

# Reducing the Complexity of Semantic Data Translation

Owen Gilson  
University of Wales Swansea, UK

[www.cs.swan.ac.uk/~csowen](http://www.cs.swan.ac.uk/~csowen)

7 November 2005



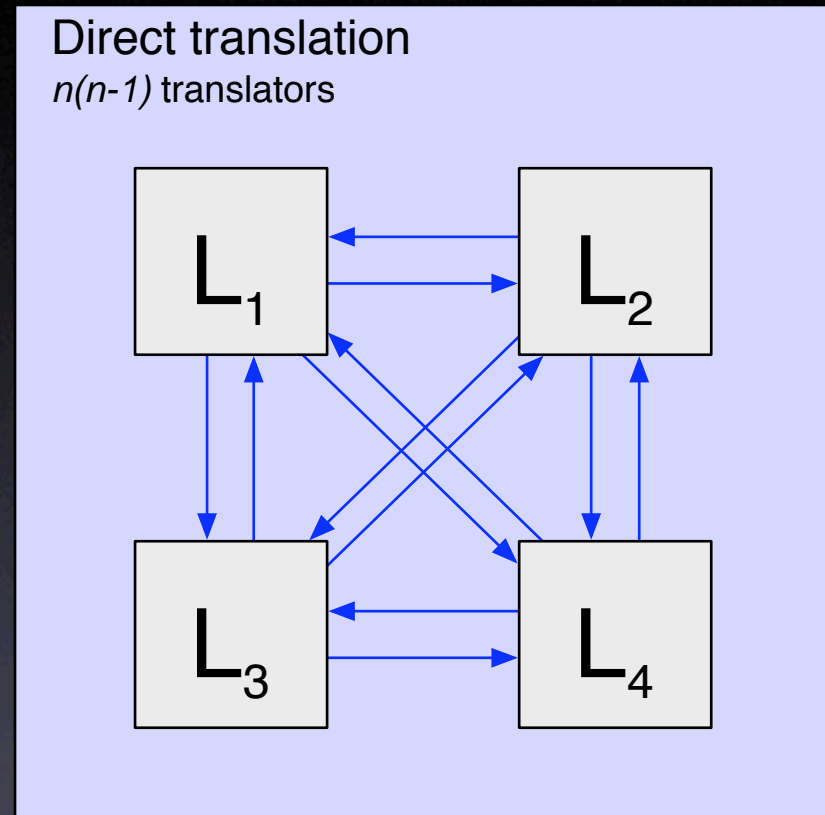
PRIFYSGOL CYMRU ABERTAWE  
UNIVERSITY OF WALES SWANSEA

# Semantically Rich Data

- Variety of formats emerging
- Many different domain areas
- What happens when we want to translate between formats?
  1. Direct translation
  2. Intermediate translation

# Direct translation

- Pros
  - Accurate
- Cons
  - Expensive



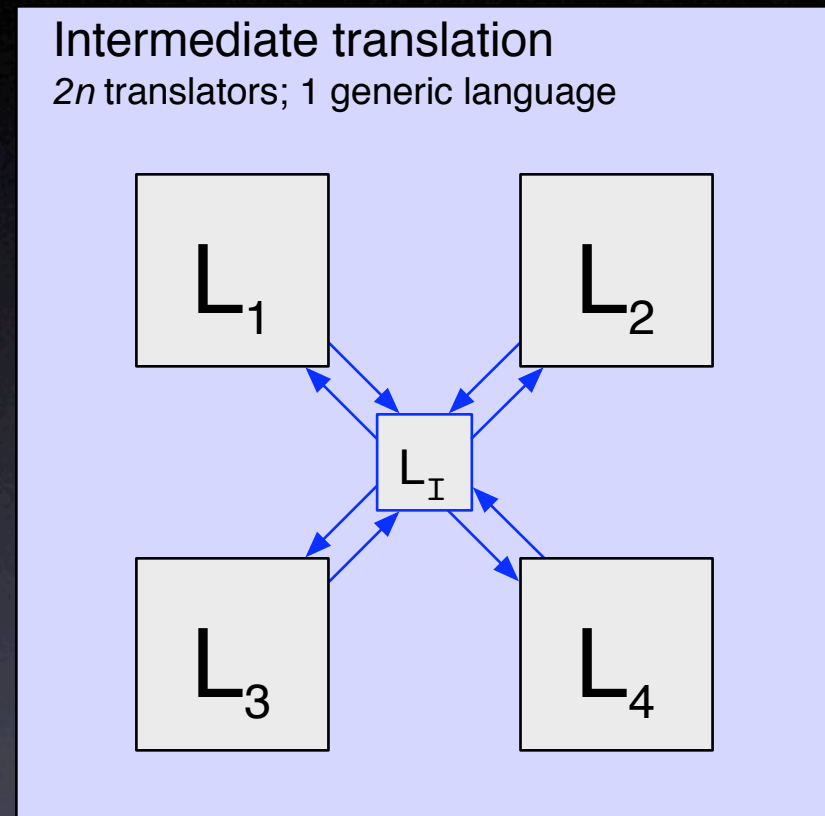
- $2(n-1)$  new translators when language added

# Intermediate translation

- Pros
  - Only 2 new translators when language added

- Cons

- Keeping intermediate language up to date
- All translators affected (potentially) by language addition

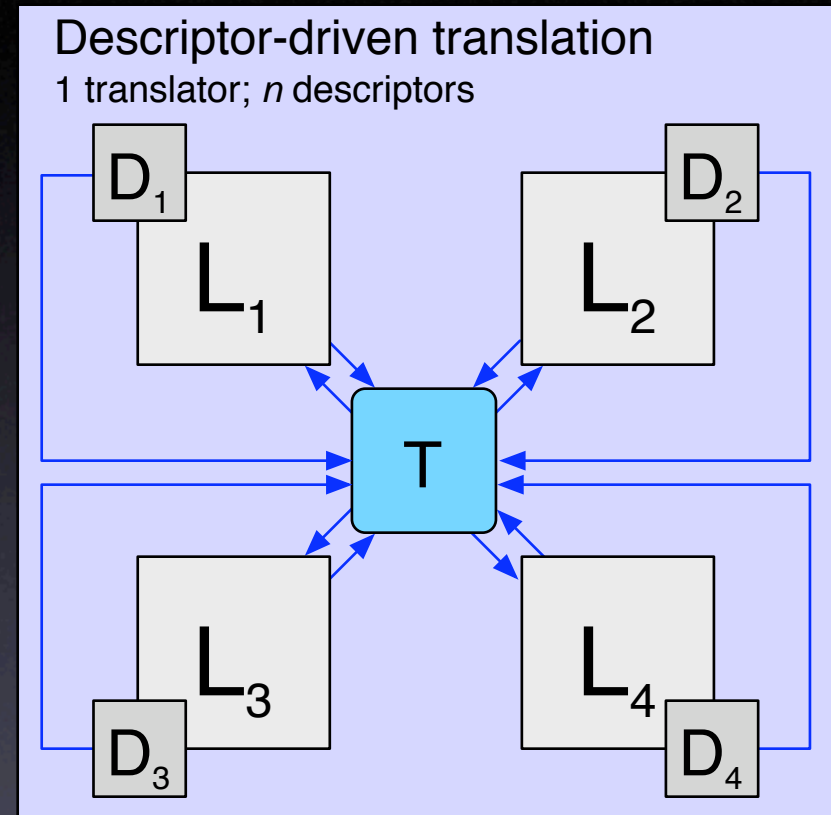


# Evaluation

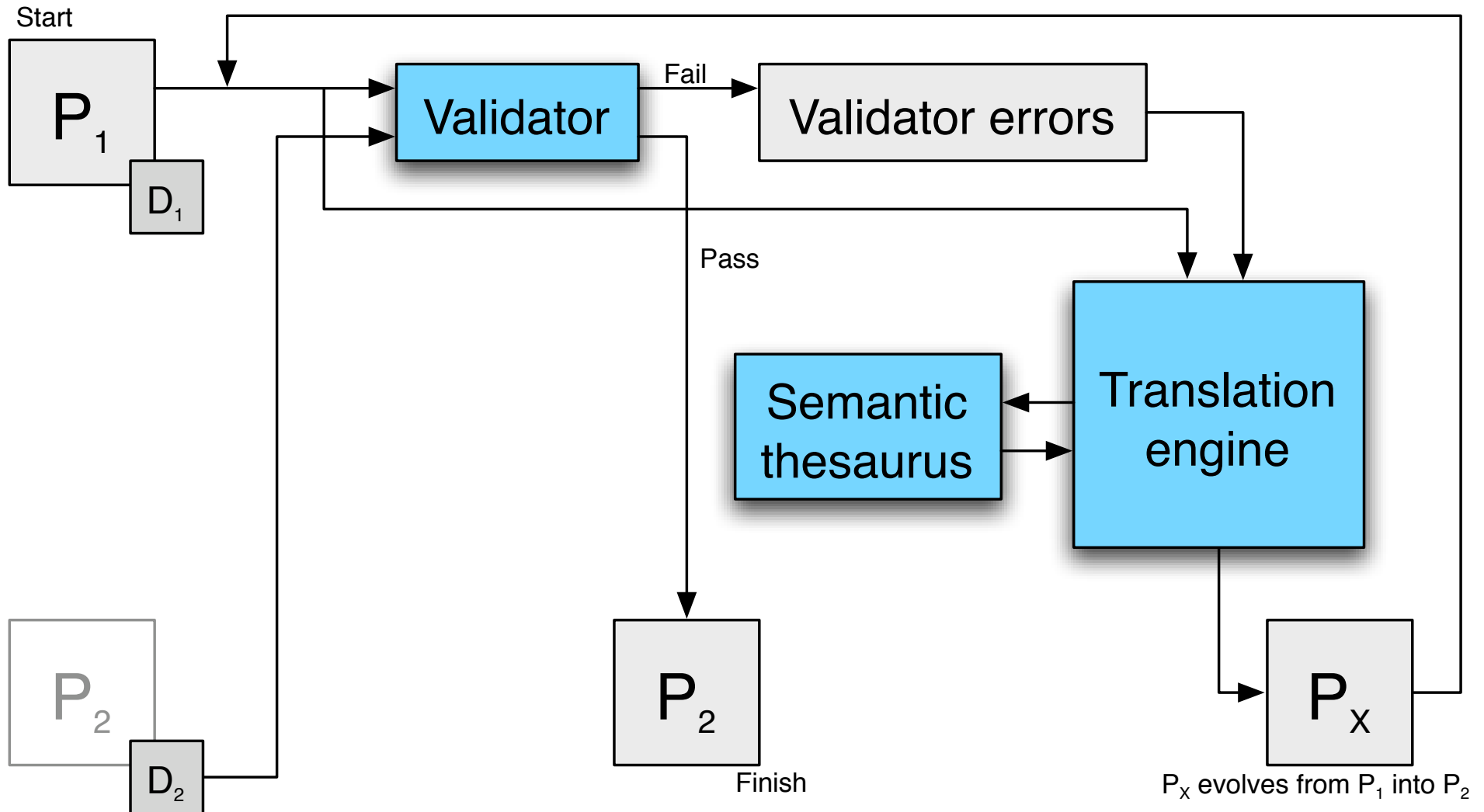
- Expensive
- Not suited to a semantically rich web
- Growing prominence and number of formats...
- Need an alternative method:
  - Descriptor Driven Translation (DDT)

# Descriptor Driven Translation (DDT)

- Pros
  - No new translators, just a descriptor
  - Additional languages added independently of others
- Cons
  - Reduced accuracy - Semantic Approximation



# DDT : Using a Schema Validator



# What will this research achieve?

- Develop techniques for:
  - extracting semantic information
  - generating mappings between semantic elements
- Exploit Semantic Web technologies



# Demonstrating the goal : Studies required

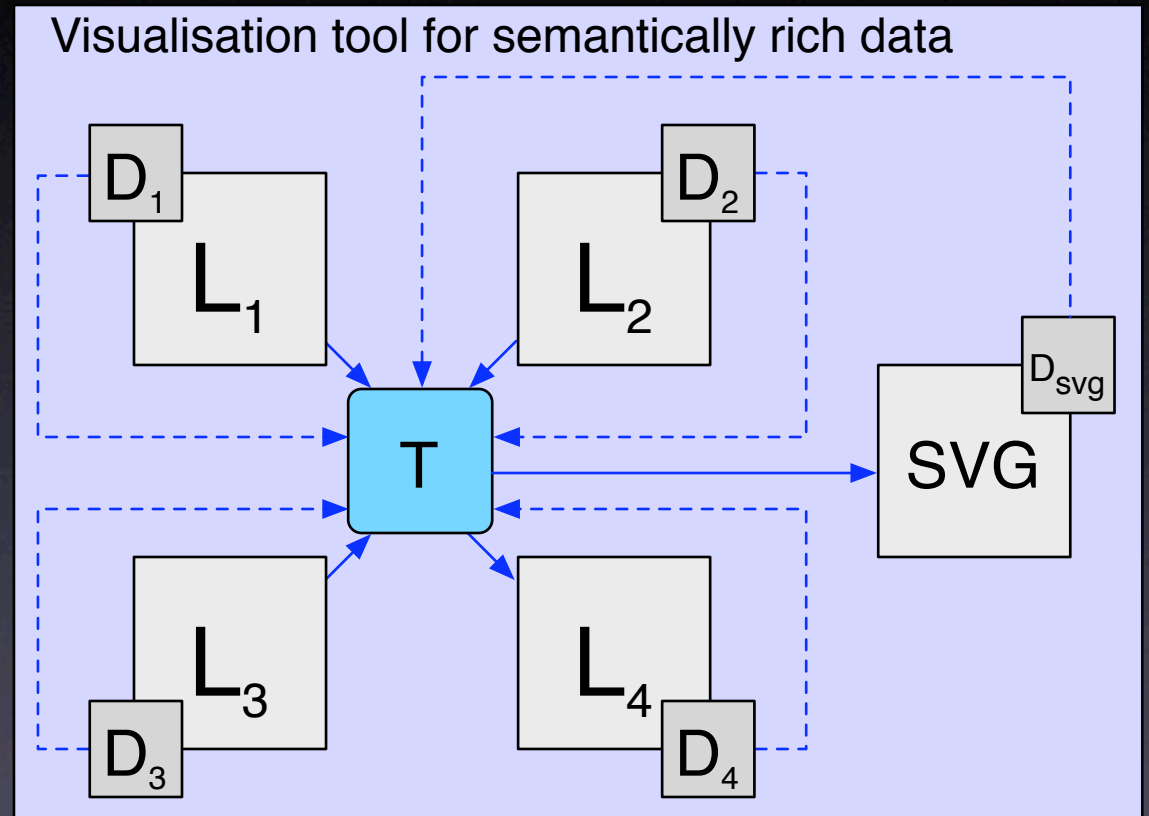
- Gather a set of example data in different formats
- Rationalise transformation types
- Improve the Schema Validator technique
- Develop further techniques
  - Structural analysis

# Metrics of success

- Compare translation accuracy with other methods
- Add a new language to the system and see how well it translates to/from existing languages

# Alternatively...

- A visualisation tool and editor for semantically rich data
- Separate structural aspects of the data from the syntax



# Summary

- Semantically rich formats = new challenges
- New translation method:
  - Descriptor Driven Translation
- First technique developed:
  - Schema Validation
- Proof of Concept produced
- Further work required beyond simplest examples

Thank you