

## 1. Background & Goals

Question-Answer driven SRL (QA-SRL; He et. al., 2015) — a natural, easily attainable SRL flavor for verbs.

### Research Goals:

- Extend QA-SRL to **deverbal nominalizations**  
Analogous to NomBank complementing PropBank
- Collect high-quality annotations via crowdsourcing
- Measure extrinsic utility

## 2. Our Proposal – QANom

### Nominal Predicate criteria – “verbal noun” iff

- Lexically, the noun has a verbal counterpart (*proposal*–*propose*, *organization*–*organize*)
- In context, the noun denotes a verb-related event  $\equiv$  it is natural to ask verbal questions about it

“the **organization** of conferences and seminars...”  
*What is being organized?*

“Health care **organization**”  
*What is being organized?*

**Role questions** - leverage QA-SRL question format, centered by the corresponding verb

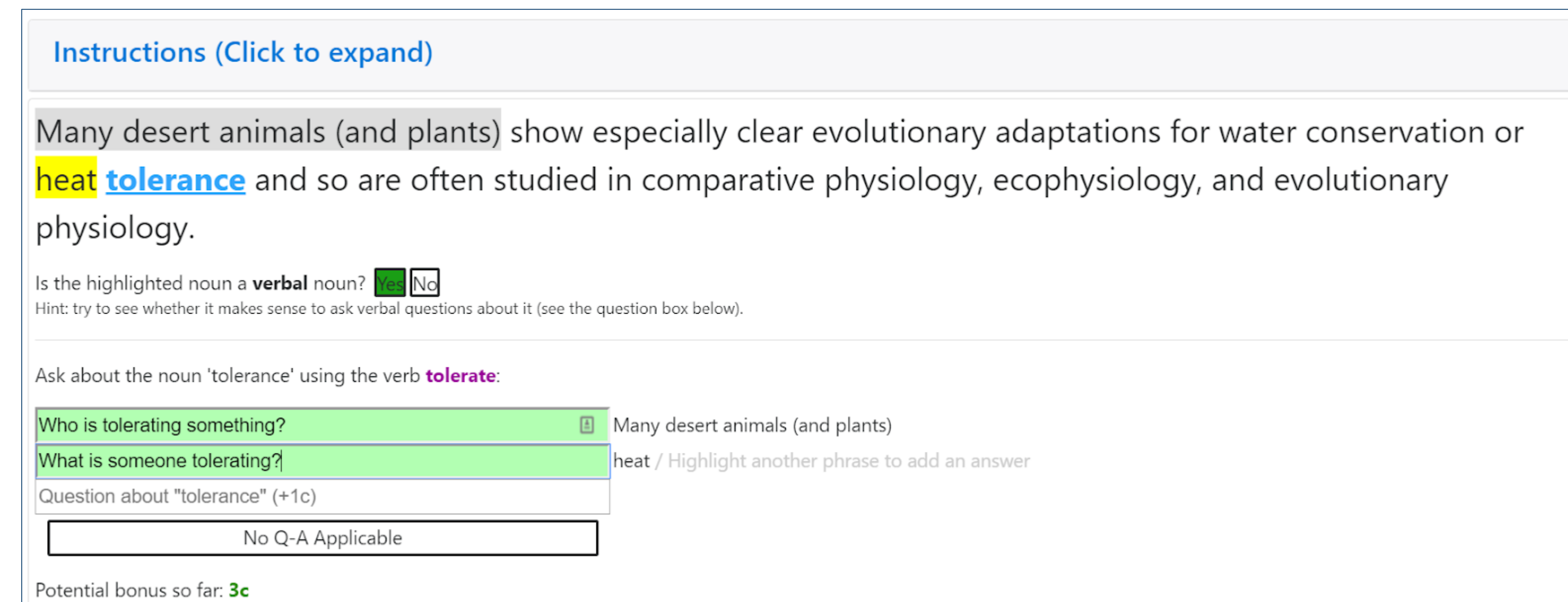
- ✓ Intuitive and natural representation
- ✓ Easily crowdsourcable
- ✓ Unified annotation scheme for verbs & nominal

Thomas has <b>proved</b> in different ways that God exists, including an assertion dubbed “the Ontological argument”.				
PropBank	ARGO	ARG1	Who has proved something?	Thomas
		ARGM-MNR	How did someone prove something?	in different ways   the Ontological argument
Thomas has provided different <b>proofs</b> for the existence of God, including an assertion dubbed “the Ontological argument”.				
NomBank	ARGO	ARG1	Who has proved something?	Thomas
		-	How did someone prove something?	the Ontological argument

## 3. Crowd Annotation Methodology

- Adapted from previous QA-SRL works (FitzGerald et. al., 2018; Roit et. al., 2020)
- Added a predicate detection question, for lexically-filtered candidates

### Interface



### Controlled Crowdsourcing

- Crowd workers are screened and trained with guidelines and feedbacks

- 2 *Generators* ask questions and highlight answer spans
- A *Consolidator* removes redundancies or erroneous questions (roles), and modifies argument spans

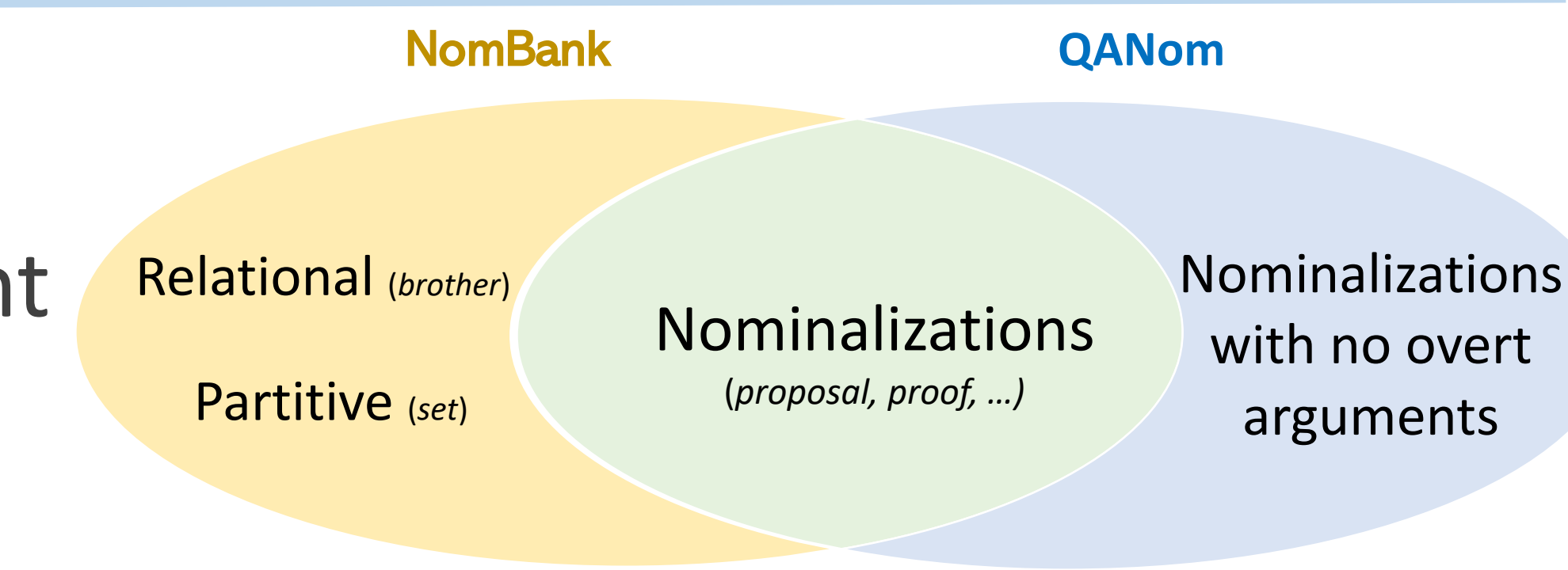
- Inter-Annotator Agreement:

	Generation	Consolidated
UA (F1)	67.2	77.1
Role (F1)	72.3	80.5
Is Verbal (Acc.)	81.8	85.6

## 4. Comparison to NomBank

### Predicate Detection:

NomBank targets a different scope of noun predicates.



Manually controlling for intended scope, QANom annotations reach **P=.97, R=.81** against NomBank

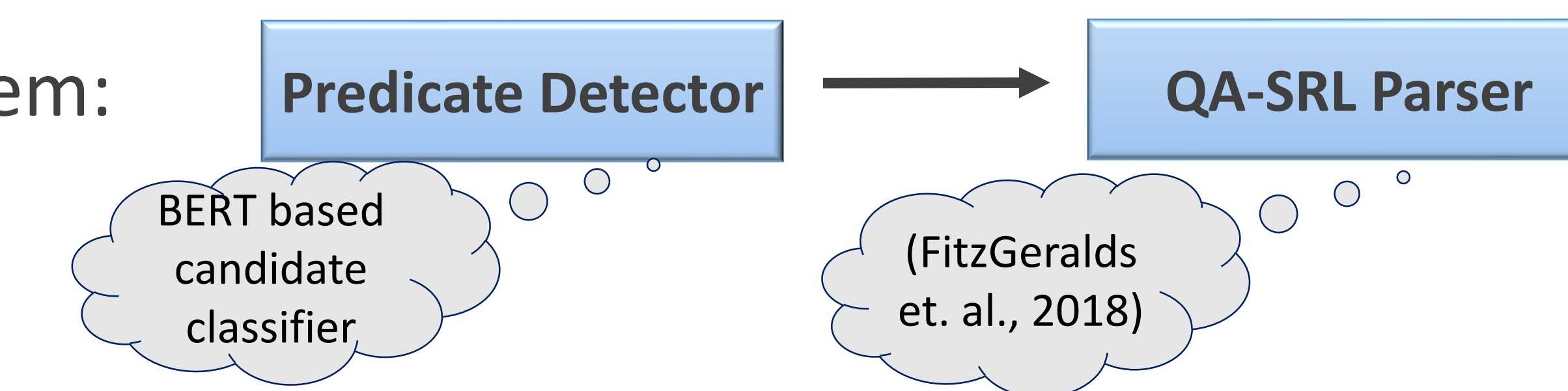
### Argument Detection:

Many formalism mismatches (**implicit arguments** or **syntactic markers**);

Manually controlled evaluation: **P=.86, R=.93** against NomBank

## 5. Baseline parser

Pipeline system:



	Gold		Predicted	
	UA	LA	UA	LA
P	45.1	29.6	47.2	31.6
R	61.5	40.4	49.7	33.3
F1	52.0	34.2	48.4	32.4

Performance of **QA-SRL parser** on QANom given either gold or predicted predicates.

## 6. Extrinsic Evaluations

**As pre-training** - *QuASE* (He et. al., 2020):

- Pre-training on QANom improves BERT for semantic tasks – SDP, PropBank, NomBank, Ontonotes, ACE

- Results are similar to pre-training on QA-SRL and QAMR

**For Zero-Shot**

**Event Extraction (ACE):**  
• Better than NomBank

Train Set	Predicate Detection			Argument Detection		
	P	R	F1	P	R	F1
NomBank	21.3	77.1	33.4	20.0	25.3	22.3
QANom	24.9	74.9	37.4	24.8	29.7	27.0
ACE	73.2	82.9	77.7	51.3	65.6	57.5