

# Moses, past, present, future

Hieu Hoang

XRCE 2013

# Timeline

2002	Pharoah decoder, precursor to Moses
2005	Replacement for Pharoah
2006	JHU Workshop extends Moses significantly
since late 2006	Funding by EU projects EuroMatrix, EuroMatrixPlus
2012	MosesCore

# What is Moses?

## Common Misconceptions

- Only the decoder
- Only for Linux
- Difficult to use
- Unreliable
- Only phrase-based
- No sparse features
- Developed by one person
- Slow

# Only the decoder

- replacement for Pharoah
- Training
- Tuning
- Decoder
- Other
  - XML Server. Phrase-table pruning/filtering.
  - Domain adaptation. Experiment management system

# Only works on Linux

- Tested on
  - Windows 7 (32-bit) with Cygwin 6.1
  - Mac OSX 10.7 with MacPorts
  - Ubuntu 12.10, 32 and 64-bit
  - Debian 6.0, 32 and 64-bit
  - Fedora 17, 32 and 64-bit
  - openSUSE 12.2, 32 and 64-bit
- Project files for
  - Visual Studio
  - Eclipse on Linux and Mac OSX

# Difficult to use

- Easier compile and install
  - Boost bjam
  - No installation required
- Binaries available for
  - Linux
  - Mac
  - Windows/Cygwin
  - Moses + Friends
    - IRSTLM
    - GIZA++ and MGIZA
- Ready-made models trained on Europarl

# Unreliable

- Monitor check-ins
- Unit tests
- More regression tests
- Nightly tests
  - Run end-to-end training
  - <http://www.statmt.org/moses/cruise/>
- Tested on all major OSes
- Train Europarl models
  - Phrase-based, hierarchical, factored
  - 8 language-pairs
  - <http://www.statmt.org/moses/RELEASE-1.0/models/>

# Only phrase-based model

- ~~replacement for Pharaoh~~
- extension of Pharaoh
- From the beginning
  - Factored models
  - Lattice and confusion network input
  - Multiple LMs, multiple phrase-tables
- since 2009
  - Hierarchical model
  - Syntactic models



# No Sparse Features

- Large number of sparse features
  - 1+ millions
  - Sparse AND dense features
- Available sparse features

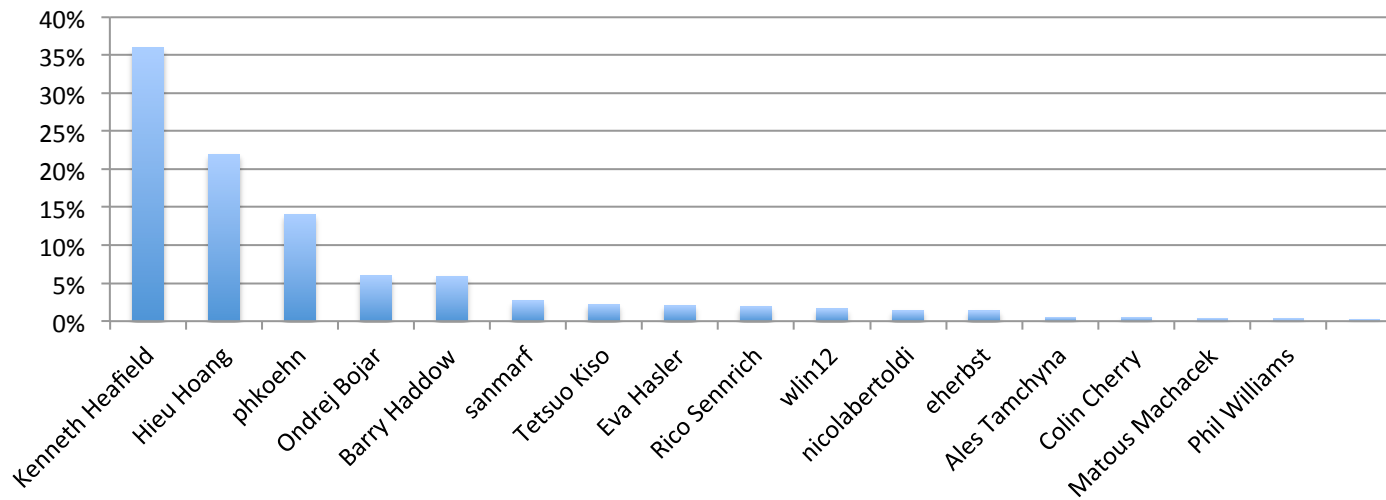
Target Bigram	Target Ngram	Source Word Deletion
Sparse Phrase Table	Phrase Boundary	Phrase Length
Phrase Pair	Target Word Insertion	Global Lexical Model

- Different tuning
  - MERT
  - Mira
  - Batch Mira (Cherry & Foster, 2012)
  - PRO (Hopkins and May, 2011)

# Developed by one person

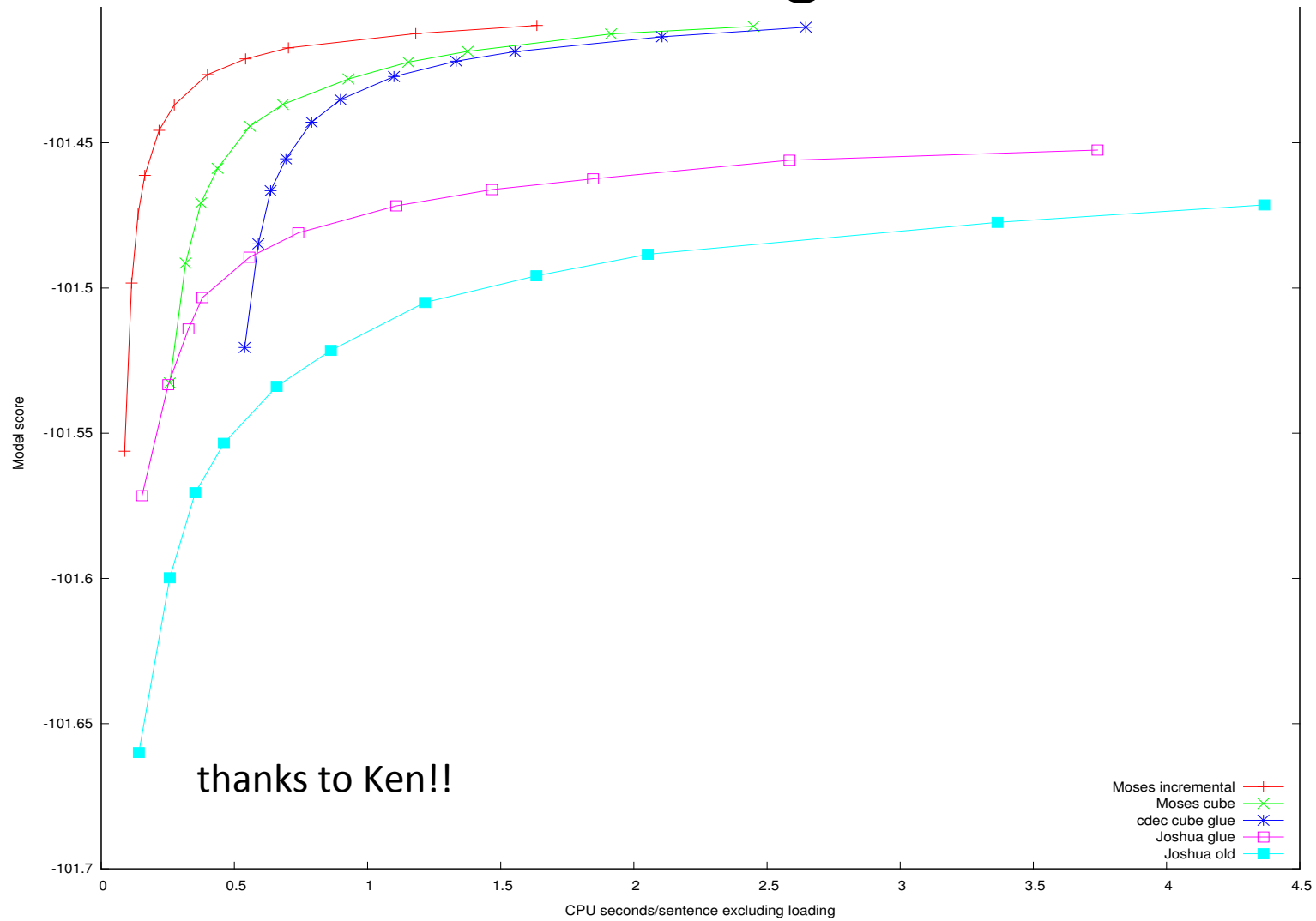
- ANYONE can contribute

'git blame' of Moses repository



– 50 contributors

# Slow Decoding



thanks to Ken!!

- Moses incremental
- Moses cube
- cdec cube glue
- Joshua glue
- Joshua old

# Slow Training

- Multithreaded

Time (mins)	1-core	2-cores	4-cores	8-cores	Size (MB)
Phrase-based	60	47 (79%)	37 (63%)	33 (56%)	893
Hierarchical	1030	677 (65%)	473 (45%)	375 (36%)	8300

- Reduced disk IO
  - compress intermediate files
- Reduce disk space requirement

# What is Moses?

## Common Misconceptions

- Only the decoder
- Only for Linux
- Difficult to use
- Unreliable
- Only phrase-based
- No sparse features
- Developed by one person
- Slow

# What is Moses?

## Common Misconceptions

- ~~Only the decoder~~ Decoding, training, tuning, server
- ~~Only for Linux~~ Windows, Linux, Mac
- ~~Difficult to use~~ Easier compile and install
- ~~Unreliable~~ Multi-stage testing
- ~~Only phrase based~~ Hierarchical, syntax model
- ~~No sparse features~~ Sparse AND dense features
- Developed by ~~one person~~ everyone
- ~~Slow~~ Fastest decoder, multithreaded training, less IO

# Future priorities

- Code cleanup
- MT applications
  - Computer-Aided Translation
  - Speech-to-speech
- Incremental Training
- Better translation
  - smaller model
  - bigger data
  - faster training and decoding

# Code cleanup

- Framework for feature functions
  - Easier to add new feature functions
- Cleanup
  - Refactor
  - Delete old code
  - Documentation



# MT Applications

- Computer-Aided Translation



- integration with front-ends
- better user of user-feedback information

# MT Applications

- Speech-to-speech



- ambiguous input
  - lattices and confusion networks
- translate prosody
  - factored word representation

# Incremental Training

- Incremental word alignment
- Dynamic suffix array
- Phrase-table update
  
- Better integration with rest of Moses

# Smaller files

- Smaller binary
  - phrase-tables
  - language models
- Mobile devices
- Fits into memory
  - faster decoding!
- Efficient data structures
  - suffix arrays
  - compressed file formats

# Better Translations

- Consistently beat phrase-based models for every language pair

	Phrase-Based	Hierarchical
en-es	24.81	24.20
es-en	23.01	22.37
en-cs	11.04	10.93
cs-en	15.72	15.68
en-de	11.87	11.62
de-en	15.75	15.53
en-fr	22.84	22.28
fr-en	25.08	24.37
zh-en	27.46	23.91
ar-en	47.90	46.56

The End