

# Closing the Semantic Gap

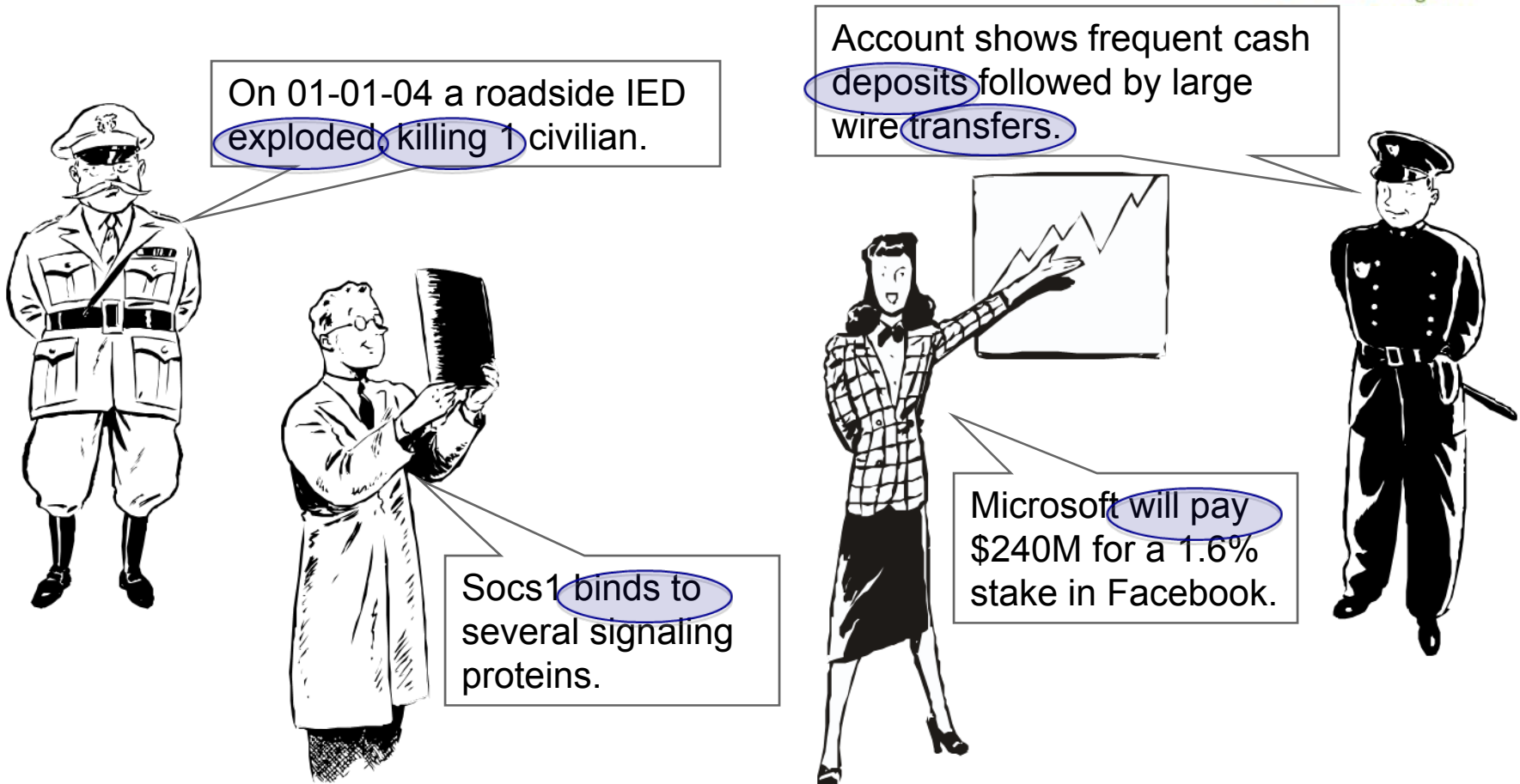
Marc Vilain

781-271-2151 • [mbv@mitre.org](mailto:mbv@mitre.org)

MITRE-sponsored Research (MSR)



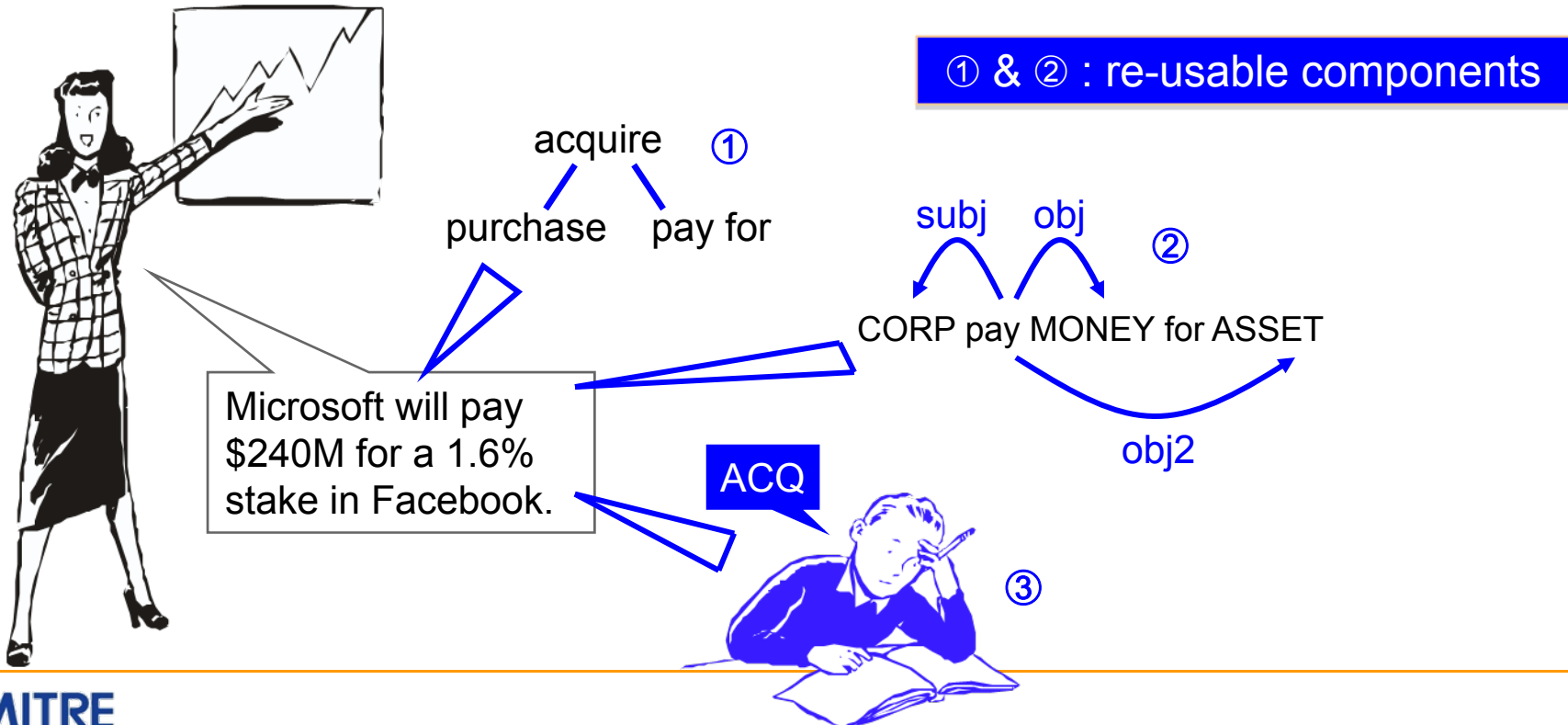
# Problem: finding events in text



- Analytic tasks often require modeling events in text.
- There are few opportunities for cross-task re-use of event models.
- The crux: creating effective event models for new tasks.

# Background

- Work to date identified elements of practical event extraction
  - ① Lexical hierarchies (generalization)
  - ② Grammatical relations parsing (normalization)
  - ③ “No-sweat” sentence-level event training (simplification).



# Continuation project objectives



- **Bring maturity to lexical hierarchies.**
- **Improve acuity of grammatical relations parser.**
- **Tighten integration of lexicon and parser.**
- **Continue validation experiments**
  - **Increase base of experience with methods**
  - **Apply methods to sponsor data**
  - **Measure differential value of methods.**
- **Publish.**

# Activities

- **Exploitation of machine readable dictionary**

- Infer lexical hierarchies
- Identify/normalize complex verb forms.



“set fire to X” =  
“set X on fire” =  
cause to burn

- **Improvements to grammatical relations parser**

- Integrate complex verb forms
- Surface indirect linguistic generalizations.



“X begins to Y”  
“X is believed to Y”  
“X promises to Y”  
→ X subj of Y

- ***Many* experiments with sponsor data.**

- **Fun and games**

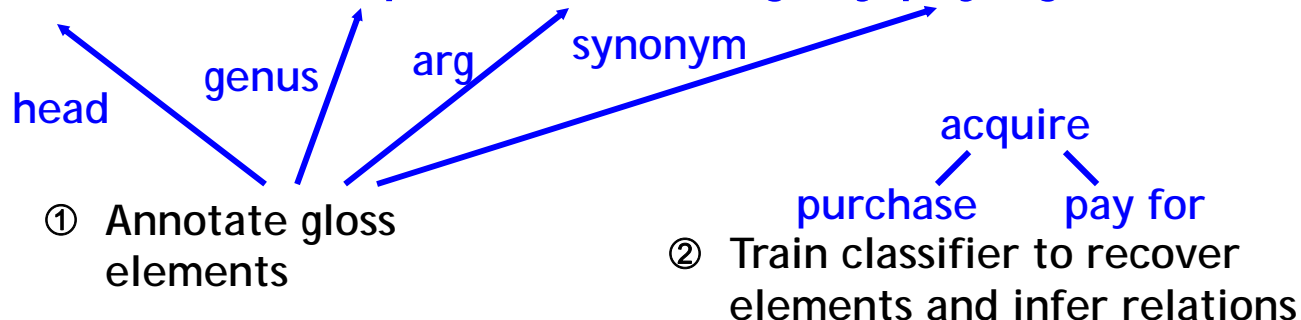
- Exploit additional dictionary markers (topic, formal/informal)
- Low-brow Web-harvested technical vocabulary
- Integration with visualization tools.

# Highlight: trainable gloss parsing



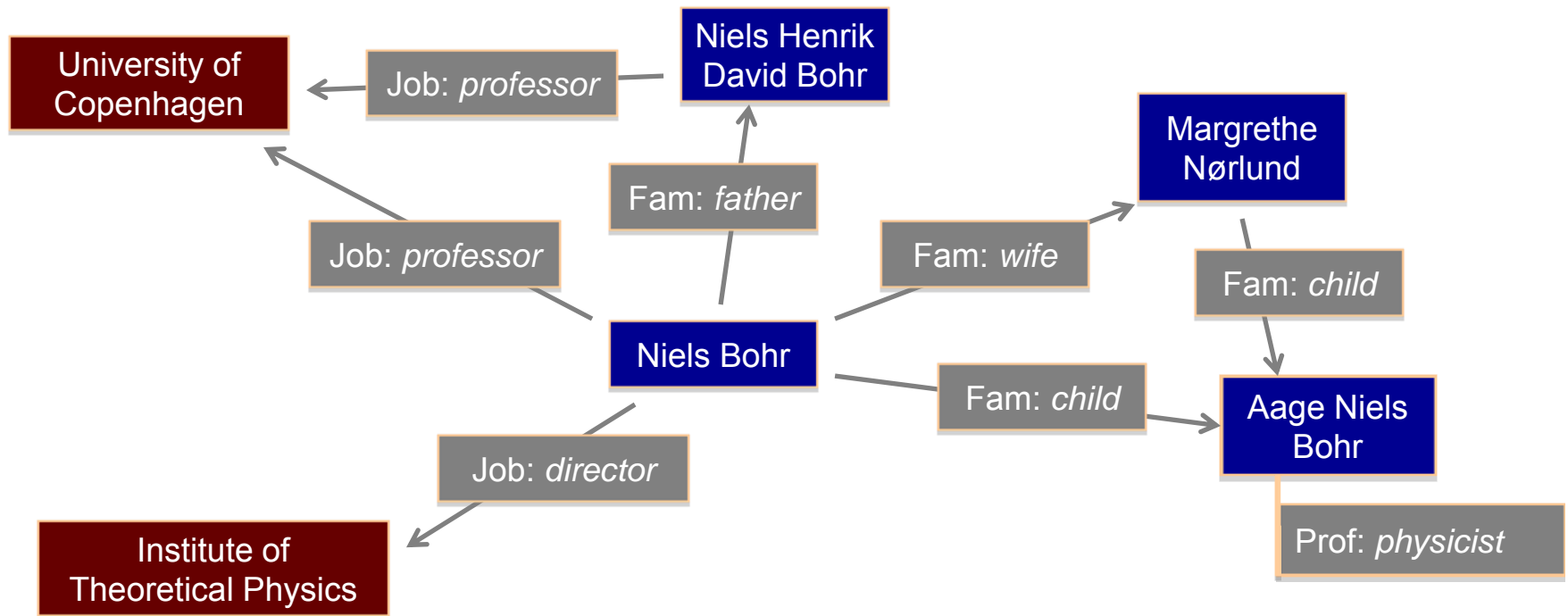
- Approach to date: extract genus of glosses
  - **Tractor**: a powerful motor **vehicle** with large rear wheels...
  - **Acquire**: **buy** or obtain.
- Complications
  - Disjunctions (“buy **or** obtain”)
  - Classifiers (“**a class of** predatory mollusks”).
- Use learning to improve identification.

Purchase (vb): acquire (something) by paying for it.



# Demo: Wikipedia biographies

Niels Henrik David Bohr was born in Copenhagen, Denmark in 1885.  
His father, Christian Bohr ... was a professor of physiology at the University of Copenhagen.



Niels Bohr and his wife Margrethe Nørlund had six children ....  
One, Aage Niels Bohr, also became a very successful physicist

In 1916, Niels Bohr became a professor at the University of Copenhagen ...  
founding the Institute of Theoretical Physics in 1921, of which he became its director.

# Impacts

- **Broad range of sponsor data experiments pave way for technology transitions**
  - Identify solved/solvable problems
  - Identify remaining hard problems
  - Provide burgeoning “best practices” experience.
- **Technology developments undergird future research and applications**
  - Parser, lexicon.
- **Sponsor-data prototypes enable light bulb moments and vision formation.**



# Future Plans



- **Continue technology transitions already in progress**
  - UAV chat logs
  - Law enforcement biographies
  - Intelligence reports
  - Financial transaction reports.
  
- **Support sponsor problems**
  - Intelligence analysis, law enforcement.
  
- **Explore research and application frontiers**
  - Empower end users to create event models
  - Explore common multilingual approaches
  - Explore intersection of lexicon and latent semantics.