Empirically Evaluating the Semantic Qualities of Language Vocabularies

Sotirios Liaskos
York University
Canada

Shakil M. Khan
University of Regina
Canada

John Mylopoulos
University of Toronto
Canada

Outline

1. Background and Motivation
   - Concepts and conceptualizations vs. terms and vocabularies.
   - When is a vocabulary appropriate good for a given conceptualization?
2. Key Idea
   - Empirically evaluate the appropriateness of terms we use to refer to concepts.
   - Identify and precisely describe vocabulary problems using an existing misalignment characterization framework.
3. Application

Conceptualizations, Languages and Ontological Commitments

We define[1]:

A system *S* we are interested in modeling.

A set *D* of distinguished elements of *S*.

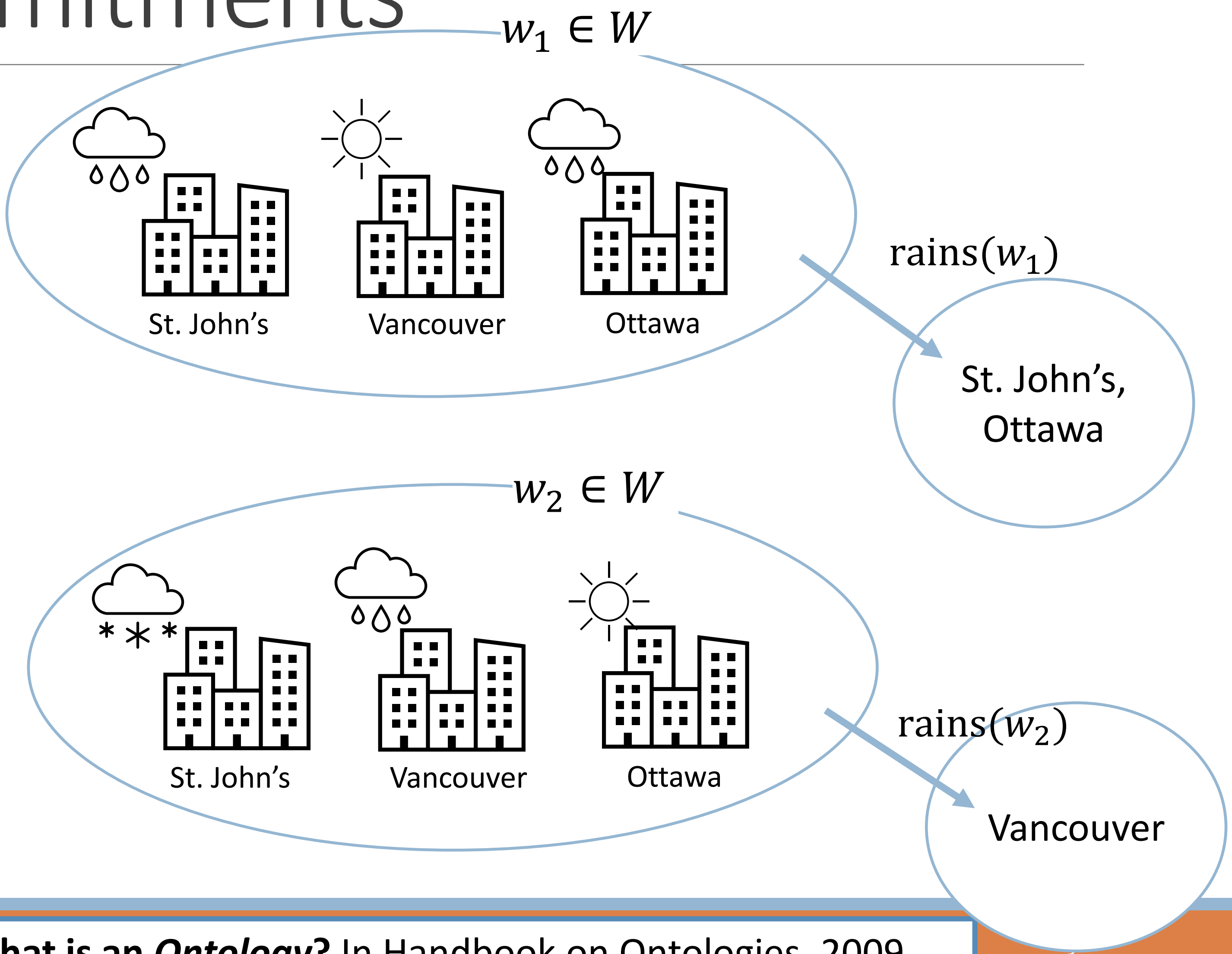
A set *W* of possible worlds (states of *S*)

Conceptual Relation (**Concept**)

- $\rho^n : W \longmapsto 2^{D^n}$

Example

- *ρ = rains*
- *D* = {St. John's, Vancouver, Ottawa}



$w_1 \in W$

St. John's

Vancouver

Ottawa

$\text{rains}(w_1)$

St. John's, Ottawa

$w_2 \in W$

St. John's

Vancouver

Ottawa

$\text{rains}(w_2)$

Vancouver

[1] N. Guarino, D. Oberle, S. Staab. **What is an *Ontology*?** In Handbook on Ontologies, 2009

Conceptualizations, Languages and Ontological Commitments

We need to use common terms to represent and communicate concepts.

- Example: use the string “rains” to describe the concept *rains*
  - Could have used “βρέχει” or “باران میبارد”.

The language L consists of (among other things):

- A set of concepts $\mathfrak{R}$
- Terms for representing the concepts in $\mathfrak{R}$: $V_{\mathfrak{R}}$ (the **vocabulary**)
- E.g., we use the English term $V_{\mathfrak{R}}$ = {“rains”} to represent the concept *rains*.

Let also a UoD *D* represented using vocabulary V_D

- E.g., terms V_D = {“St. John’s”, “Vancouver”, “Ottawa”} represent the corresponding cities.

Extension $I(v)$ of a term in $v \in V_{\mathfrak{R}}$*:*

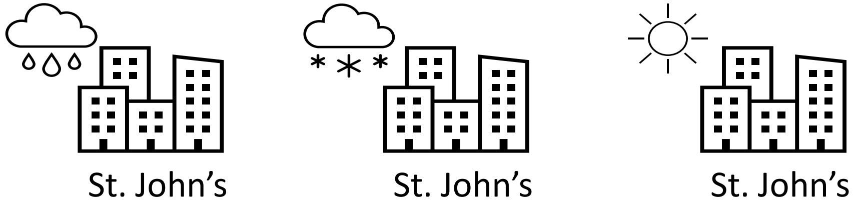
- *A subset of* $V_D \cup$ **R** to which v maps. [**R** is the set of n-tuples from V_D].

Extension of $I(\text{“rains”})$ can be, e.g., {“St. John’s”, “Vancouver”}, { “Vancouver”} or {}.

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Conceptualizations, Languages and Ontological Commitments

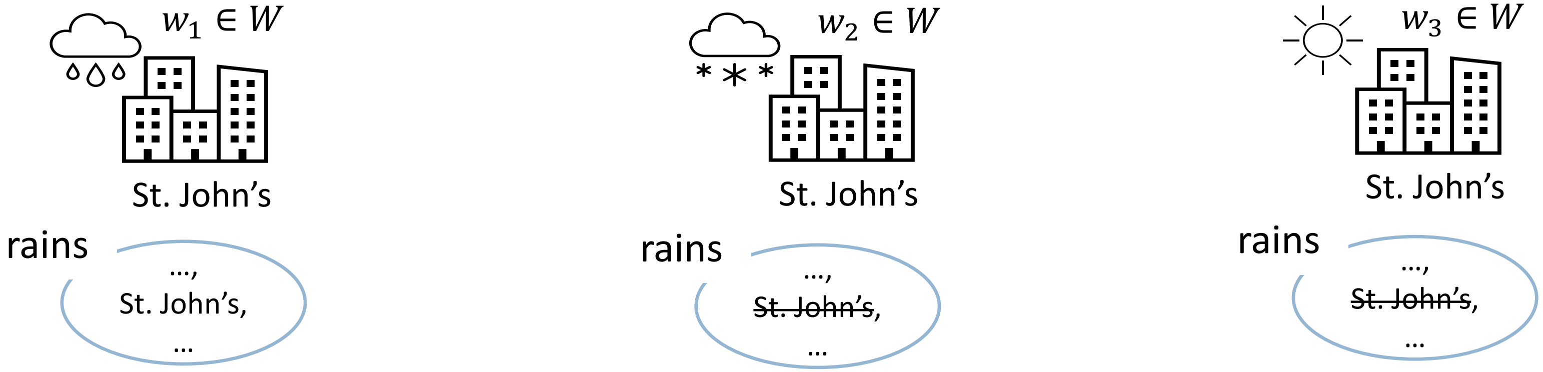
L is agnostic wrt. how the term “rains” is supposed to be used.



“St. Johns”$\in I$(“rains”) ???

Ontological commitment:

- Map language terms from $V_{\mathfrak{R}}$ to concepts in $\mathfrak{R}$. The terms are then *meaningful*: given a state of the world some extensions of “rains” are admissible while others are not.



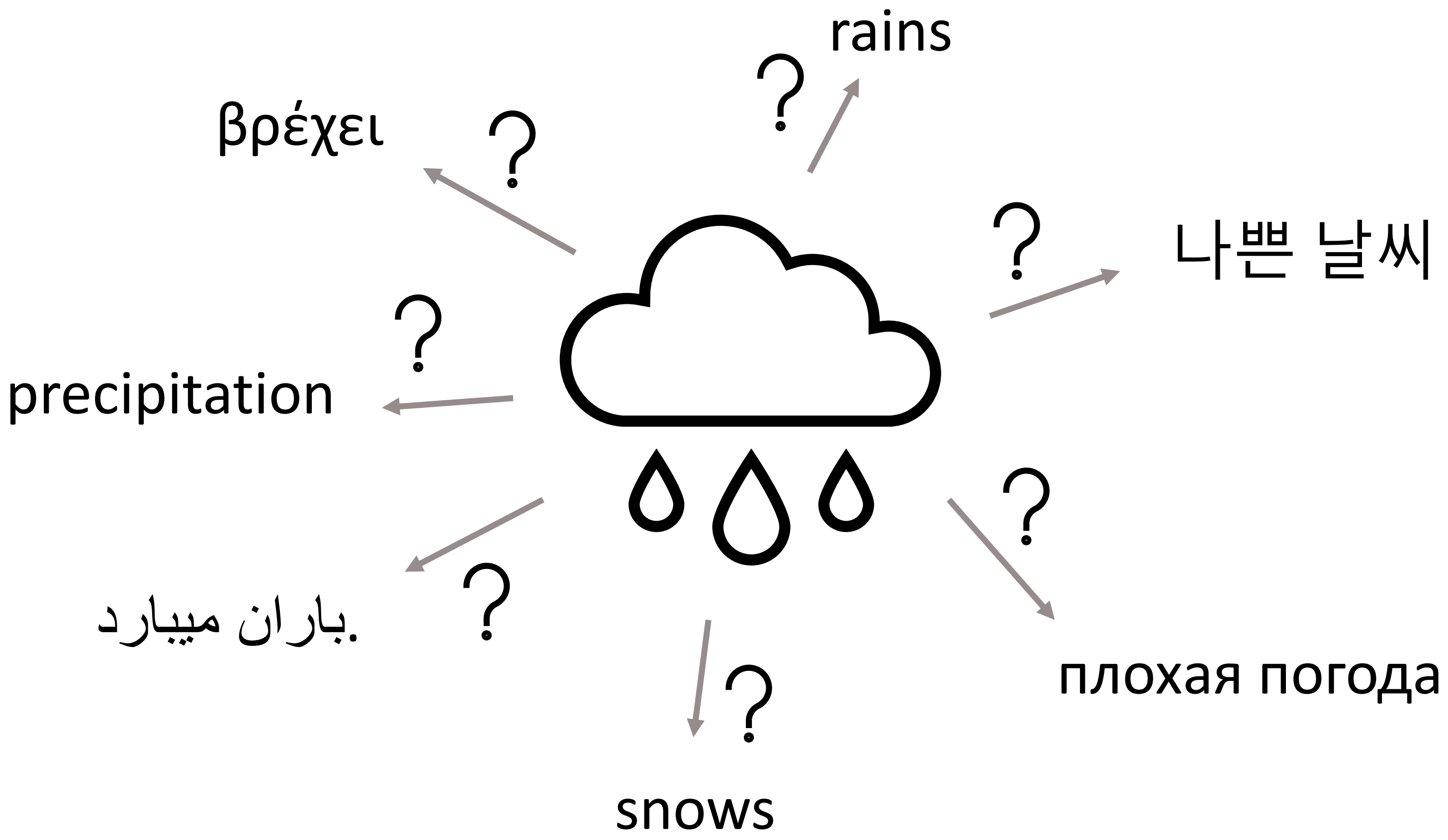
so: “St. Johns”$\in I$(“rains”)

so: “St. Johns”$\notin I$(“rains”)

so: “St. Johns”$\notin I$(“rains”)

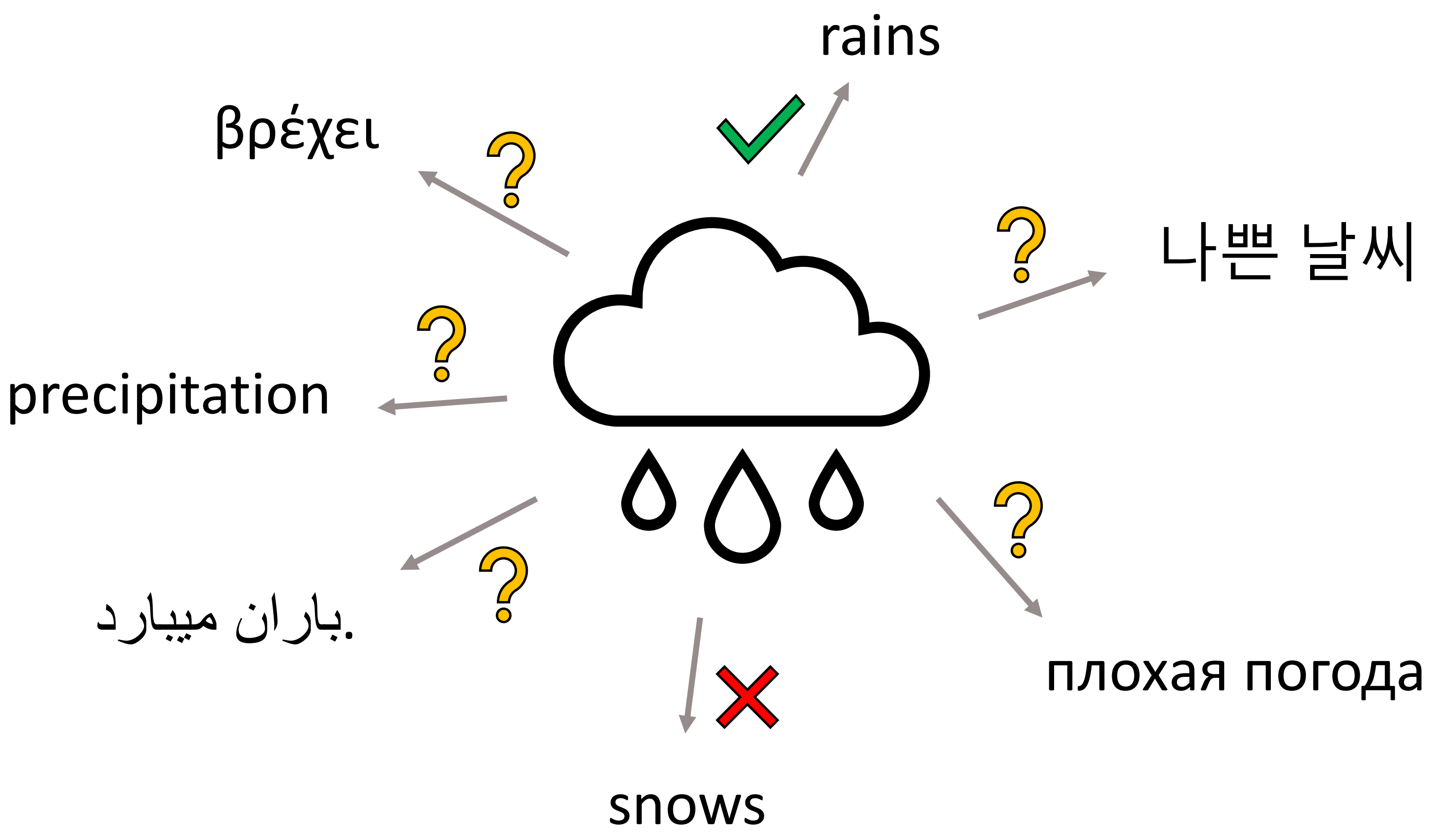
S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Research Question



S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Research Question



rains ✓

βρέχει ?

나쁜 날씨 ?

precipitation ?

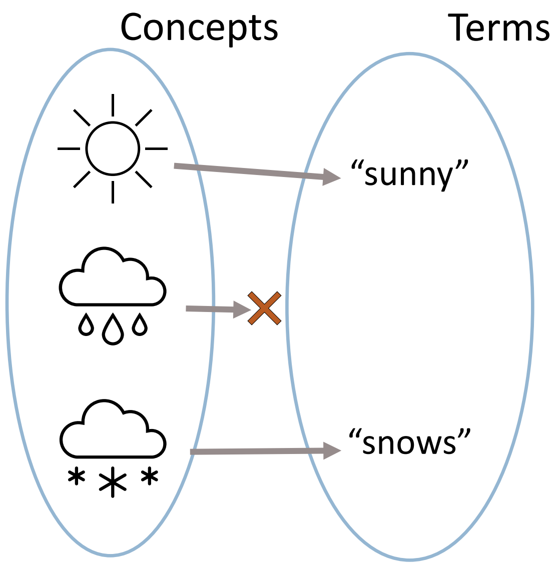
باران میبارد. ?

плохая погода ?

snows ✗

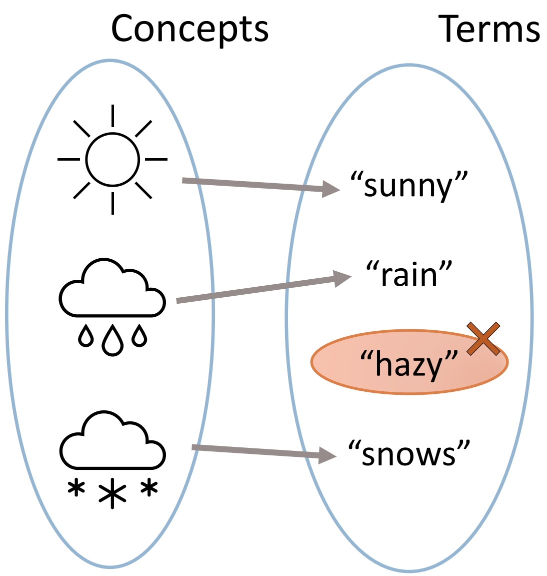
S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Characterising Vocabulary Quality[2]



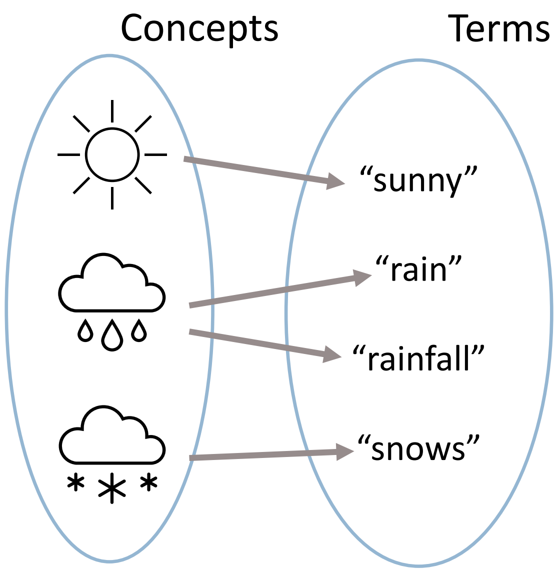
Construct Deficit

There are concepts without terms.



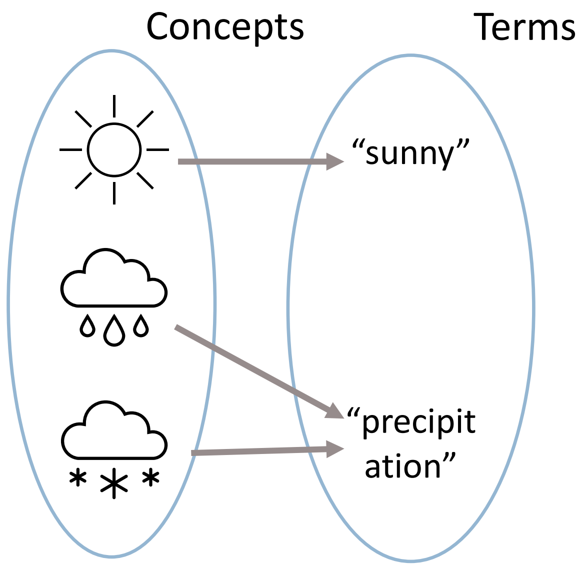
Construct Excess

There are terms not representing a concept of interest.



Construct Redundancy

There are concepts represented by more than one terms.



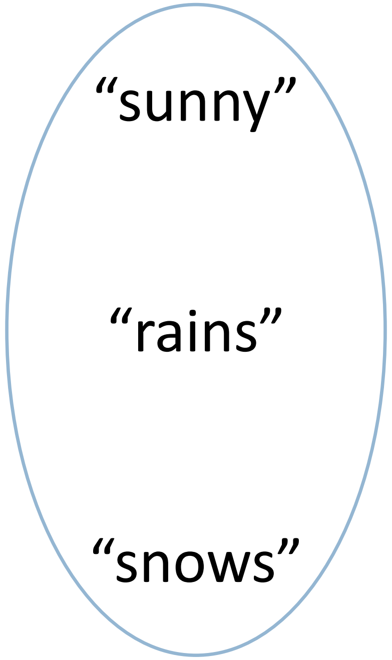
Construct Overload

There are terms representing more than one concept.

[2] Y. Wand, R. Weber. **On the ontological expressiveness of information systems analysis and design grammars**. Journal of Information Systems. 3(4). 1993.

Key Idea

Vocabulary $V_{\mathfrak{R}}$



“sunny”

“rains”

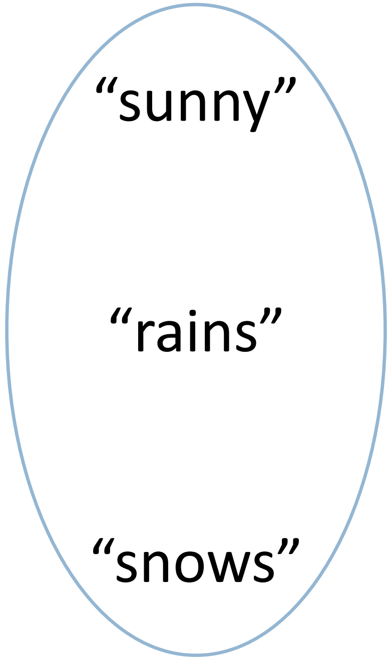
“snows”

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea

1. **Descriptions** $e \in E$ describing $w \in W$
2. **Elements** $d \in D$ worth modeling

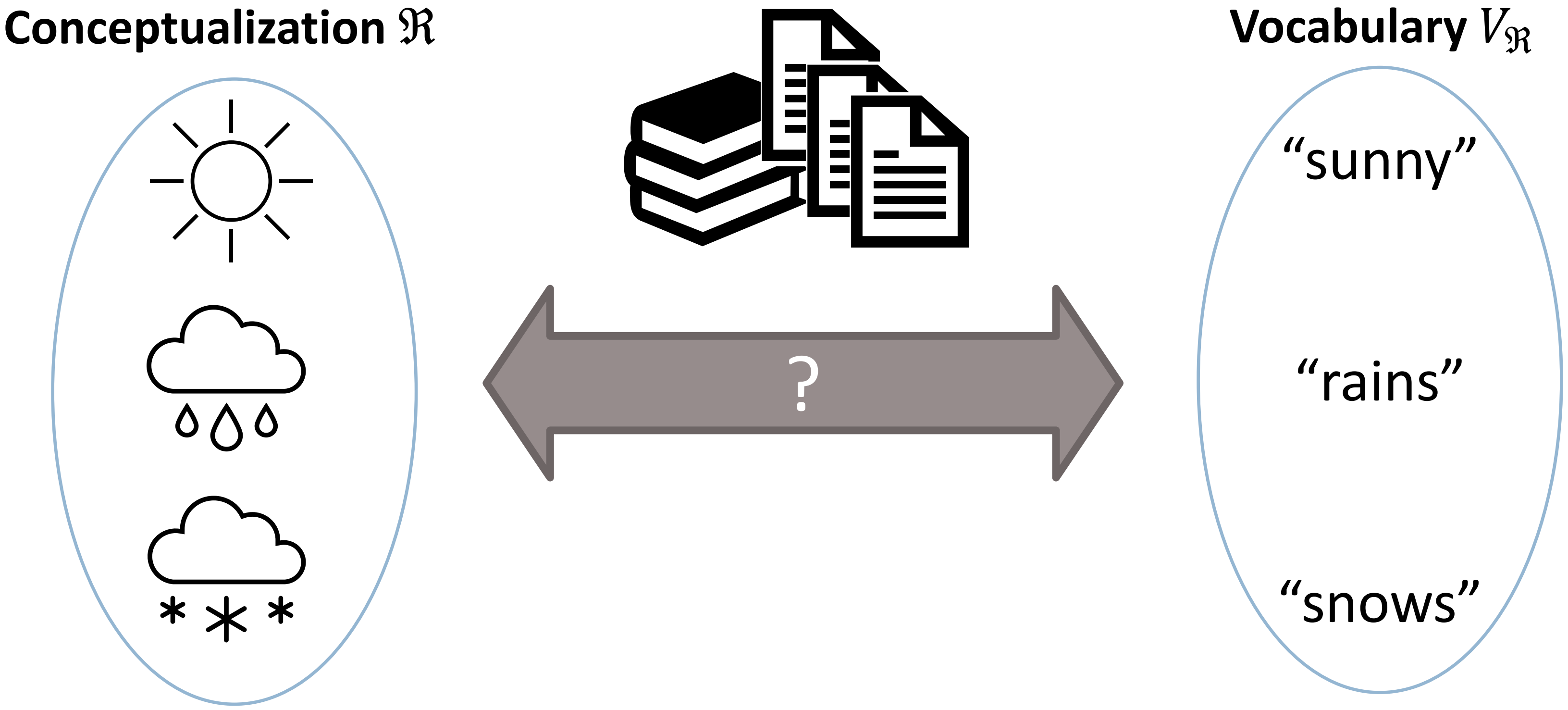
Vocabulary $V_{\mathfrak{R}}$



S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea

1. **Descriptions** $e \in E$ describing $w \in W$
2. **Elements** $d \in D$ worth modeling



Conceptualization $\mathfrak{R}$

?

Vocabulary $V_{\mathfrak{R}}$

“sunny”

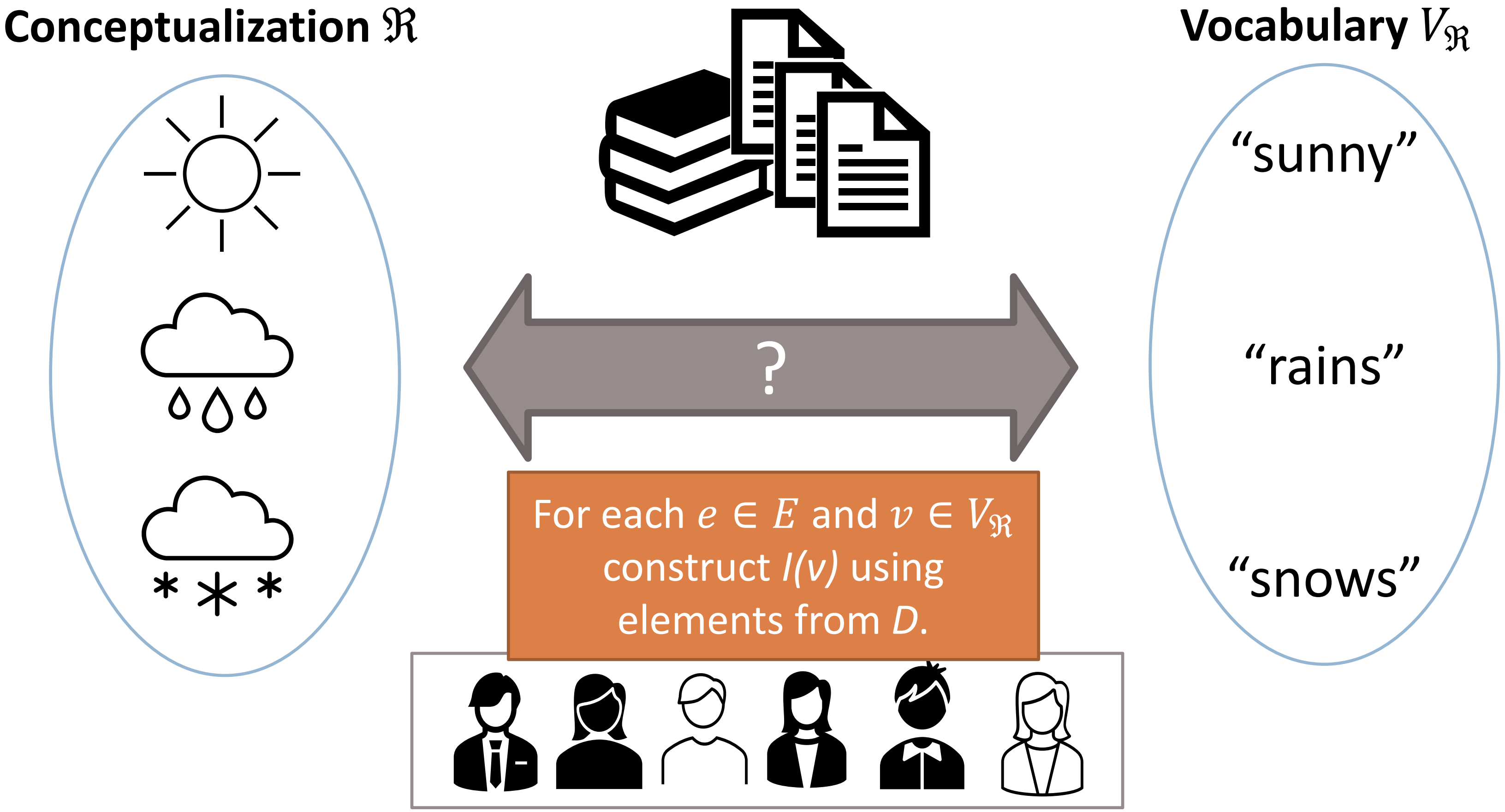
“rains”

“snows”

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea

1. **Descriptions** $e \in E$ describing $w \in W$
2. **Elements** $d \in D$ worth modeling



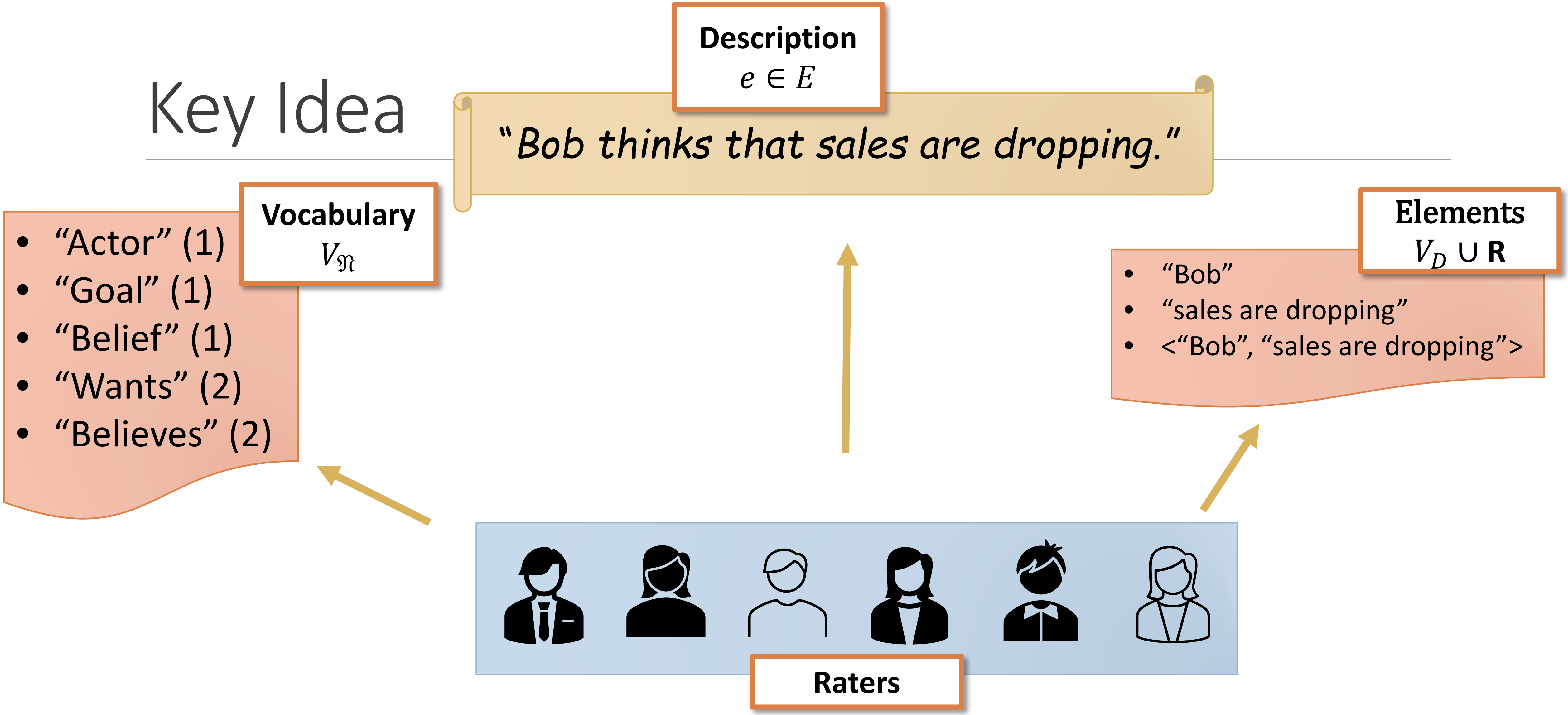
Conceptualization $\mathfrak{R}$

Vocabulary $V_{\mathfrak{R}}$

For each $e \in E$ and $v \in V_{\mathfrak{R}}$ construct *I(v)* using elements from *D*.

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea



Description
$e \in E$

"Bob thinks that sales are dropping."

Vocabulary
$V_{\mathfrak{N}}$

- "Actor" (1)
- "Goal" (1)
- "Belief" (1)
- "Wants" (2)
- "Believes" (2)

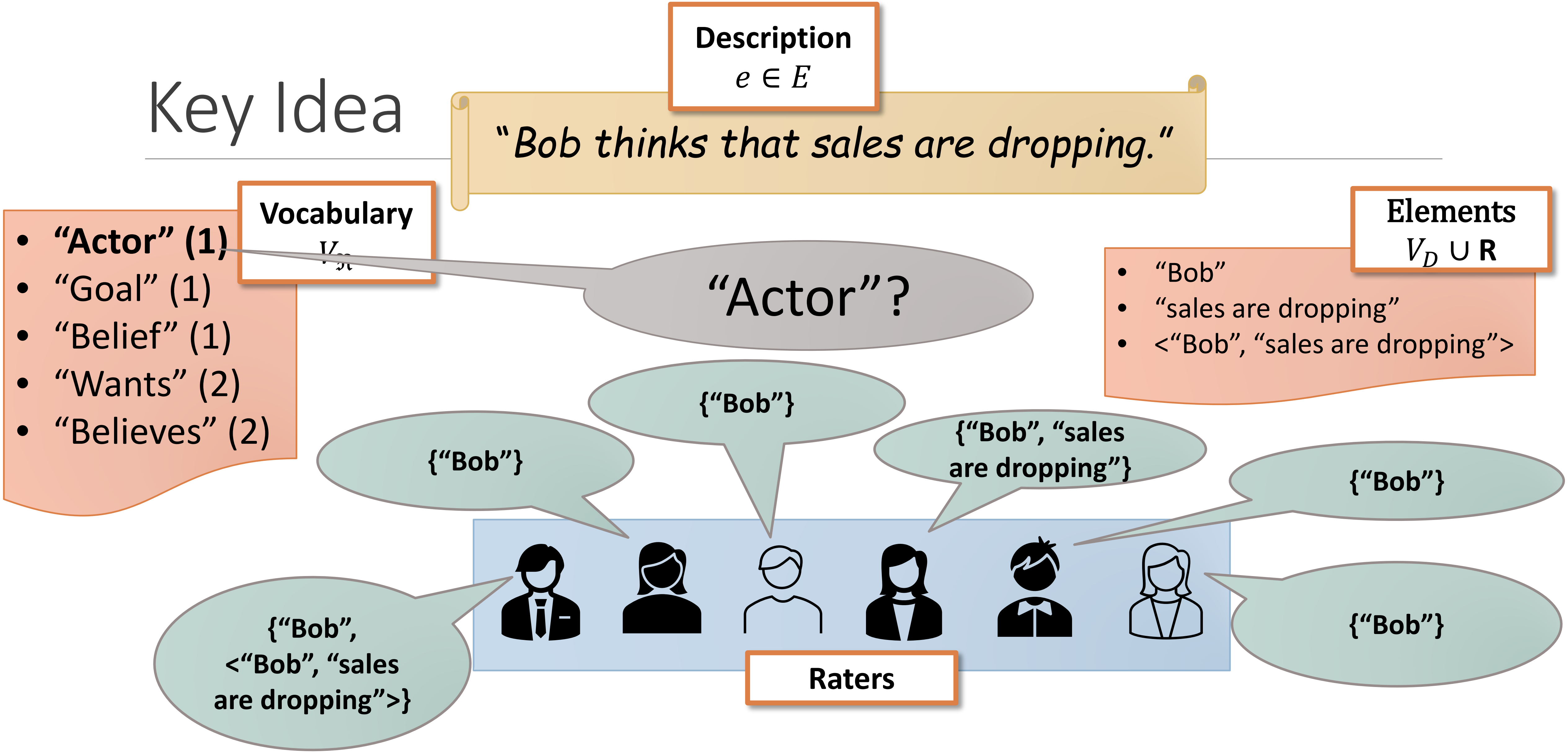
Elements
$V_D \cup \mathbf{R}$

- "Bob"
- "sales are dropping"
- <"Bob", "sales are dropping">

Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea



Description
$e \in E$

"Bob thinks that sales are dropping."

Vocabulary
$V_{\mathfrak{M}}$

- **"Actor" (1)**
- "Goal" (1)
- "Belief" (1)
- "Wants" (2)
- "Believes" (2)

"Actor"?

Elements
$V_D \cup \mathbf{R}$

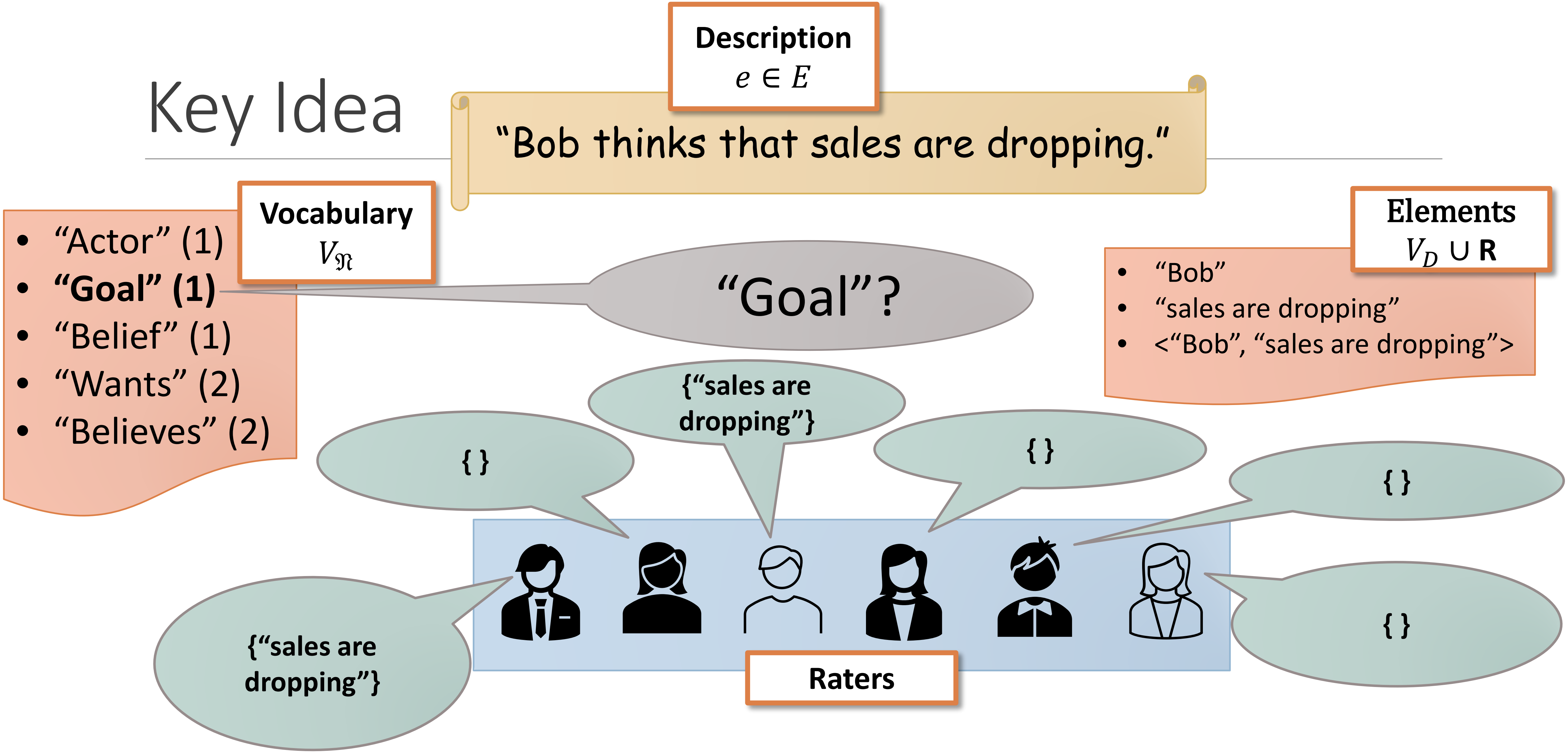
- "Bob"
- "sales are dropping"
- <"Bob", "sales are dropping">

{"Bob"}
{"Bob"}
{"Bob", "sales are dropping"}
{"Bob"}
{"Bob", <"Bob", "sales are dropping">}
{"Bob"}

Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea



Description
$e \in E$

"Bob thinks that sales are dropping."

Vocabulary
$V_{\mathfrak{N}}$

- "Actor" (1)
- **"Goal" (1)**
- "Belief" (1)
- "Wants" (2)
- "Believes" (2)

"Goal"?

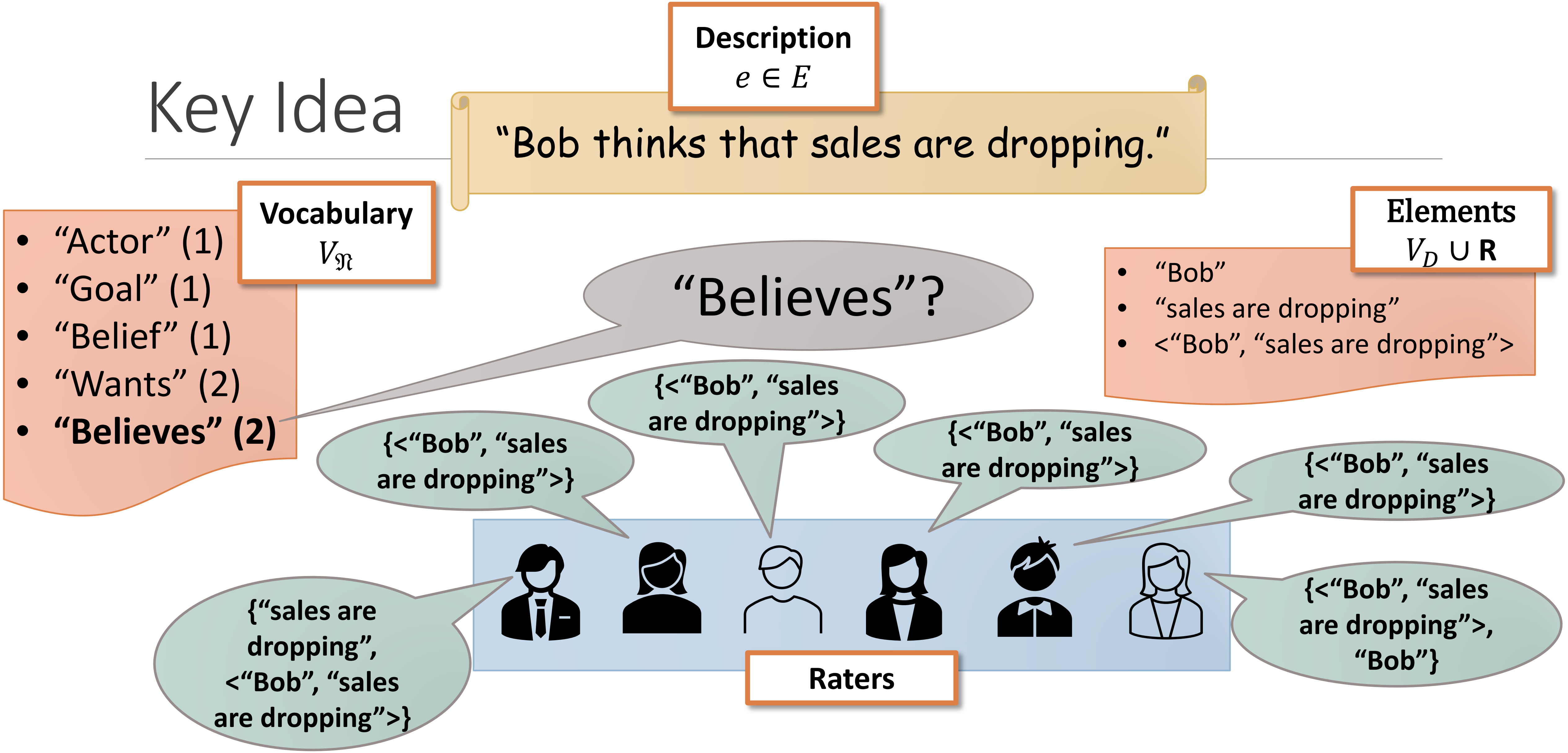
Elements
$V_D \cup \mathbf{R}$

- "Bob"
- "sales are dropping"
- <"Bob", "sales are dropping">

Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Key Idea



Description
$e \in E$

"Bob thinks that sales are dropping."

Vocabulary
$V_{\mathfrak{R}}$

- "Actor" (1)
- "Goal" (1)
- "Belief" (1)
- "Wants" (2)
- **"Believes" (2)**

"Believes"?

Elements
$V_D \cup \mathbf{R}$

- "Bob"
- "sales are dropping"
- <"Bob", "sales are dropping">

{"sales are dropping", <"Bob", "sales are dropping">}

{<"Bob", "sales are dropping">}

{<"Bob", "sales are dropping">}

{<"Bob", "sales are dropping">}

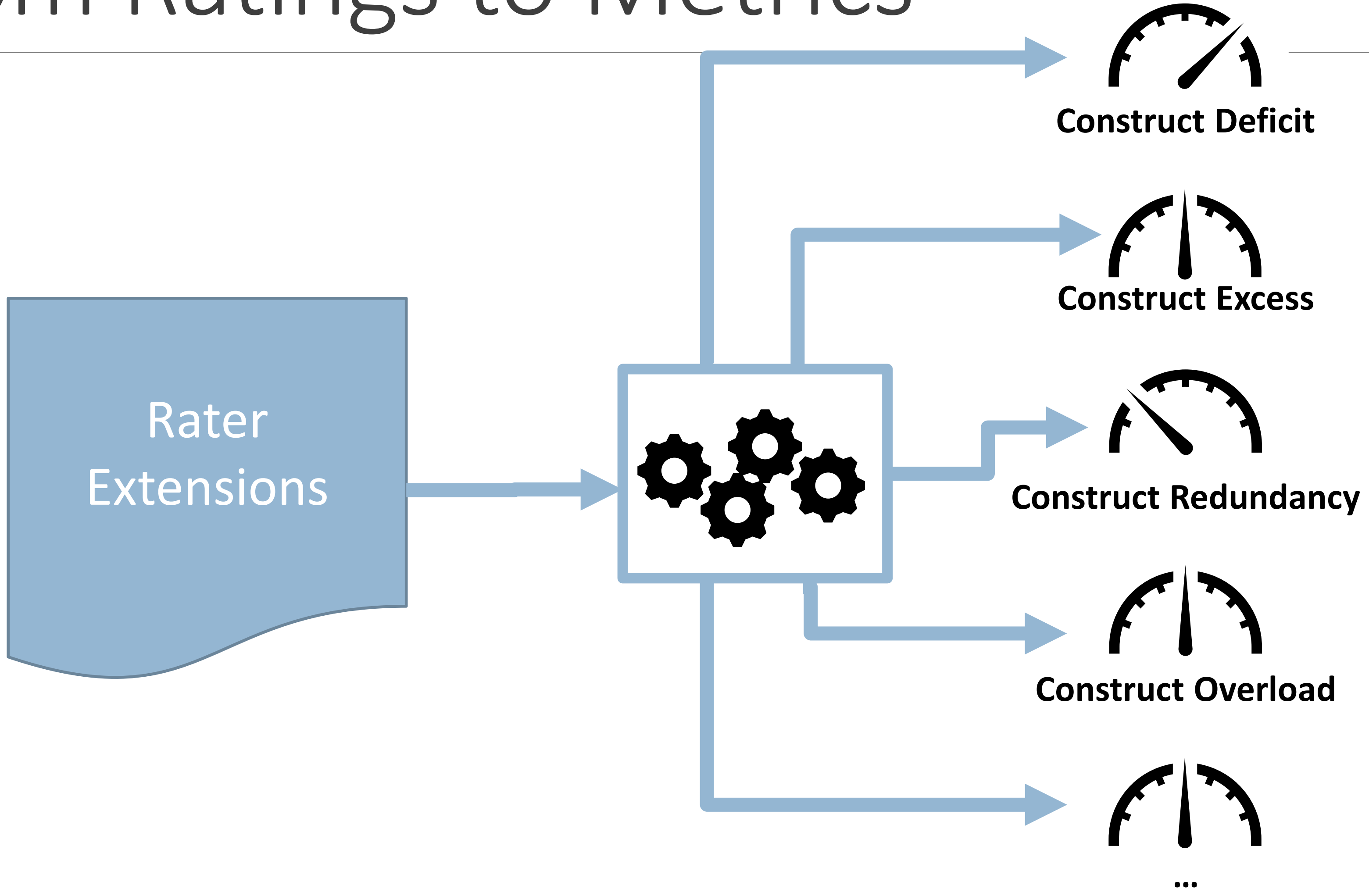
{<"Bob", "sales are dropping">}

{<"Bob", "sales are dropping">, "Bob"}

Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

From Ratings to Metrics



S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Construct Deficit

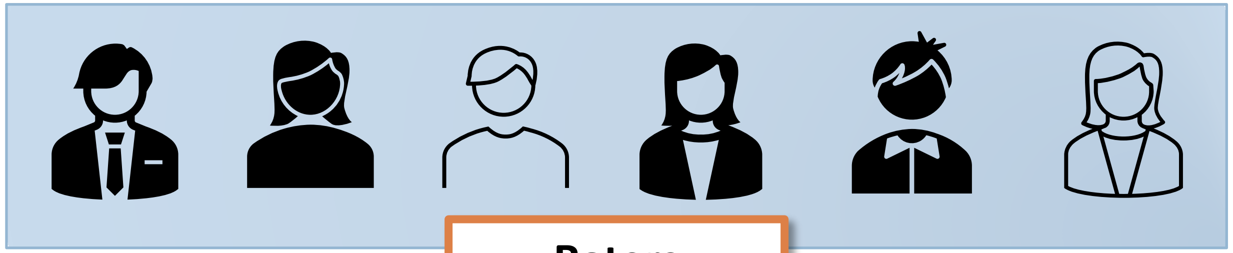
$V_{\mathfrak{N}}$

- “Goal” (1)
- “Belief” (1)
- “Wants” (2)
- “Believes” (2)

Elements (V_D ∪ R)	**Ratings #**
“Bob”	0/15
“sales are dropping”	8/15
<“Bob”, “sales are dropping”>	6/15

V_D ∪ **R**

- “Bob”
- “sales are dropping”
- <“Bob”, “sales are dropping”>



Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Construct Excess

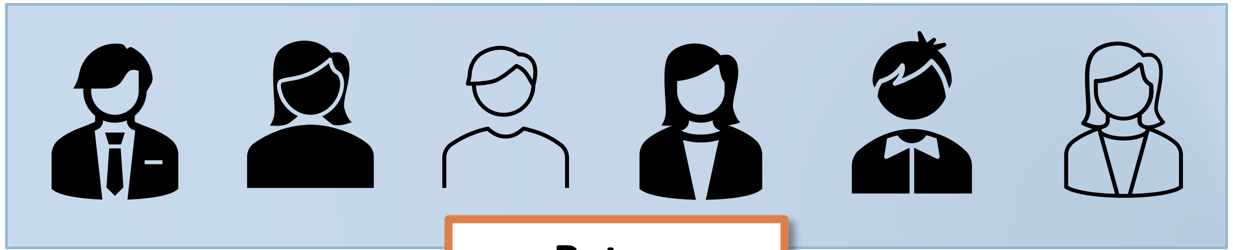
$V_{\mathfrak{N}}$

- “Goal” (1)
- “Belief” (1)
- “Wants” (2)
- “Believes” (2)

	# Times term was used					
Terms ($V_{\mathfrak{N}}$)	👤	👤	👤	👤	👤	👤
“Goal”	0	0	0	1	0	1
“Belief”	7	7	6	7	8	7
“Wants”	1	0	0	0	0	0
“Believes”	14	12	13	13	13	12

$V_D \cup \mathbf{R}$

- “Bob”
- “sales are dropping”
- <“Bob”, “sales are dropping”>



Raters

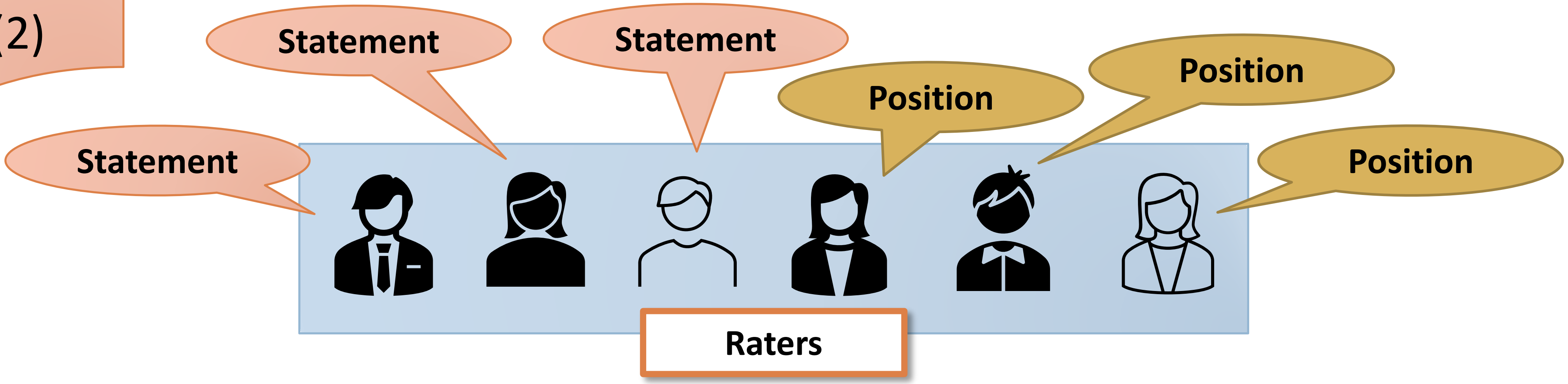
S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Overlap

$V_{\mathfrak{N}}$

- “Goal” (1)
- **“Statement” (1)**
- **“Position” (1)**
- “Wants” (2)
- “Believes” (2)

“sales are dropping”



Statement

Statement

Statement

Position

Position

Position

Raters

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Construct Redundancy

$V_{\mathfrak{N}}$

- “Goal” (1)
- **“Statement” (1)**
- **“Position” (1)**
- “Wants” (2)
- “Believes” (2)

	# Judges rated it as:	
Elements ($V_D \cup$ R)	**Statement**	**Position**
“sales are dropping”	3	3
“market is saturated”	2	4
“customers are happy”	4	2
“employees are dissatisfied”	3	3
“Bob”	0	1
“Alice”	0	0

Every time the term is substantially used, there is overlap with some other term.

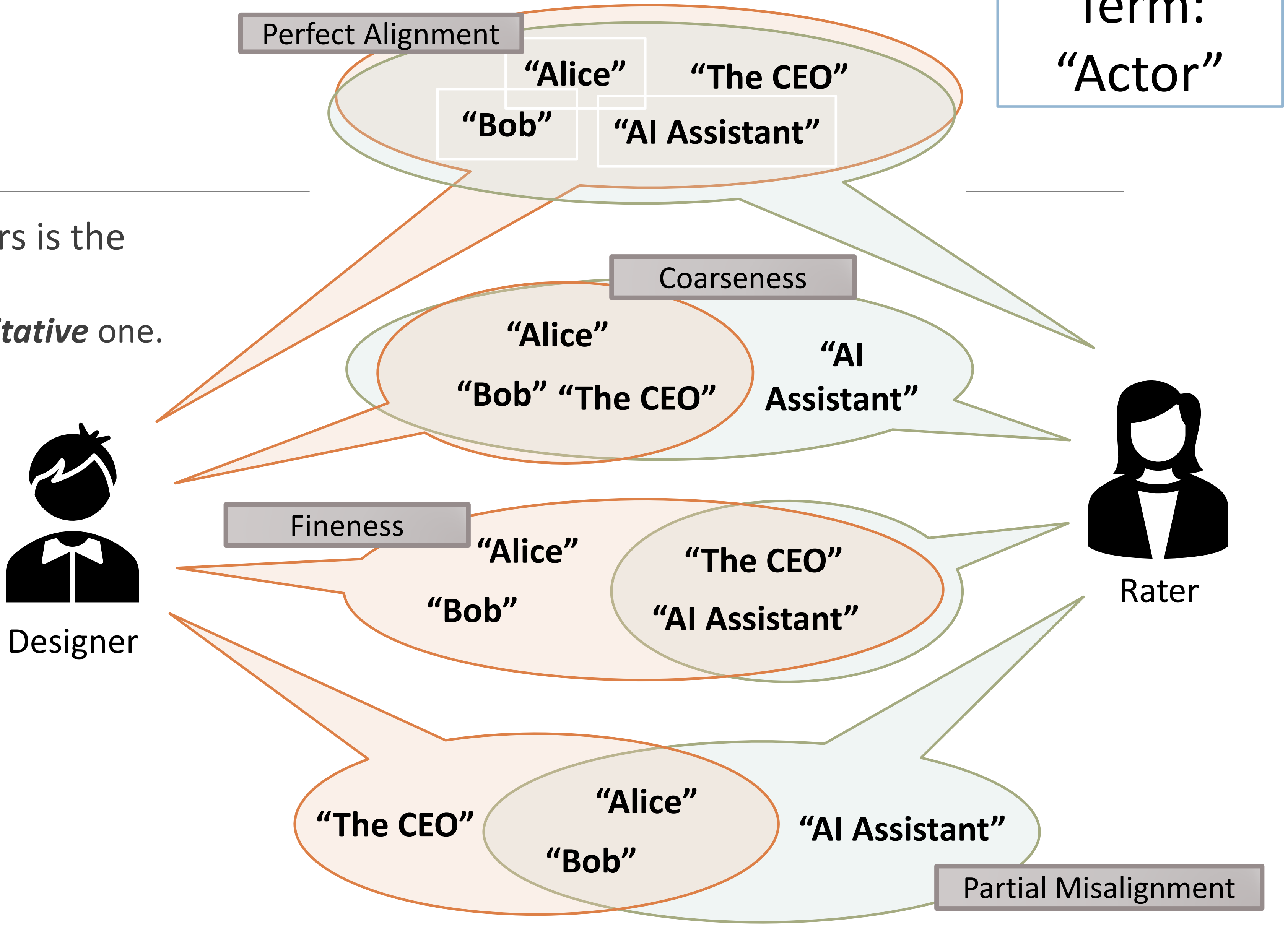
S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Measures of Accuracy

Term: "Actor"

Assume that one of two raters is the designer of the language.

- I.e. their rating is the ***authoritative*** one.



S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Application

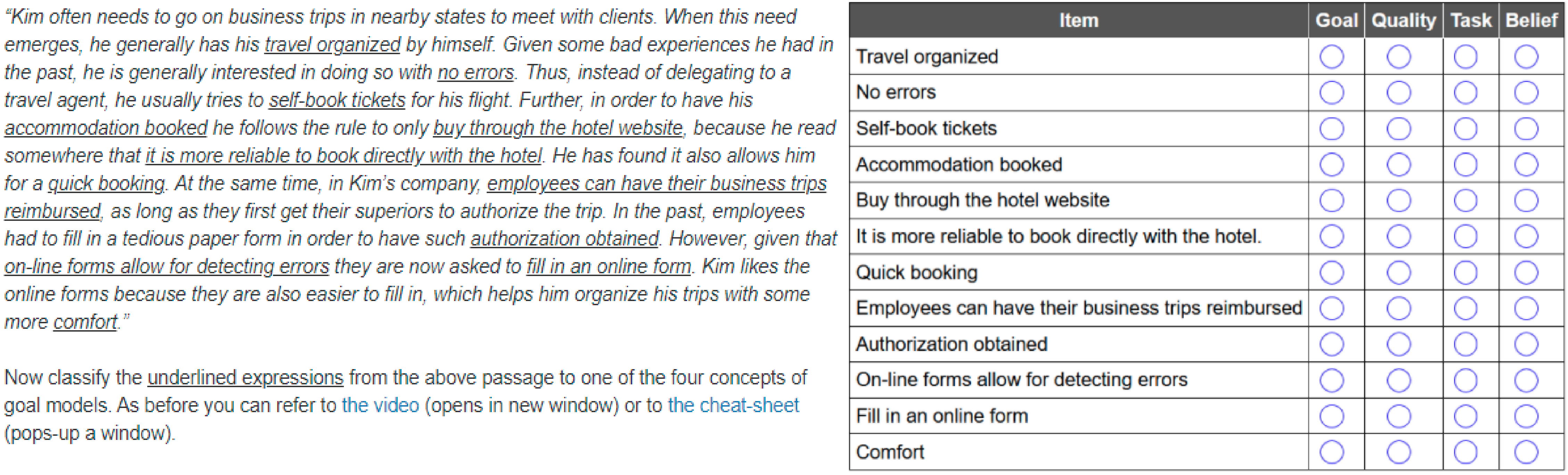
Data from previous study augmented/edited with simulated data.

Original study:

- Language: $V_{\mathfrak{N}_0}$ = {“goal”, “task”, “quality”, “belief”}
- Four different descriptions of 250 words each.
- Data collected from 20 Mechanical Turk participants trained to the language through videos.

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

Data Collection Instrument



"Kim often needs to go on business trips in nearby states to meet with clients. When this need emerges, he generally has his travel organized by himself. Given some bad experiences he had in the past, he is generally interested in doing so with no errors. Thus, instead of delegating to a travel agent, he usually tries to self-book tickets for his flight. Further, in order to have his accommodation booked he follows the rule to only buy through the hotel website, because he read somewhere that it is more reliable to book directly with the hotel. He has found it also allows him for a quick booking. At the same time, in Kim's company, employees can have their business trips reimbursed, as long as they first get their superiors to authorize the trip. In the past, employees had to fill in a tedious paper form in order to have such authorization obtained. However, given that on-line forms allow for detecting errors they are now asked to fill in an online form. Kim likes the online forms because they are also easier to fill in, which helps him organize his trips with some more comfort."

Now classify the underlined expressions from the above passage to one of the four concepts of goal models. As before you can refer to the video (opens in new window) or to the cheat-sheet (pops-up a window).

Item	Goal	Quality	Task	Belief
Travel organized	○	○	○	○
No errors	○	○	○	○
Self-book tickets	○	○	○	○
Accommodation booked	○	○	○	○
Buy through the hotel website	○	○	○	○
It is more reliable to book directly with the hotel.	○	○	○	○
Quick booking	○	○	○	○
Employees can have their business trips reimbursed	○	○	○	○
Authorization obtained	○	○	○	○
On-line forms allow for detecting errors	○	○	○	○
Fill in an online form	○	○	○	○
Comfort	○	○	○	○

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.

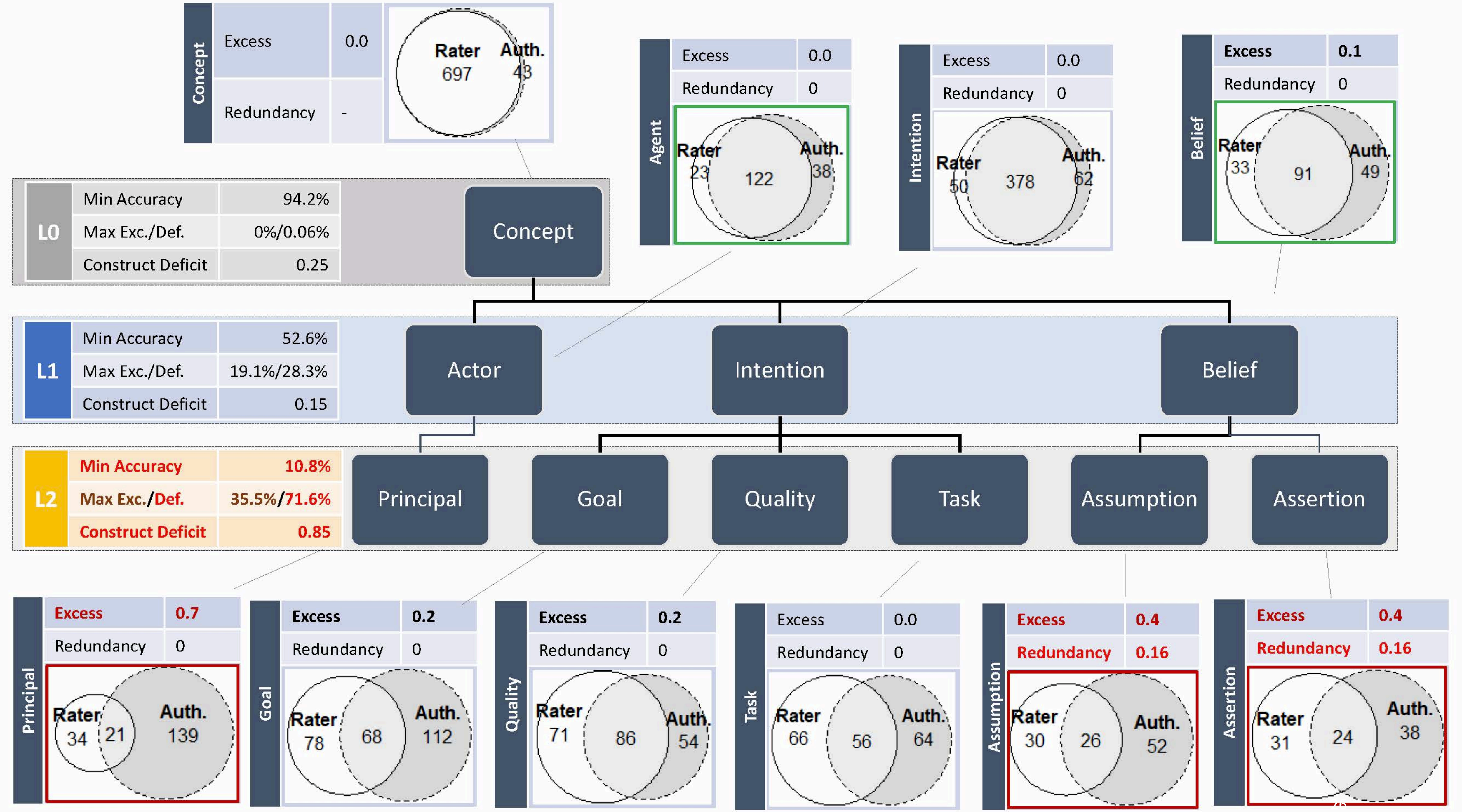
Application

Language tweaked to test detection of issues:

- $V_{\mathfrak{N}_1}$ = {“goal”, “task”, “quality”, “assumption”, “assertion”, “principal”}
  - Simulate overlap between “assumption” and “assertion” and difficulty to understand “principal” as a synonym for “actor”.
- $V_{\mathfrak{N}_2}$ = {“actor”, “intention”, “belief”}
  - “Principal” replaced by “actor”, “assumption” and “assertion” merged into “belief”, “goal”, “task”, and :quality” merged into intention.
- $V_{\mathfrak{N}_3}$ = {“concept”}

Precise operationalizations of the metrics were developed.

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.



Concept

Excess	0.0
Redundancy	-

Agent

Excess	0.0
Redundancy	0

Intention

Excess	0.0
Redundancy	0

Belief

Excess	**0.1**
Redundancy	0

L0	Min Accuracy	94.2%
	Max Exc./Def.	0%/0.06%
	Construct Deficit	0.25

L1	Min Accuracy	52.6%
	Max Exc./Def.	19.1%/28.3%
	Construct Deficit	0.15

L2	**Min Accuracy**	**10.8%**
	Max Exc./Def.	**35.5%/71.6%**
	Construct Deficit	**0.85**

Principal

Excess	**0.7**
Redundancy	0

Goal

Excess	**0.2**
Redundancy	0

Quality

Excess	**0.2**
Redundancy	0

Task

Excess	0.0
Redundancy	0

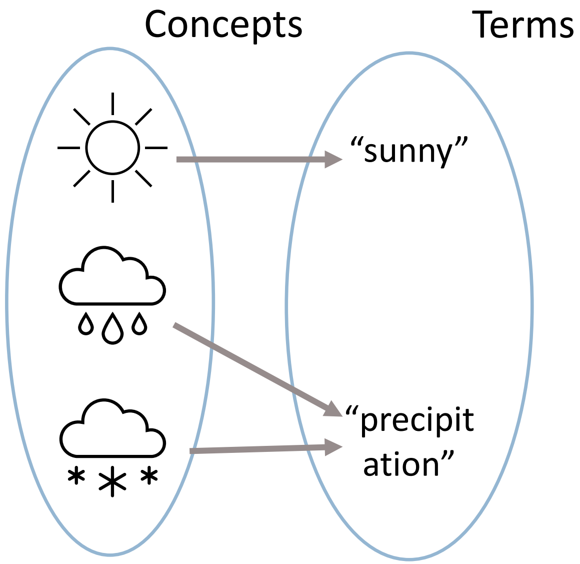
Assumption

Excess	**0.4**
Redundancy	**0.16**

Assertion

Excess	**0.4**
Redundancy	**0.16**

Construct Overload

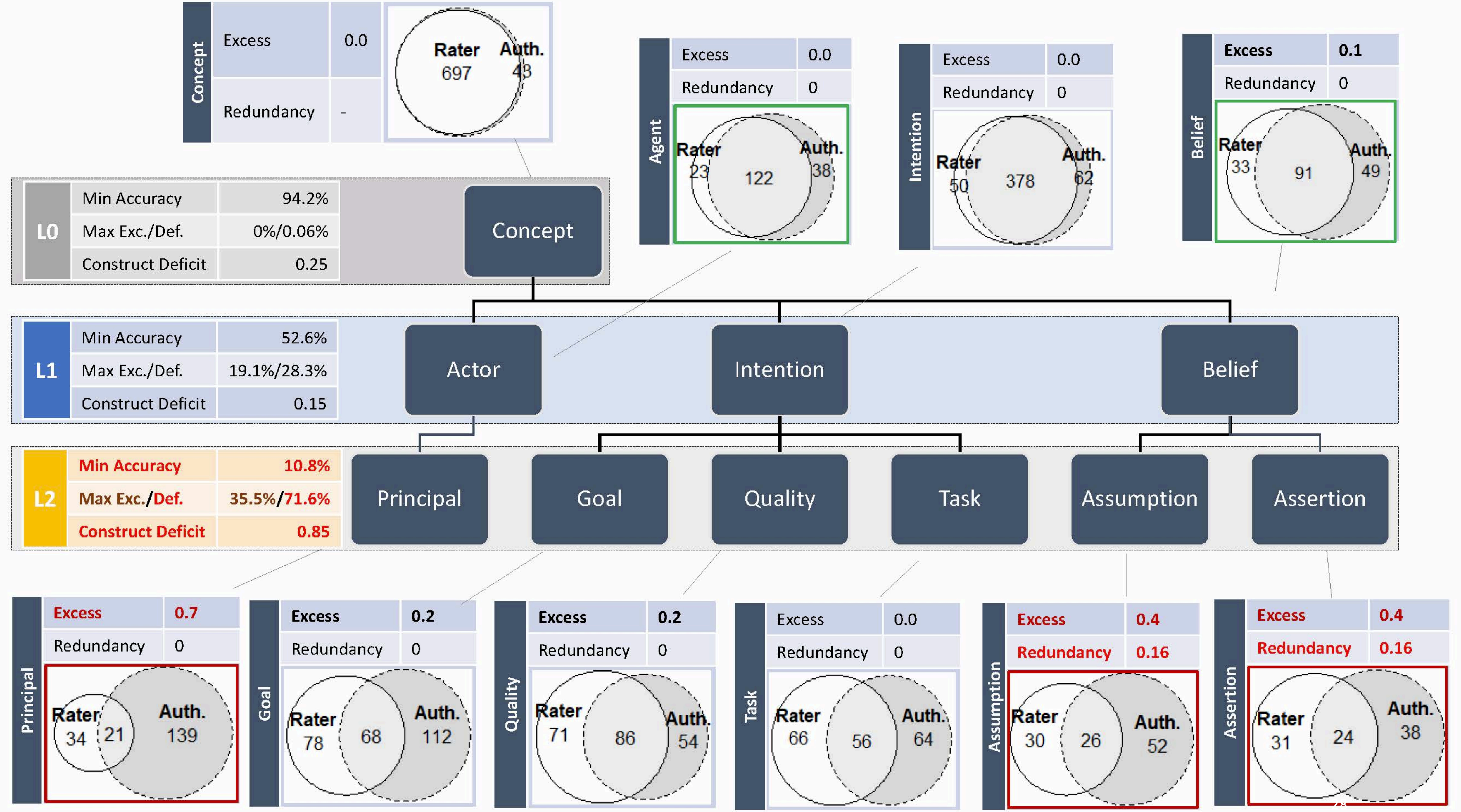


Construct Overload

There are terms
representing
more than one concept.

When refinement of the language is attempted and the result is a language that performs well in all other aspects, then we can hypothesize the presence of remediable construct overload in the original language.

S. Liaskos, S. M. Khan, J. Mylopoulos. *Empirically Evaluating the Semantic Qualities of Language Vocabularies.* ER 2021.



L0		
	Min Accuracy	94.2%
	Max Exc./Def.	0%/0.06%
	Construct Deficit	0.25

L1		
	Min Accuracy	52.6%
	Max Exc./Def.	19.1%/28.3%
	Construct Deficit	0.15

L2		
	Min Accuracy	10.8%
	Max Exc./Def.	35.5%/71.6%
	Construct Deficit	0.85

Concept

Excess	0.0
Redundancy	-

Agent

Excess	0.0
Redundancy	0

Intention

Excess	0.0
Redundancy	0

Belief

Excess	0.1
Redundancy	0

Principal

Excess	0.7
Redundancy	0

Goal

Excess	0.2
Redundancy	0

Quality

Excess	0.2
Redundancy	0

Task

Excess	0.0
Redundancy	0

Assumption

Excess	0.4
Redundancy	0.16

Assertion

Excess	0.4
Redundancy	0.16

Concept
Actor
Intention
Belief
Principal
Goal
Quality
Task
Assumption
Assertion
Concept: Rater 697 Auth. 43
Agent: Rater 23 122 Auth. 38
Intention: Rater 50 378 Auth. 62
Belief: Rater 33 91 Auth. 49
Principal: Rater 34 21 Auth. 139
Goal: Rater 78 68 Auth. 112
Quality: Rater 71 86 Auth. 54
Task: Rater 66 56 Auth. 64
Assumption: Rater 30 26 Auth. 52
Assertion: Rater 31 24 Auth. 38

Summary

A framework for empirically measuring vocabulary qualities

Based on examining how raters associate elements in the domain with concept-describing terms, under world descriptions.

Able to measure:

- Construct Deficit
- Construct Redundancy
- Construct Excess
- Accuracy, if authoritative data is available.
- Implicitly: Construct Overload

An application shows how to derive concrete operationalizations.

Thank you!

(questions?)