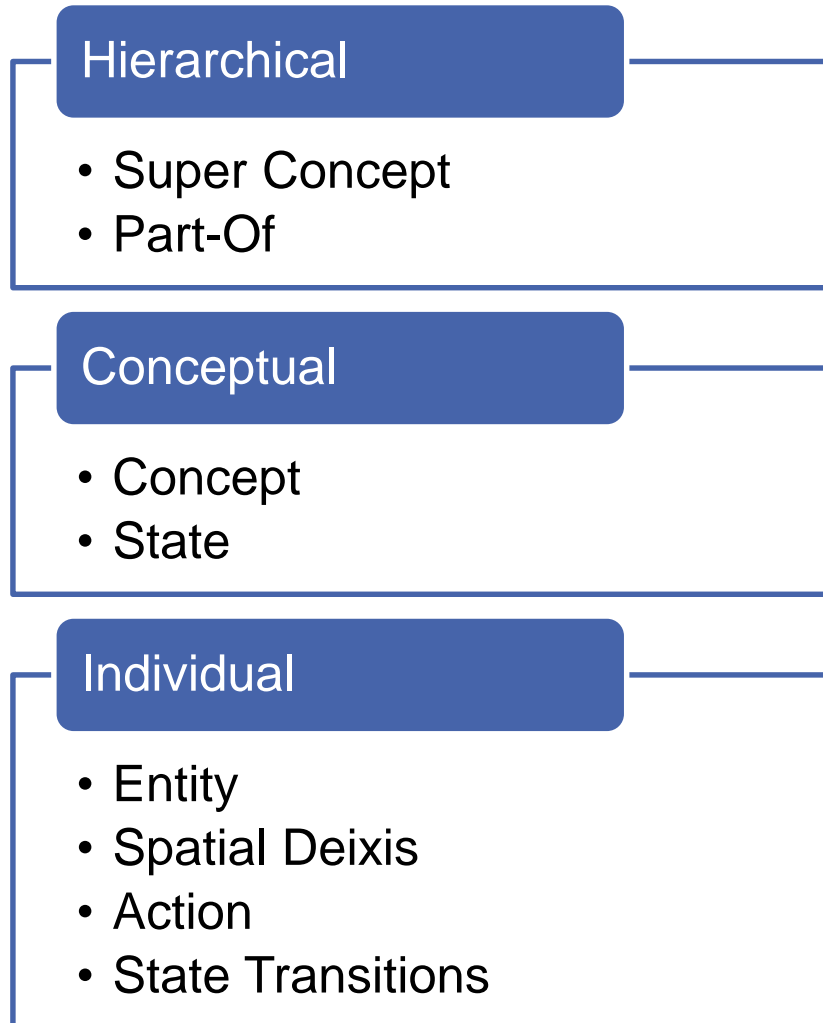
- Entity
- Spatial Deixis
- Action
- State Transitions

- “Go to *the fridge*...”
- “...*next to* the cupboard...”
- “...*open* its door...”
- “...and *close it* again”

# Context Acquisition – Prerequisites

- Input: Long (and complex) spoken Utterances (no dialog)
- Domain: Open (Examples: Robot in a kitchen setting)
- Preprocessing: Shallow parsing, POS, SRL, NER, Stemming & WSD
- Knowledge: WordNet, generic interchangeable domain model

# Context Acquisition – Analysis



- Knowledge-based
  - WordNet: Entities (Synsets)  
Synonyms, Hypernyms, Hypernyms, Meronyms,...
  - Domain ontology: States, Concepts, Meronyms,...
- Rule-based
  - Parsing: Chunk sequences (artifacts)
  - POS: Part of speech patterns
  - NER: Entities
  - Stemming: Word base forms

# Context Acquisition – Analysis (Example)

## Hierarchical

- Super Concept
- Part-Of

## Conceptual

- Concept
- State

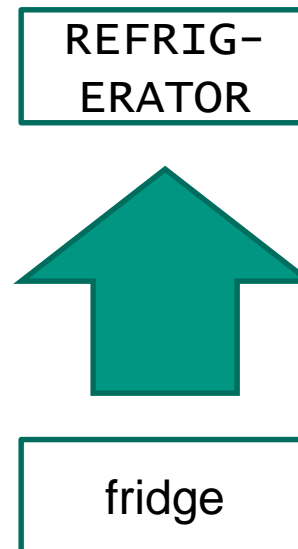
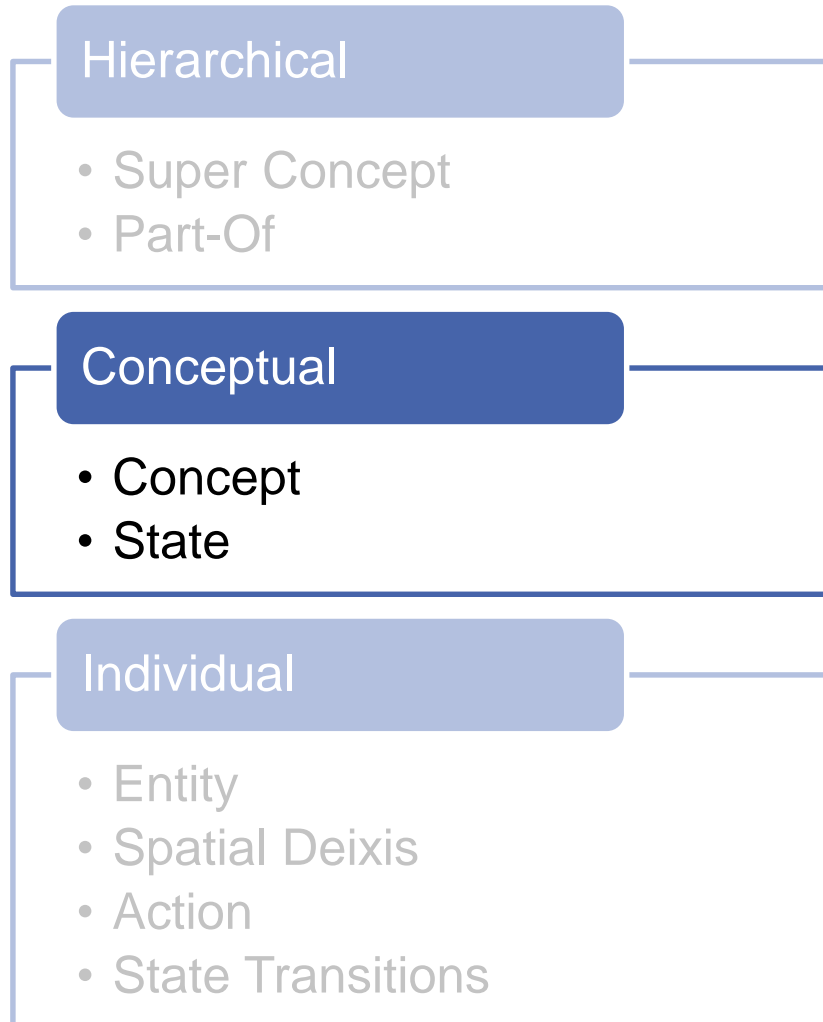
## Individual

- Entity
- Spatial Deixis
- Action
- State Transitions

fridge

1. Extract noun phrases (Parsing)  
→ Entity := head of noun phrase
2. Extract additional information  
Chunks:
  - i. Determiner
  - ii. Adjectives
  - iii. Number
 WordNet:
  - iv. Synonyms

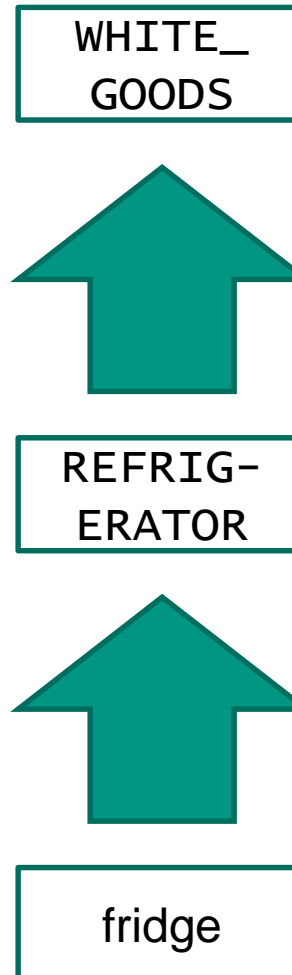
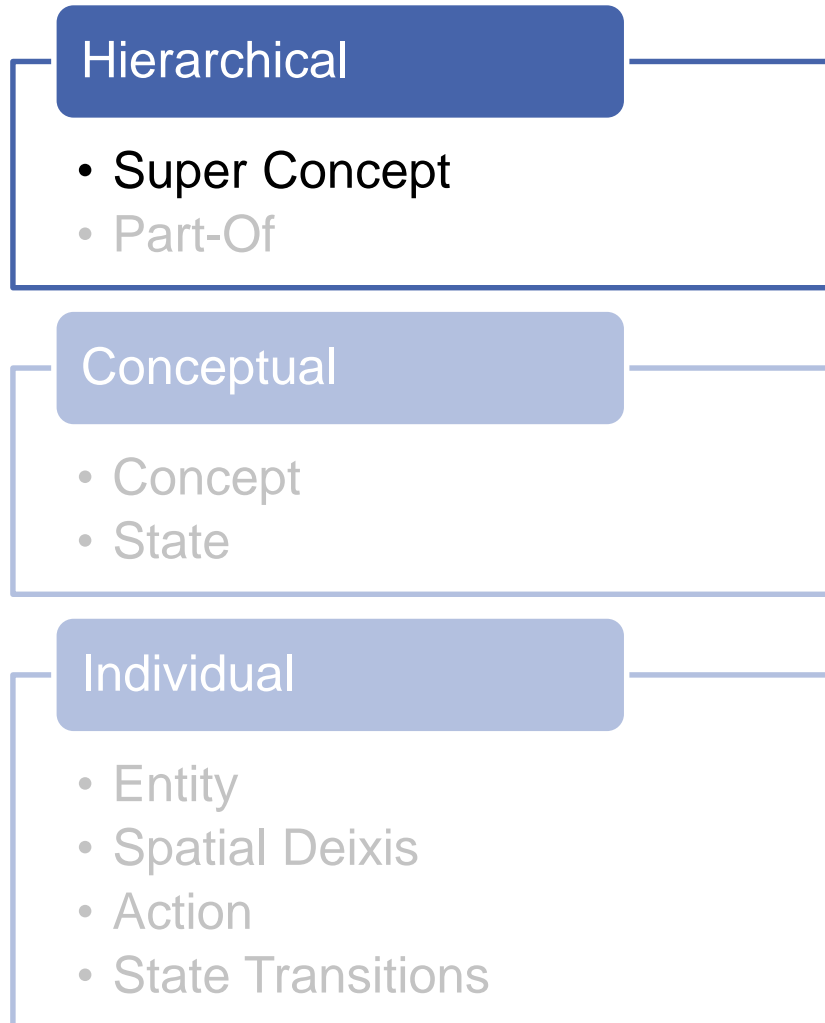
# Context Acquisition – Analysis (Example)



1. Matching (Entities → Concepts)\*
  - i. Existing concept
  - ii. Ontology concept
    - i. Direct match
    - ii. Synonym Match
    - iii. Sub-phrases
  - iii. WordNet concept
2. Extract additional Information
  - i. Synonyms
  - ii. States (Domain ontology)

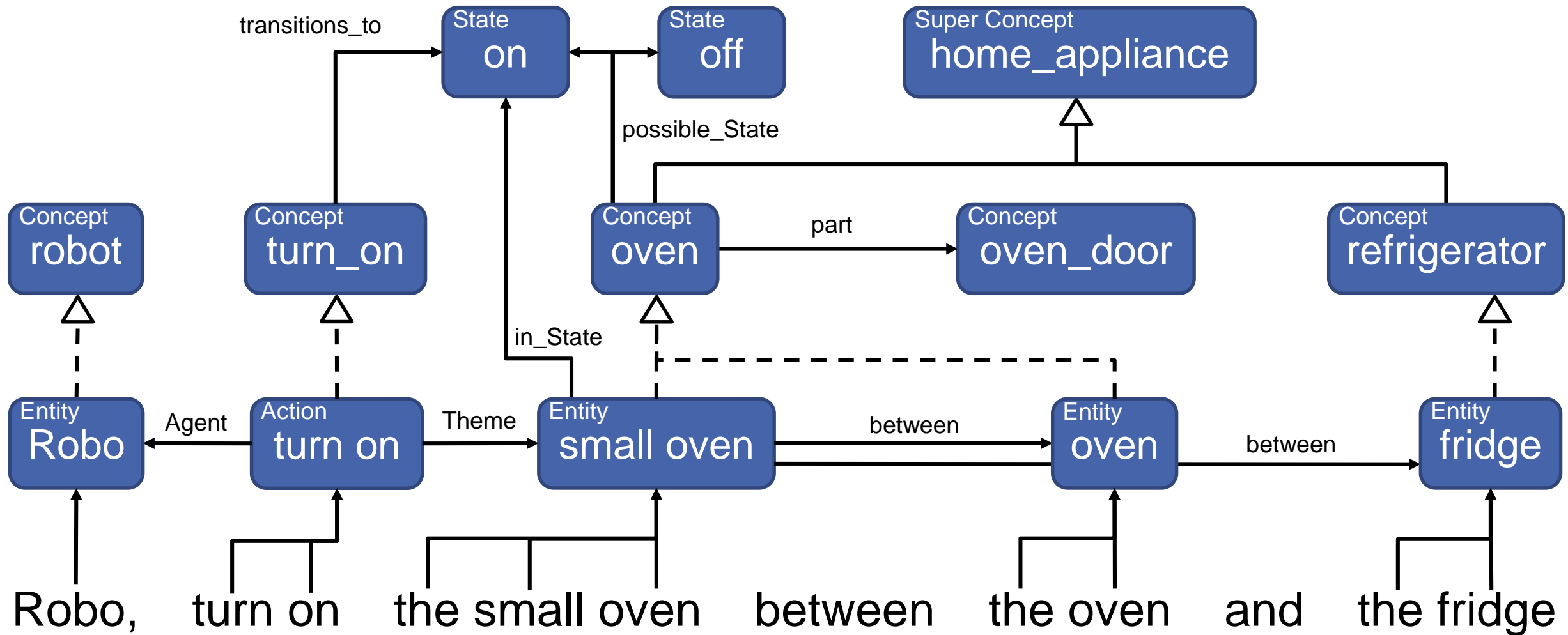
\*Jaro-Winkler distance threshold: 0.92

# Context Acquisition – Analysis (Example)



1. Extract super concept  
For all entity pairs:
  - i. Determine lowest common subsumer (LCS) in WordNet
  - ii. Filter generic terms
    - i. Hierarchy Threshold
    - ii. Wu & Palmer similarity (threshold 0.7)
2. Repeat 1. for super concepts

# Context Acquisition – Output Example



# Evaluation – Setting



- 10 subjects
- 2 tasks/scenarios
- Long and complex descriptions

*Go to the fridge open the door. Take the water bottle.  
 Close the door and go to the table. Open the water bottle.  
 Fill the green cup with water. [...]*

	Scenario 1	Scenario 2	Total
Recordings	10	11	21
Words	734	811	1545
Phrases	467	543	1010
Instructions	121	143	234



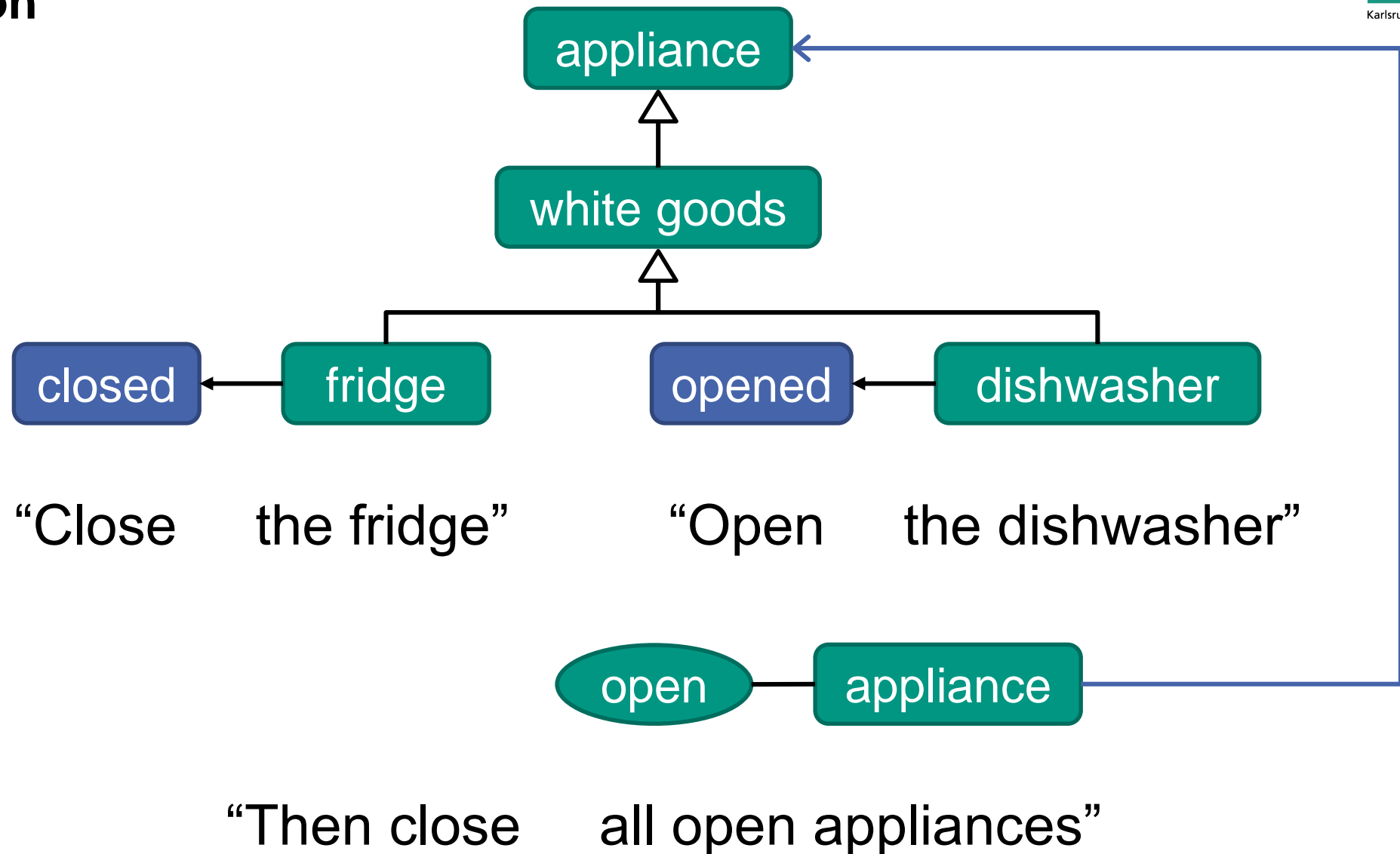
# Evaluation – Results

Context Type	Quantity
Entity	432
Spatial Deixis	84
Action	274
State Transition	81
Concept	594
State	18
Super Concept	144
Part-Of Relation	78

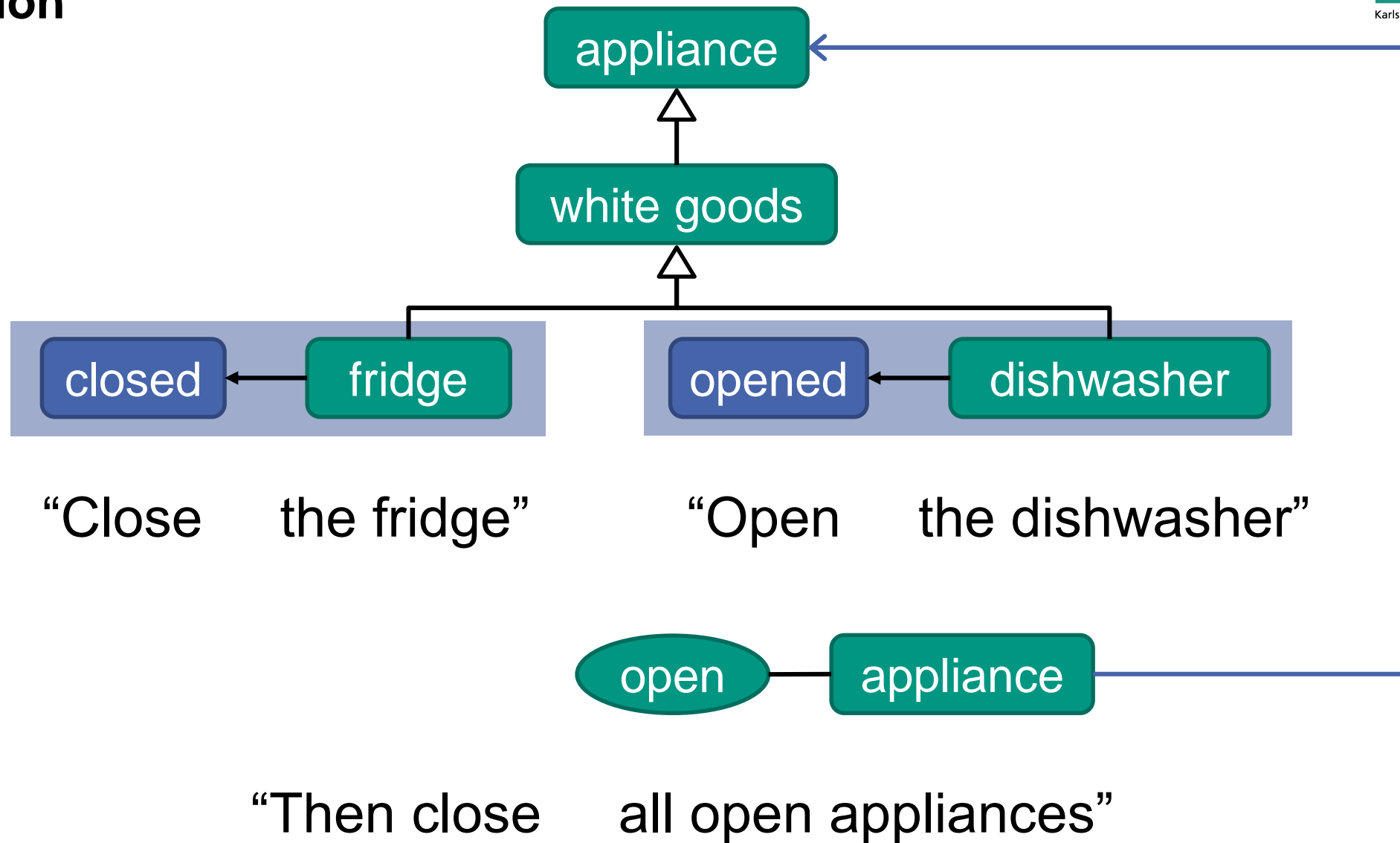
# Evaluation – Results

Context Type	Quantity	Precision	Recall	F <sub>1</sub>
Entity	432	0.972	0.975	0.973
Spatial Deixis	84	0.945	0.793	0.862
Action	274	0.852	0.762	0.804
State Transition	81	0.854	0.627	0.723
Concept	594	0.986	0.974	0.981
State	18	1.000	0.955	0.977
Super Concept	144	0.680	0.932	0.786
Part-Of Relation	78	0.897	0.959	0.927

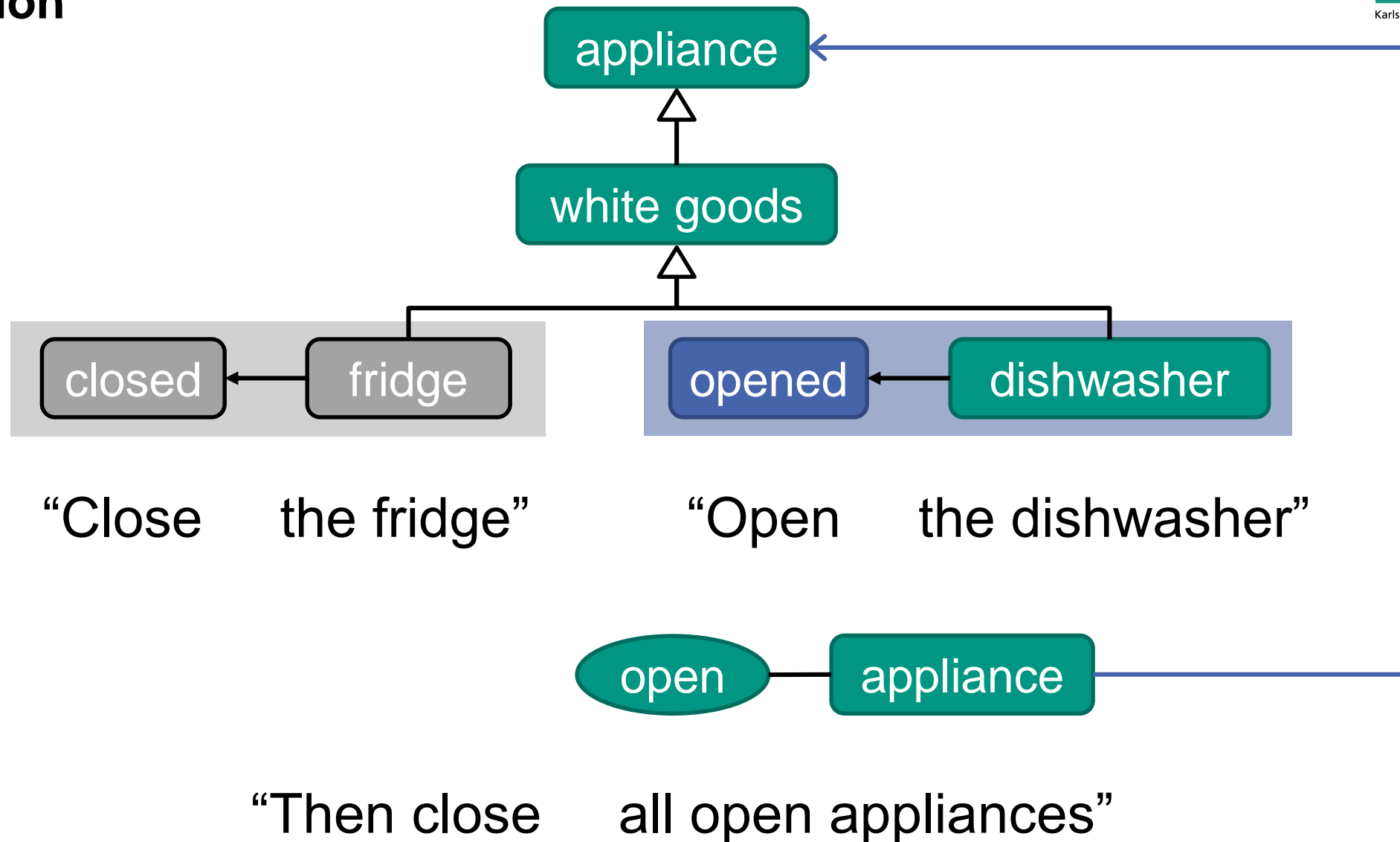
# Application



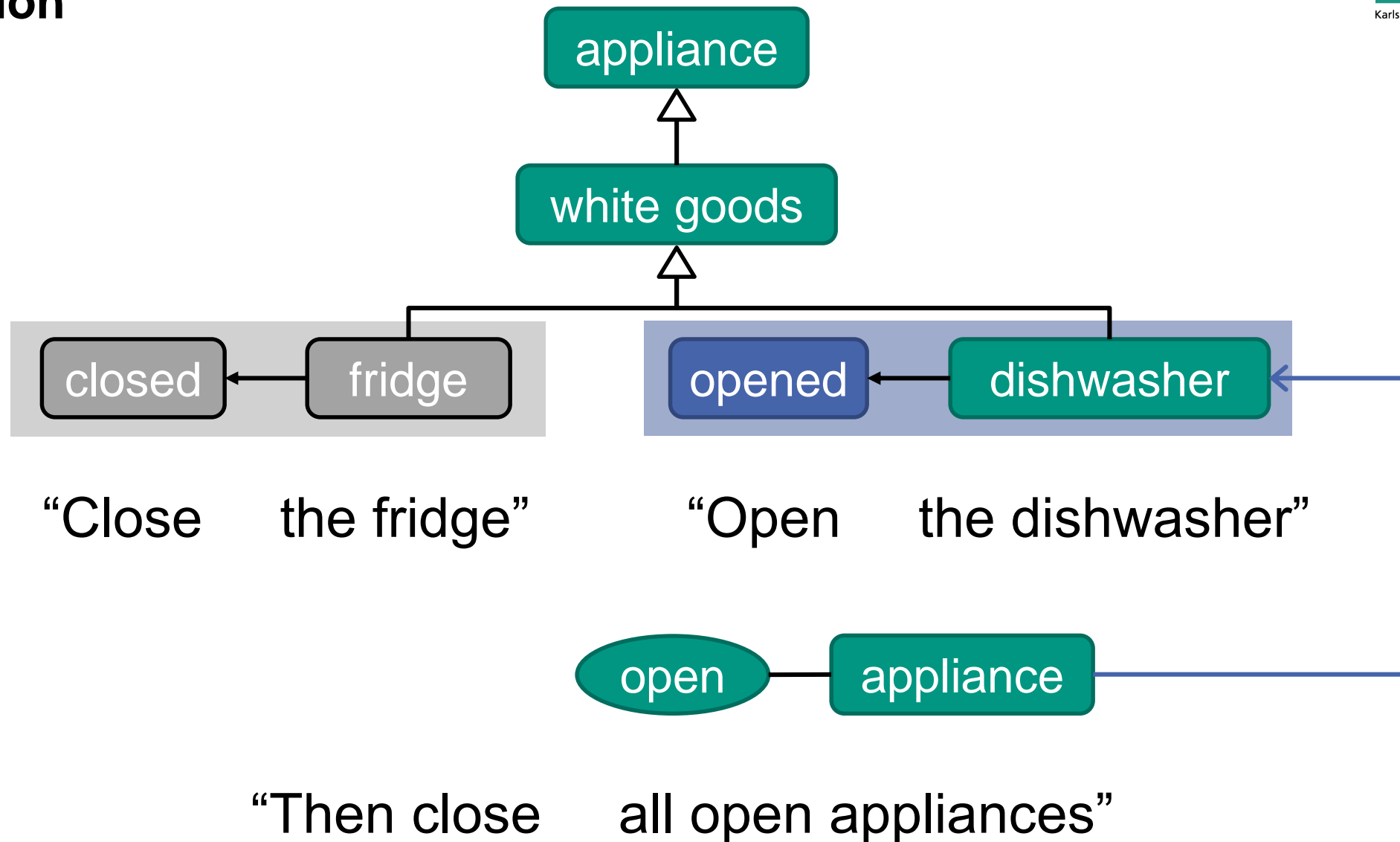
# Application



# Application



# Application



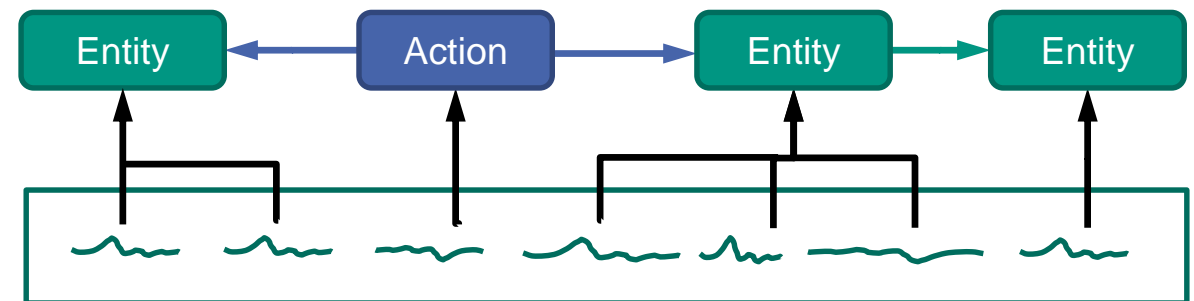
# Conclusion

- Context model acquisition from spoken utterances



# Conclusion

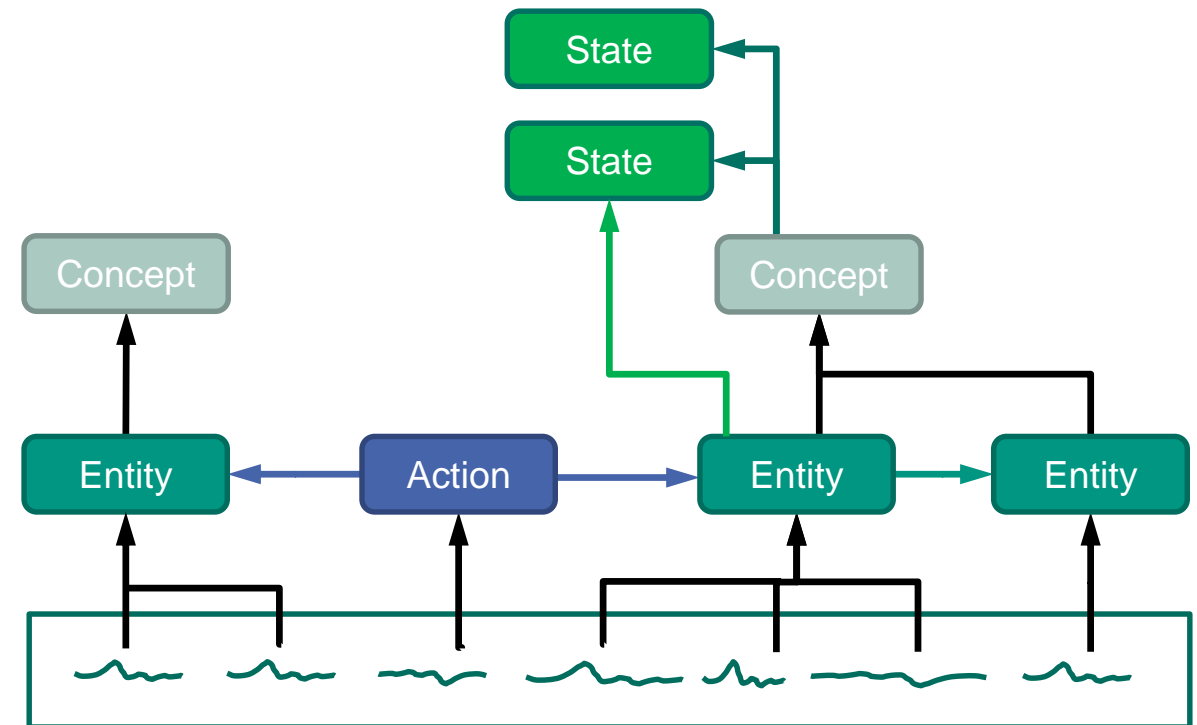
- Context model acquisition from spoken utterances
  - Entities and actions





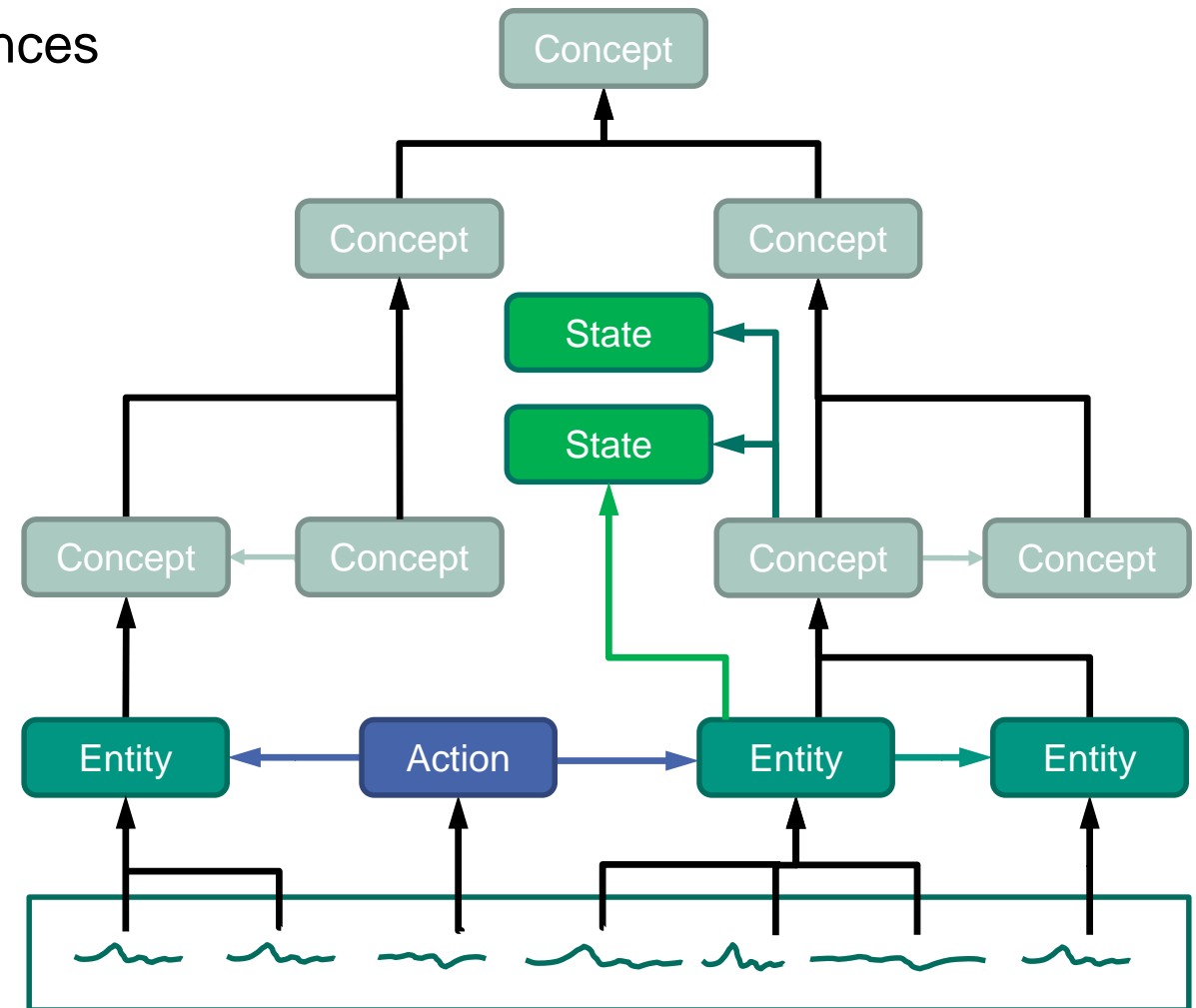
# Conclusion

- Context model acquisition from spoken utterances
  - Entities and actions
  - Concepts and states



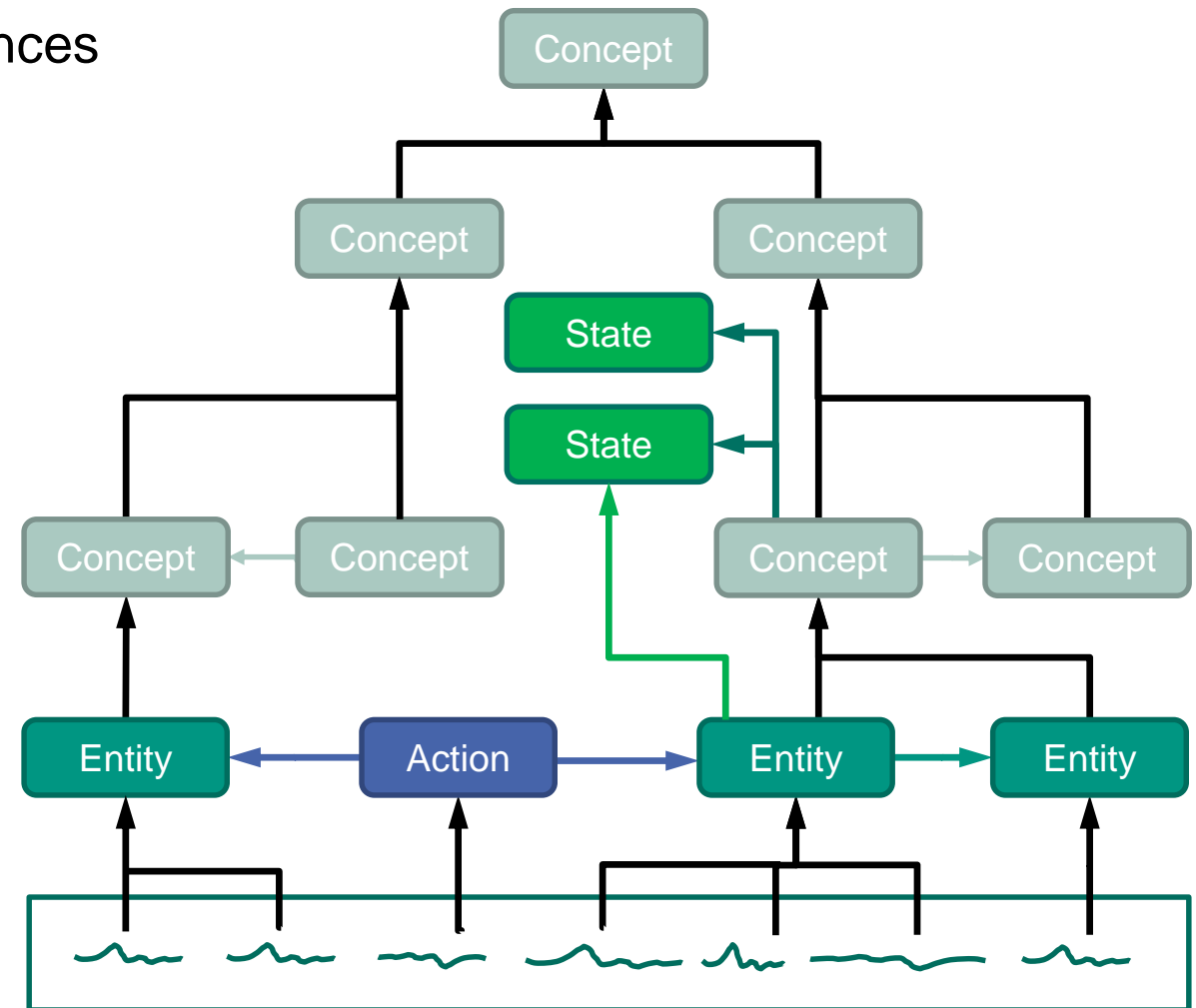
# Conclusion

- Context model acquisition from spoken utterances
  - Entities and actions
  - Concepts and states
  - Part-of and super concepts



# Conclusion

- Context model acquisition from spoken utterances
  - Entities and actions
  - Concepts and states
  - Part-of and super concepts
  
- Understanding of long spoken utterances
  
- Rule- and knowledge-based approach (F1: 0.723 - 0.981)
  
- Future Work
  - Extension (more types of context)
  - Application (support other analysis)
  - Statical methods (machine learning)



# References

- [D1992] A. Duranti, Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, May 1992.
- [F2004] A. Fetzer, Recontextualizing Context: Grammaticality Meets Appropriateness. John Benjamins Publishing, 2004.
- [S1999] R. Stalnaker, Context and Content: Essays on Intentionality in Speech and Thought. Oxford University Press, 1999.
- [XS2014] P. Xu and R. Sarikaya, “Contextual domain classification in spoken language understanding systems using recurrent neural network,” in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2014.
- [YGKMSTY2010] S. Young, M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu, “The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management,” Computer Speech & Language, vol. 24, no. 2, Apr. 2010.
- [GBC2007] D. Guzzoni, C. Baur, and A. Cheyer, “Modeling Human-Agent Interaction with Active Ontologies.” in AAAI Spring Symposium: Interaction Challenges for Intelligent Assistants, 2007.
- [MTLS2015] D. K. Misra, K. Tao, P. Liang, and A. Saxena, “Environment-Driven Lexicon Induction for High-Level Instructions,” ACL (1), 2015.
- [FR2005] M. Fleischman and D. Roy, “Intentional Context in Situated Natural Language Learning,” in Proceedings of the Ninth Conference on Computational Natural Language Learning, ser. CONLL '05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005.
- [BUC2010] A. Bordes, N. Usunier, R. Collobert, and J. Weston, “Towards understanding situated natural language,” in International Conference on Artificial Intelligence and Statistics, 2010.