

wordrobe

Gamification for Word Sense Labeling

Game

- **Senses** is a **Game With a Purpose**
- Using **crowdsourcing** for linguistic annotation
- Players are asked to select the senses of **nouns** and **verbs**
- Game material is generated from the **Groningen Meaning Bank**



More electrical power capacity is needed for additional investment **projects** in titanium extraction and processing and garment manufacturing that could further close the import/export gap.

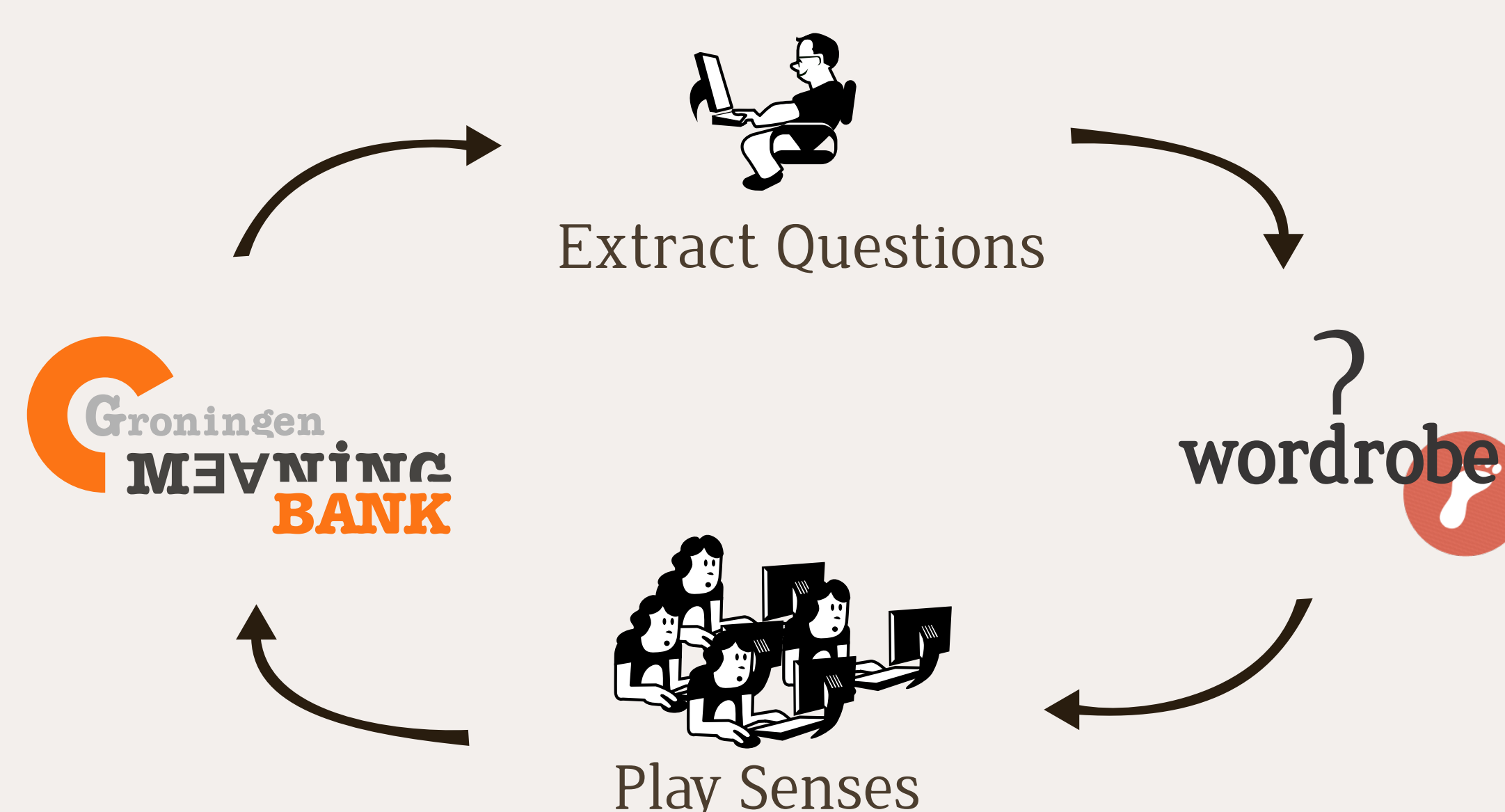
any piece of work that is undertaken or attempted (synonyms: undertaking, task, labor)
 a planned undertaking

Place your bet: low high

answer

Purpose

- **Goal:** to use answers from the game for Word Sense Labeling
Players assign senses from WordNet 3.1 to words
- **Challenge:** players are no experts
We cannot expect consistent, high quality answers
- **Solution:** check for high agreement on answers
Consider only answers upon which many players agree



Evaluation

- **Gold Standard:** 115 questions manually annotated by experts
- Several measures of **agreement**
- Only questions that received exactly 6 answers

Precision and recall based on different agreement measures

Strategy	Precision	Recall	F-score
Relative majority	0.880	0.834	0.857
Absolute majority ($t=0.5$)	0.882	0.782	0.829
Absolute majority ($t=0.7$)	0.945	0.608	0.740
Unanimity ($t=1$)	0.975	0.347	0.512
Chi-square test ($p < 0.05$)	0.923	0.521	0.666

When taking into account only **confident** answers, precision is **higher** and recall is **lower**

Precision and recall for questions with average bet $b \geq 80\%$

Strategy	Precision	Recall	F-score
Relative majority	0.917	0.478	0.629
Absolute majority ($t=0.5$)	0.930	0.461	0.616
Absolute majority ($t=0.7$)	0.956	0.383	0.547
Unanimity ($t=1$)	0.961	0.217	0.355
Chi-square test ($p < 0.05$)	0.950	0.330	0.355

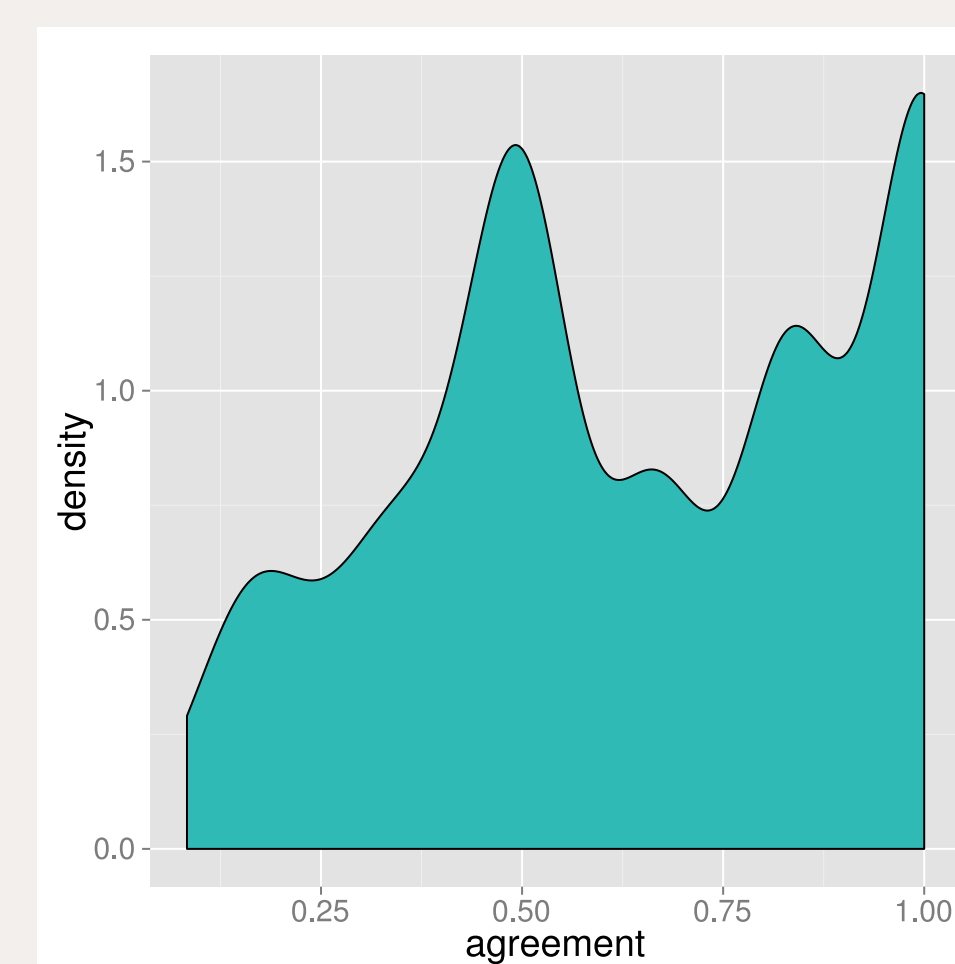
Conclusion

Results (update 12/3/2013)

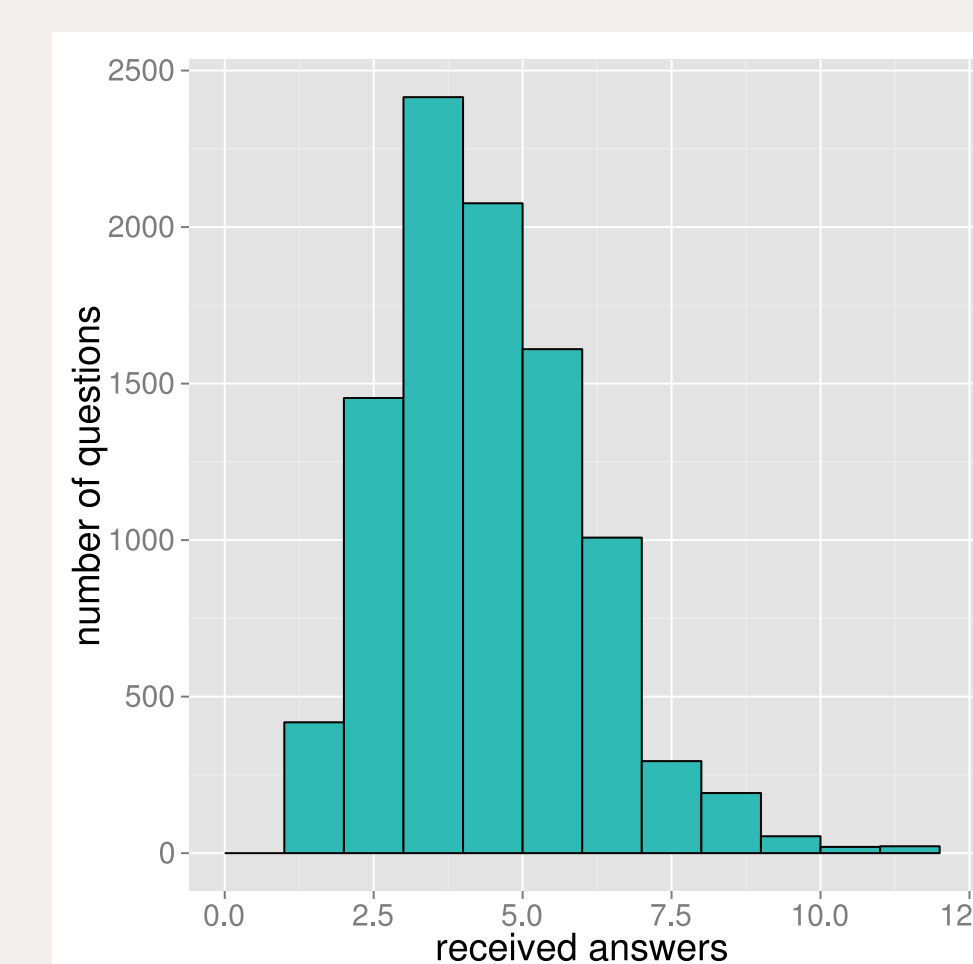
- 9,563 single answers
- 3,035 questions
- 508 players

Future work

- Players can bet (measure of confidence)
- Different strategies for agreement assessment
- Coarse vs fine-grained sense inventory



Density of agreement distribution on questions with 6 answers



Distribution of number of received answers per question

