Abkhazia – ASR experiments made easy

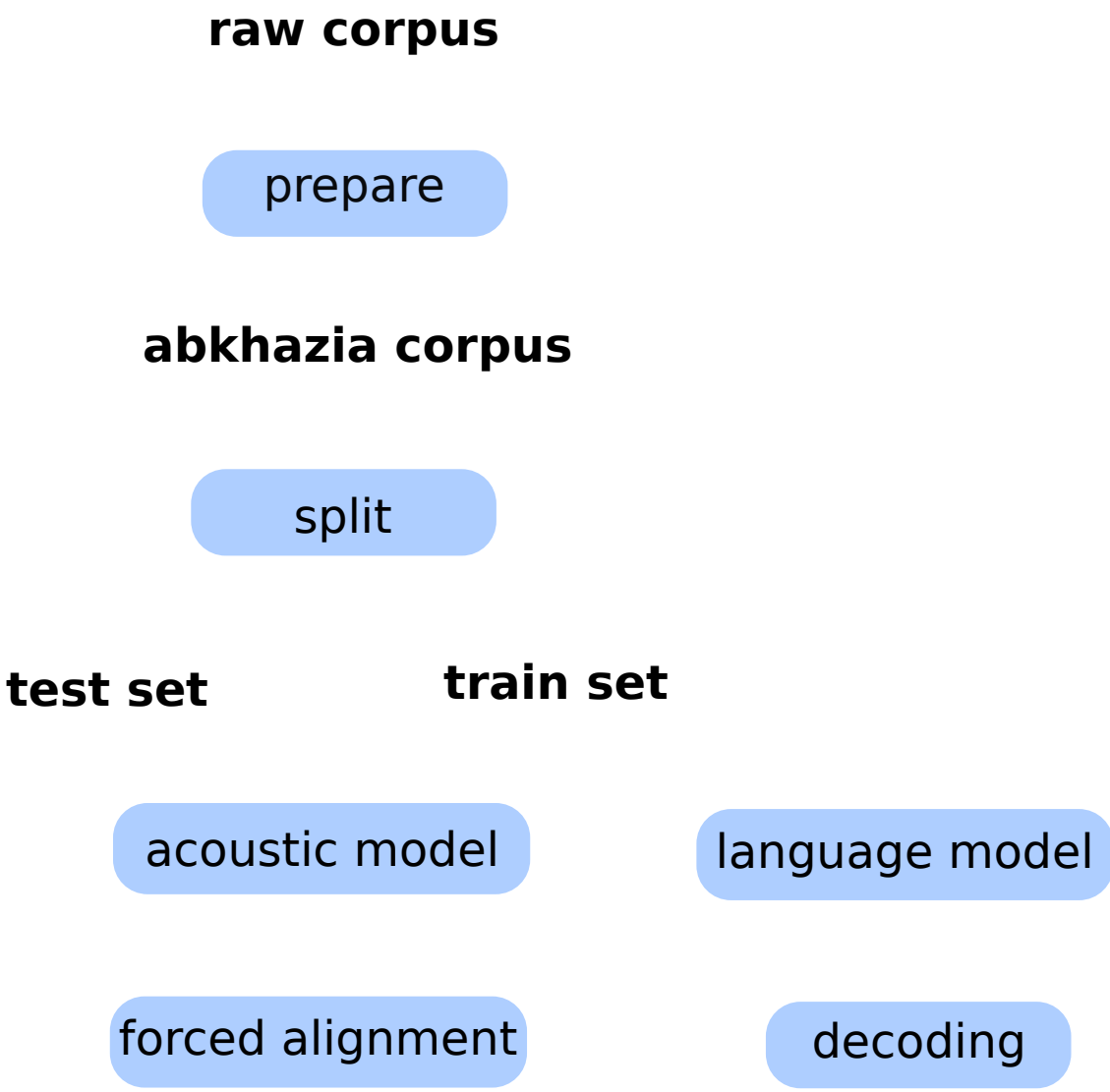
Xuan Nga Cao Mathieu Bernard

LSCP Bootphon Team Meeting – March 24, 2016

Abkhazia – ASR experiments made easy

- Abkhazia is a Python library and a **command-line tool**
  - sources and installation instructions are available at *https://github.com/bootphon/abkhazia*
  - once installed, get in with `abkhazia --help`
- Performs **various ASR tasks**
  - use the Kaldi toolkit for speech recognition (kaldi-asr.org)
  - language and acoustic models, forced-alignment, decoding
  - computation on the cluster or locally
  - all from a uniform command-line syntax
- Defines and rely on a **standard speech corpus format**
  - inspired by Kaldi format
  - support for several corpora (WSJ, Buckeye, etc. . .)
  - possible extension to new corpora

The big picture – what abkhazia can do



Standard speech corpus format

TODO XN

Supported corpora

- aic - Articulation Index Corpus LSCP
- buckeye - Buckeye Corpus of conversational speech
- csj - Corpus of Spontaneous Japanese
- globalphone - GlobalPhone multilingual read speech corpus
- librispeech - LibriSpeech ASR Corpus
- wsj - Wall Street Journal ASR Corpus
- xitsonga - NCHLT Xitsonga Speech Corpus

Abkhazia commands

From `abkhazia --help`

- prepare - prepare a speech corpus for use with abkhazia
- split - split a corpus in train and test subsets
- language - compute a language model
- train - train (or retrain) an acoustic model
- decode - compute phone posteriograms or transcription
- align - compute forced-aligment

Each command have its own help message as

```
abkhazia <command> --help
```

abkhazia prepare: [raw] -> [corpus]

Convert a corpus from its raw distribution to the abkhazia format.
Exemple: Buckeye preparation

- read the buckeye corpus from `/path/to/raw/buckeye`
- convert it to `<abkhazia-data-dir>/buckeye/data`

```
abkhazia prepare buckeye /path/to/raw/buckeye
```

abkhazia split: [corpus] -> [corpus], [corpus]

Split a corpus in train and test subsets.

abkhazia language: [corpus] -> [lm]

Generate a language model from a prepared corpus. Write the directory `<corpus>/language`.

abkhazia train: [corpus], [lm] -> [model]

Train a standard speaker-adapted triphone HMM-GMM model from a prepared corpus. Write the directory `<corpus>/model`.

abkhazia align: [model], [lm] -> [result]

Generate a forced-alignment from acoustic and language models. Write the directory `<corpus>/align`.

abkhazia decode: [corpus], [model], [lm] -> [result]

Decode a prepared corpus from a HMM-GMM model and a language model. Write the directory `<corpus>/decode`.

Example – forced-alignment of a buckeye subset

TODO XN

Conclusion

- And the answer is...
- $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$