



Reproducible Results in Parsing-Based Machine Translation

The JHU Shared Task Submission

Lane Schwartz

University of Minnesota & Johns Hopkins University

lane@cs.umn.edu

Motivation: Reproducible Results

Research investigating natural language processing and computational linguistics can and should have an extremely low barrier to entry. The data with which we work is customarily available in common electronic formats. The computational techniques which we apply can typically be performed on commodity computing resources which are widely available. In short, there should be no reason why small research groups and even lone researchers should not be able to join and make substantive contributions furthering our field. The reality is less encouraging.

Many published articles describe novel techniques and provide interesting results, yet fail to describe technical details in sufficient detail to allow their results to be reproduced by other researchers. While there are notable and laudable exceptions, many publications fail to provide the source code and scripts necessary to reproduce results. The use of restricted data, not freely available for download by any interested researcher only compounds these problems. Pedersen (2010) rightly argues that the implementation details so often ignored in publications are in fact essential for our research to be reproducible science.

Reproducibility in machine translation is made more challenging by the complexity of experimental workflows. Results in machine translation tasks are dependent on a cascade of processing steps and configurations. While interesting subsets of these usually appear in experimental descriptions, many steps (preprocessing techniques, alignment parameters, translation rule extraction parameters, language model parameters, list of features used) are invariably omitted, even though these configurations are often critical to reproducing results.

Related Work

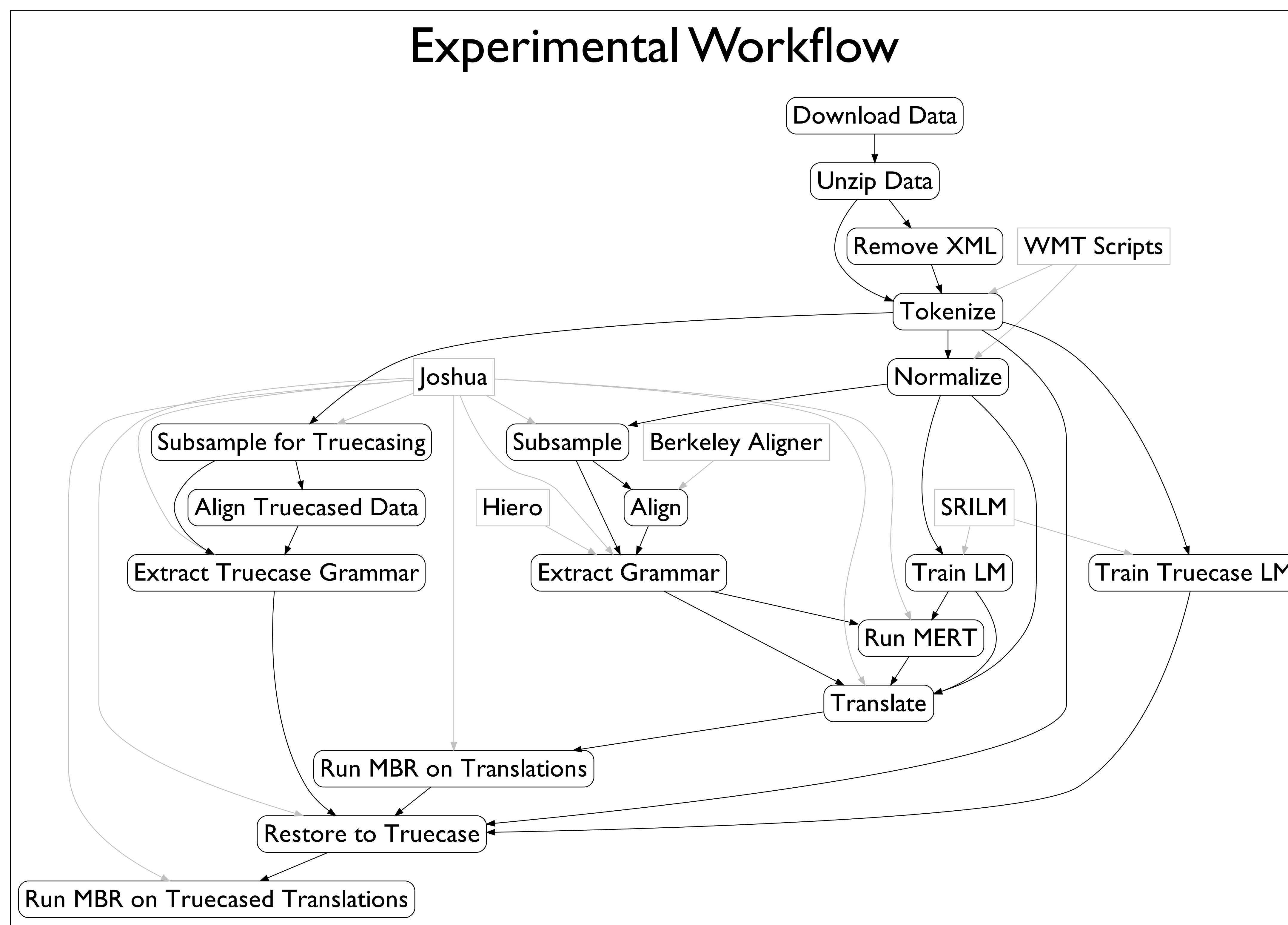
- LoonyBin (Clark et al., 2010) is a general hyperworkflow management utility for designing and running reproducible experiments.
- experiment.perl (Koehn et al., 2010) is a machine translation experiment management tool, available with the Moses machine translation system.

Translation Process

To enable fully reproducible results, we implemented a make-based machine translation workflow using open source tools and freely available data.

- Download & preprocess publicly available data
- Train lowercase & truecase target language models
- Subsample training corpus for test set
- Align bilingual parallel training corpus
- Extract hierarchical translation grammar
- Tune parameters using minimum error rate training
- Translate test set to lowercase target lang using Joshua
- Extract best lowercase translation from n-best list using Minimum Bayes Risk extraction
- Extract lowercase to truecase translation grammar
- Translate test set to truecase target lang using Joshua
- Extract best truecase translation from n-best list using Minimum Bayes Risk extraction

Experimental Workflow



Language Pair	Bilingual Parallel Training Corpora
German, English	news-commentary10.de-en europarl-v5.de-en
French, English	news-commentary10.fr-en europarl-v5.fr-en giga-fren.release2 undoc.2000.en-fr
Spanish, English	news-commentary10.es-en europarl-v5.es-en undoc.2000.en-es

Target Lang	Monolingual Language Model Corpora
English	europarl-v5.en news-commentary10.en news.en.shuffled undoc.2000.en-fr.en giga-fren.release2.en
French	europarl-v5.fr news-commentary10.fr news.fr.shuffled undoc.2000.en-fr.fr giga-fren.release2.fr
German	europarl-v5.de news-commentary10.de news.de.shuffled
Spanish	europarl-v5.es news-commentary10.es news.es.shuffled undoc.2000.en-es.en

Empirical Results

Source	Target	BLEU	TER
German	English	21.3	0.660
English	German	15.2	0.738
French	English	27.7	0.614
English	French	23.8	0.681
Spanish	English	29.0	0.595
English	Spanish	28.1	0.596

WMT newstest 2010. Our de-en submission was the best-scoring WMT system according to TER.

Reproduce these Results

To reproduce these results run these make scripts: <http://sourceforge.net/projects/joshua/files/joshua/1.3/wmt2010-experiment.tgz/download>

The scripts download training & test data, and download, compile, configure and run the software used in this submission. SRILM must be compiled and installed separately.