

## University of Groningen

### A Bigger Fish to Fry

Haagsma, Hessel

DOI:  
[10.33612/diss.131057087](https://doi.org/10.33612/diss.131057087)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Haagsma, H. (2020). *A Bigger Fish to Fry: Scaling up the Automatic Understanding of Idiomatic Expressions*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.  
<https://doi.org/10.33612/diss.131057087>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## Stellingen

behorende bij het proefschrift

# A Bigger Fish to Fry

## Scaling up the Automatic Understanding of Idiomatic Expressions

van

Hessel Haagsma

---

1. Individual idioms are rare, which makes automatically processing them difficult. Idioms, as a group, are frequent, which makes being able to process them a crucial NLP problem.
2. The lack of sizeable and varied idiom datasets has hampered previous research in many ways. A large, comprehensive dataset is an important step towards solving this.
3. The automatic extraction of potentially idiomatic expressions, an important prerequisite for large-scale corpus building, can be solved using a rule-based approach based on existing NLP tools.
4. Utilising crowdsourcing for difficult tasks like annotating the meaning of potentially idiomatic expressions can be successful, but the devil is in the details: breaking down the task into smaller bits and careful instruction are crucial.
5. Unsupervised disambiguation approaches have intuitive appeal and show promising results, but never actually **good** results.
6. Deep learning models for the disambiguation of potentially idiomatic expressions show a generalisation capability other supervised models have failed to achieve.
7. If natural language had been designed by a logician, idioms would not exist. — Philip Johnson-Laird, 1993
8. Lakoff reports surprising unanimity in informants concerning the prototypical mental image invoked by *spilling the beans*. For example, most say that the beans are uncooked and in a container about the size of the human head; that they are supposed to be in the container; that the spilling is accidental; that the beans go all over the place and are never easy to retrieve; and that the spill is messy. — Rosamund Moon, 1998
9. The use of *spill the beans* and *kick the bucket* as example idioms in scientific publications should be outlawed.
10. **beating a dead horse** (*British, idiomatic*): repeatedly rewriting and resubmitting a paper regardless of any hopes of getting it accepted.