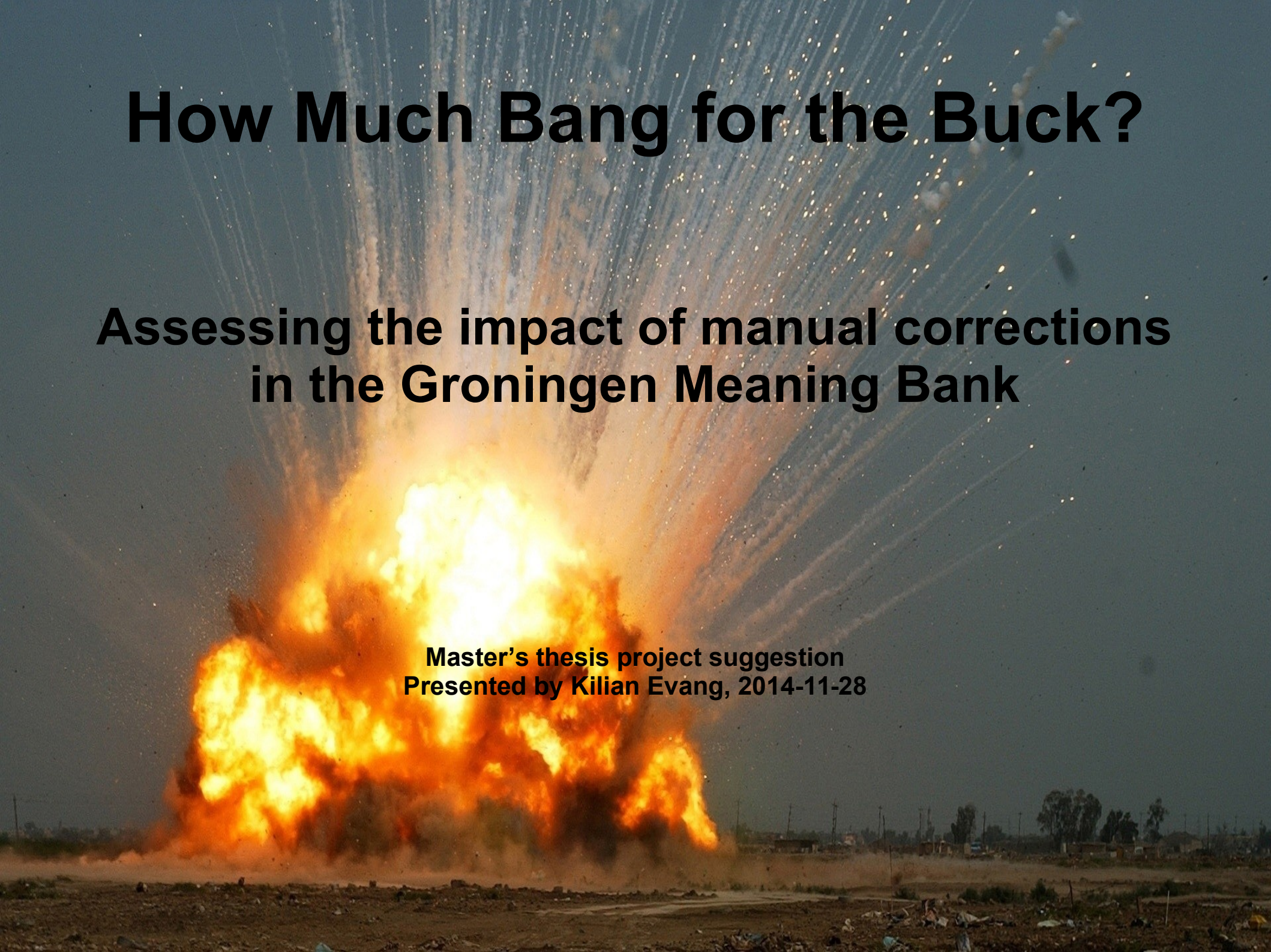


How Much Bang for the Buck?

Assessing the impact of manual corrections
in the Groningen Meaning Bank

Master's thesis project suggestion
Presented by Kilian Evang, 2014-11-28





Groningen

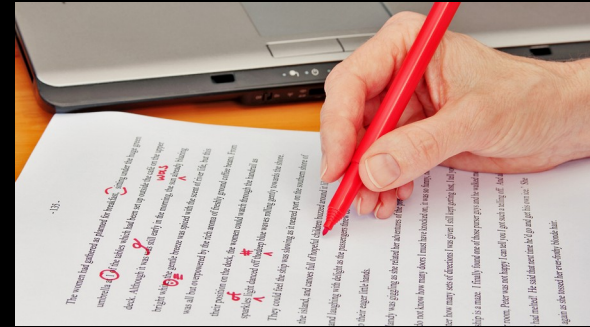
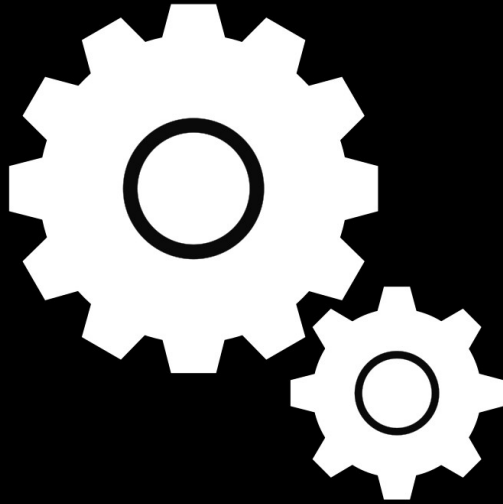
MEVNING

BANK





Groningen MEANING BANK



Active corrections

Type	#
Tokenization	3,047
Part of speech	22,475
Named entity	92,382
Animacy	1,412
Relation (e.g. in compound nouns)	1,151
Co-reference	6,499
Word sense	2,467

Research questions

- How can we effectively retrain models to generalize from corrections?
- How many mistakes does one correction fix?
- What factors does it depend on (type of mistake, source of correction, ...)?



Further reading

- Valerio Basile, Johan Bos, Kilian Evang, Noortje Venhuizen (2012): **Developing a large semantically annotated corpus**. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), pp 3196–3200, Istanbul, Turkey.
- Noortje Venhuizen, Valerio Basile, Kilian Evang, Johan Bos (2013): **Gamification for Word Sense Labeling**. Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013), pp 397–403, Potsdam, Germany