

# SGNMT – A Flexible NMT Decoding Platform for Quick Prototyping of New Models and Search Strategies



UNIVERSITY OF  
CAMBRIDGE

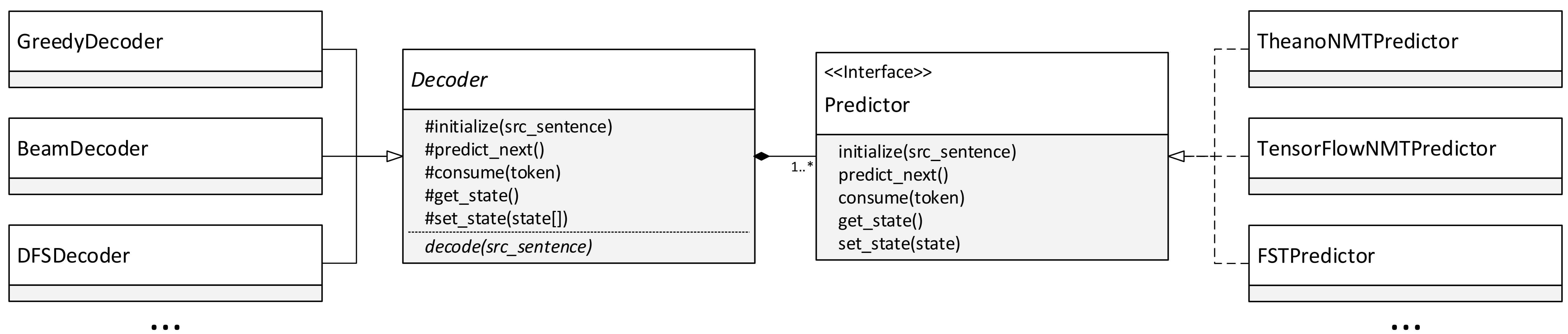
SDL\*

Felix Stahlberg, Eva Hasler, Danielle Saunders, Bill Byrne

## Introduction

- SGNMT is a decoding platform for MT research which makes it easy to combine various kinds of **neural** and **symbolic models** and **constraints**.
- Focus on **minimizing implementation effort** and supporting **multiple frameworks** (TensorFlow, Theano, NPLM, SRILM, OpenFST, ...).
- SGNMT is used for **course work** at the University of Cambridge, as well as for most of the research work in our group.

## Software architecture



### Search strategies (decoders):

- greedy: Greedy search
- beam: Beam search
- dfs: Depth-first search
- restarting: DFS with better pruning
- astar: A\* search
- sepbeam: System-level combination
- syncbeam: Beam search with sync symbol
- bucket: Multiple beam search passes
- vanilla: Fast NMT beam search

### Neural predictors:

- nmt (TensorFlow+Theano), rnnlm, nplm

### Constraining predictors:

- fst (deterministic), nfst (non-deterministic), rtn, forced (forced decoding), forcedlst (*n*-best rescoring), bow (bag of words)

### Symbolic scoring predictors:

- srilm, lrhiero (left-to-right Hiero (Siahbani et al., 2013)), wc (word penalty), unkc (Poisson UNK model), ngramc (MBR-based NMT), length

## Example decoder implementation

### Algorithm 1 Greedy(src\_sen)

```
1: initialize(src_sen)
2:  $h \leftarrow \langle \langle s \rangle \rangle$ 
3: repeat
4:    $P \leftarrow \text{predict\_next}()$ 
5:    $(t, c) \leftarrow \arg \max_{(t', c') \in P} c'$ 
6:    $h \leftarrow h \cdot t$ 
7:   consume( $t$ )
8: until  $t = \langle /s \rangle$ 
9: return  $h$ 
```

## Example predictor constellations

- nmt: Pure NMT decoding
- nmt, nmt, nmt: Ensemble of three NMT models
- fst, nmt: Lattice rescoring
- nmt, rnnlm, srilm, nplm: Combining NMT with three kinds of language models: Kneser-Ney *n*-gram, feedforward NNLM, RNNLM
- nmt, ngramc, wc: MBR-based NMT decoding

<http://ucam-smt.github.io/sgnmt/html/>