Gene Regulation Event Corpus (GREC)

- 240 MEDLINE abstracts
- created to train IE systems
- annotators: 6 biology PhD students
  - annotated by hand using WordFreak
- identified 3067 events
  - verbs or nominalized verbs
- event arguments labeled with:
  - semantic roles
  - biological concept types

Labels

- Semantic roles:
  - AGENT, THEME, MANNER, INSTRUMENT, LOCATION, SOURCE, DESTINATION, RATE, TEMPORAL, CONDITION, PURPOSE, DESCRIPTIVE-AGENT, DESCRIPTIVE-THEME

- Biological concept hierarchies:
  - Nucleic_Acids, Proteins, Living_Systems, Processes, Experimental

Event example

In *E. coli*, glnAP2 may be **activated** by NifA.

Argument	Semantic role	Biological concept
In E. Coli	LOCATION	Wild_type_bacteria
glnAP2	THEME	Gene
NifA	AGENT	Activator

Inter-annotator agreement

- subtasks
  - event identification
  - argument identification (relaxed and exact)
  - semantic role assignment
  - biological concept identification
  - biological concept category assignment (exact category match, parent match, supercategory match)
- split *E. coli* and human species
- 66-95%