University of Le Mans, France

Some recent research work at LIUM based on the use of CMU Sphinx

Yannick Estève, Paul Deléglise, Sylvain Meignier, Holger Schwenk, Loïc Barrault, Fethi Bougares, Richard Dufour, Vincent Jousse, Antoine Laurent, Anthony Rousseau

presented by Richard Dufour

CMU SPUD Workshop 2010, Dallas

Developers and users

- LIUM as developers:
  - to improve WER performance
  - to adapt Sphinx to our needs
  - SphinxTrain, Sphinx3, Sphinx4
- LIUM as users
  - Sphinx as tools for research
  - for speech recognition research work: phonetization, error correction
  - And other: spontaneous speech detection, machine translation

Development

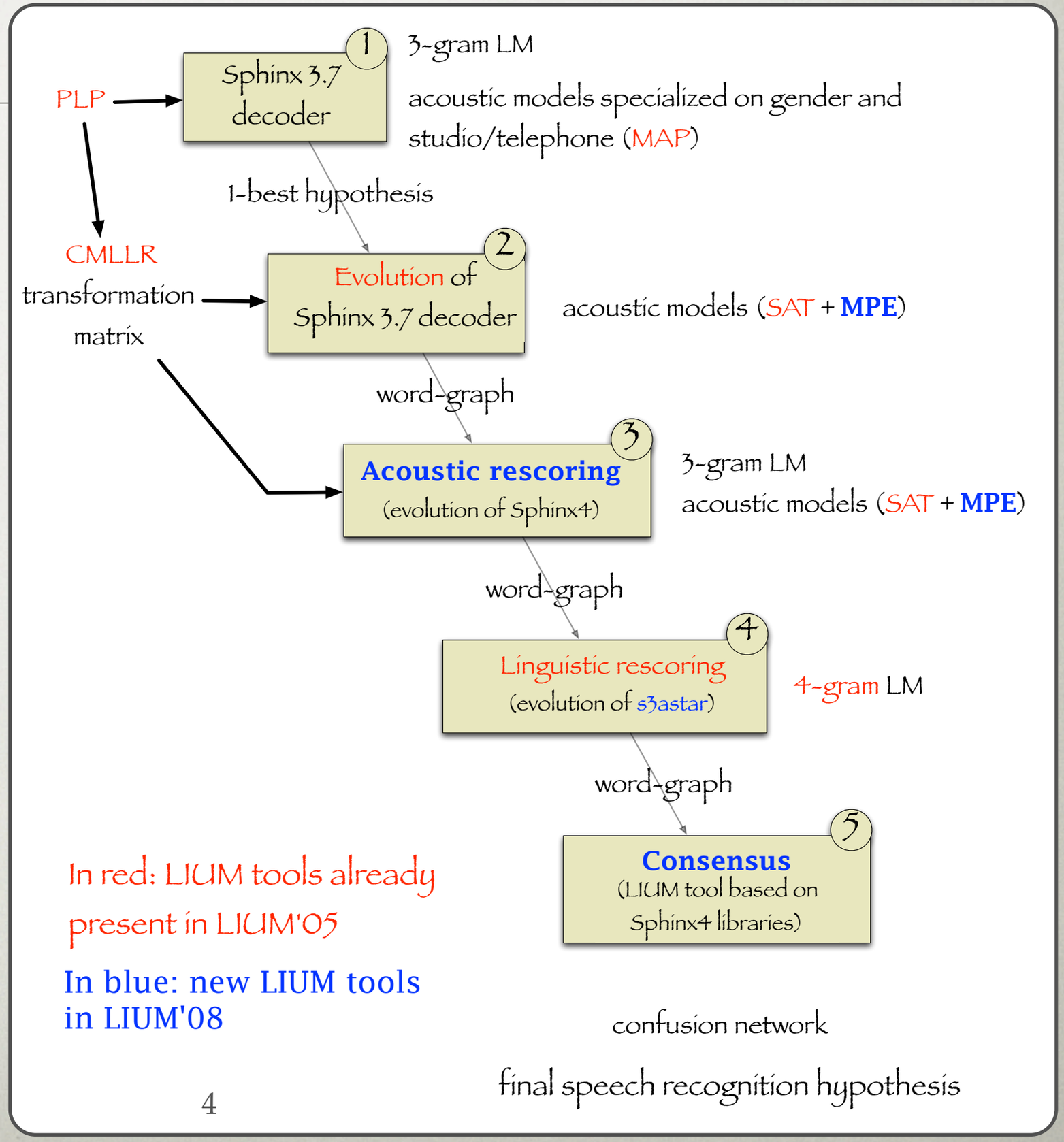
- Multi-Pass ASR system initially developed for Broadcast news
- Segmentation and Speaker diarization system (lium_spkdiarization)
- Added functionalities:
  - Training and adapting acoustic models
    - Speaker Adaptive Training (with CMLLR), Minimum Phone Error
  - Decoding process
    - Acoustic and Linguistic Word-graph Rescoring, 4-gram rescoring, Consensus using confusion network, >65K vocab (sphinx4)

LIUM'08 ASR System

Results on the test data from the ESTER2 French Evaluation Campaign

LIUM'08	WER
pass 1	27.1%
pass 2	22.5%
pass 3	20.4%
pass 4	19.4%
pass 5	19.2%

WER according to the pass in LIUM'08



LIUM participation to the CMU Sphinx project

- SphinxTrain: source code of CMLLR+SAT given to Arthur Chan (Sphinx Maintainer) in 2006 (code too messy, not integrated yet...)
- LIUM branch created in the official Sphinx SVN server in 2009 containing our Sphinx4 code (messy code, no doc, but it works !), and some Sphinx3 code (clean, especially to deal with n-gram LM with n>3 and rescore word-lattice in 4-gram)
- In 2010: first integration in the official source code of Sphinx4 (vocab >65K + class to deal with n-grams, with n>3)
  - Expecting other contributions in the official code in the future

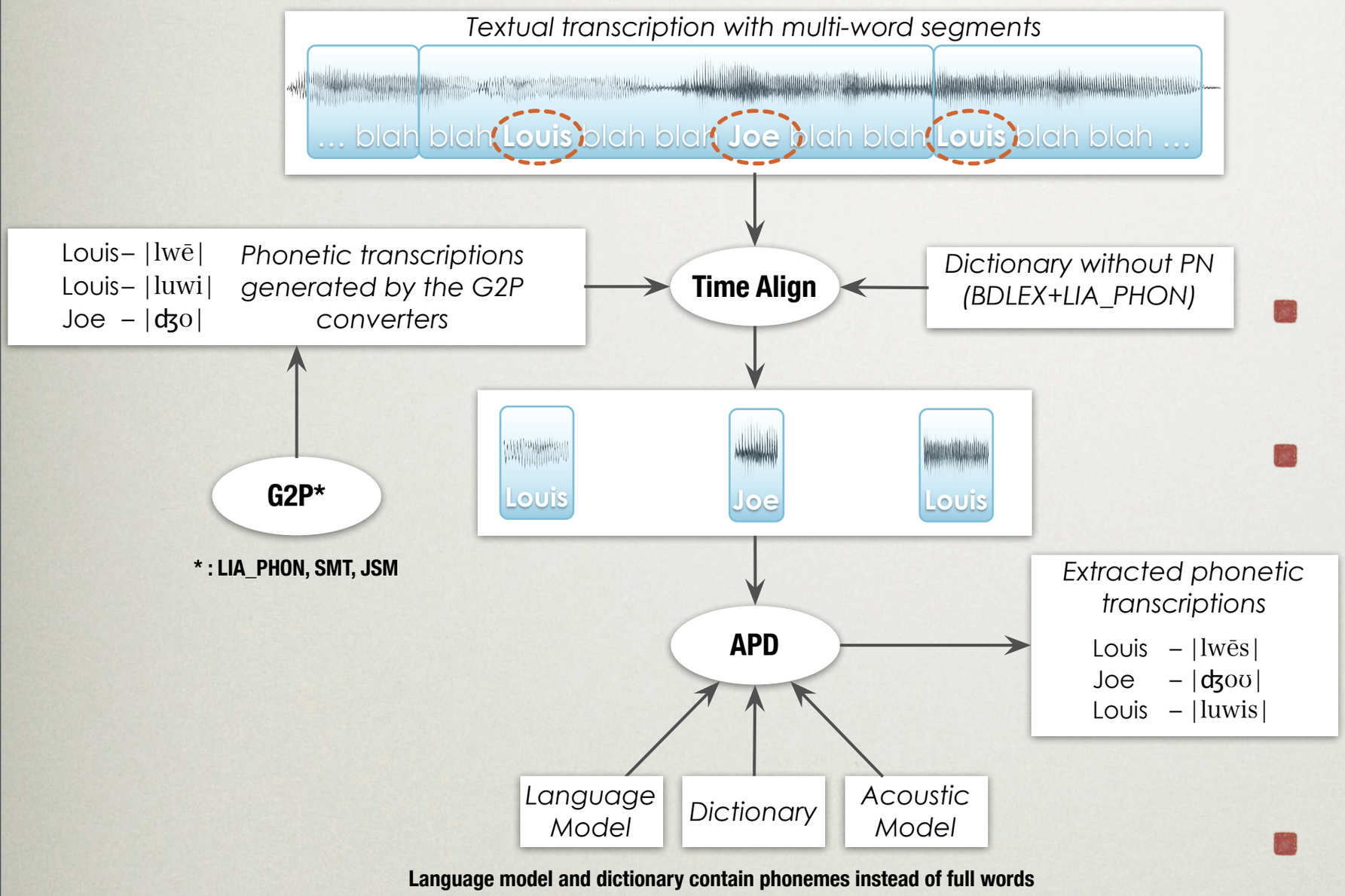
Using Sphinx

- Acoustic-Based phonetic transcription method for proper nouns
- Combining outputs of machine translation
- Specific corrections for specific errors in French
- Spontaneous speech detection in large audio database
- And other works not presented here...

Phonetic transcription of proper nouns

- Proper nouns constitute a special case when it comes to phonetic transcription
  - In French, pronunciation rules are much less normalized for proper nouns than for other categories
- We propose a method that relies on speech signal
- Process consists of two major steps:
  - Finding boundaries of proper nouns
  - Retrieving pronounced phonemes

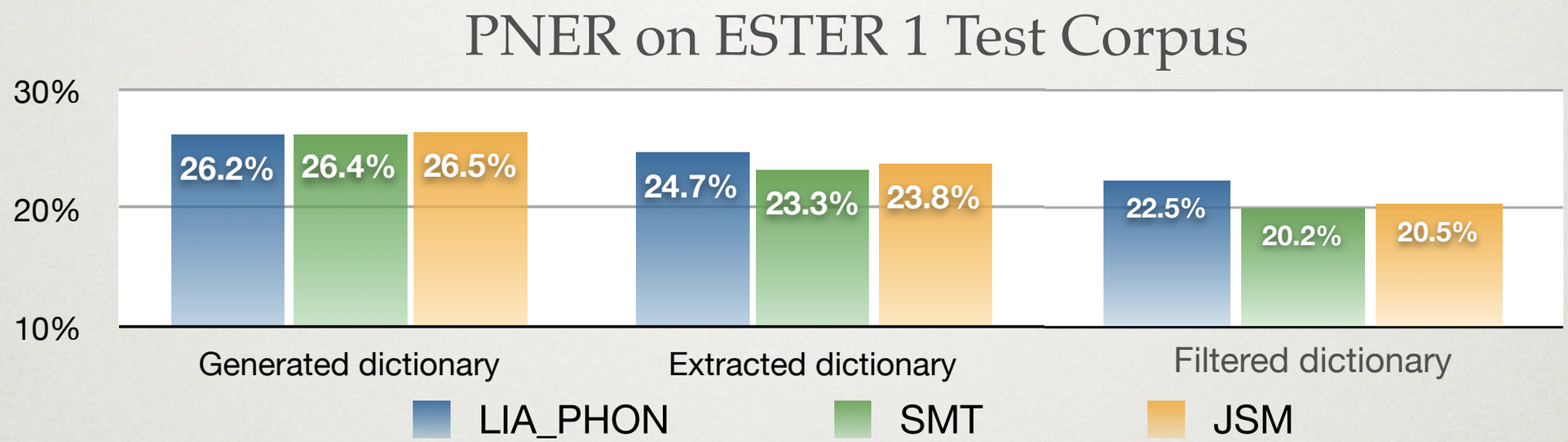
Phonetic transcription of proper nouns

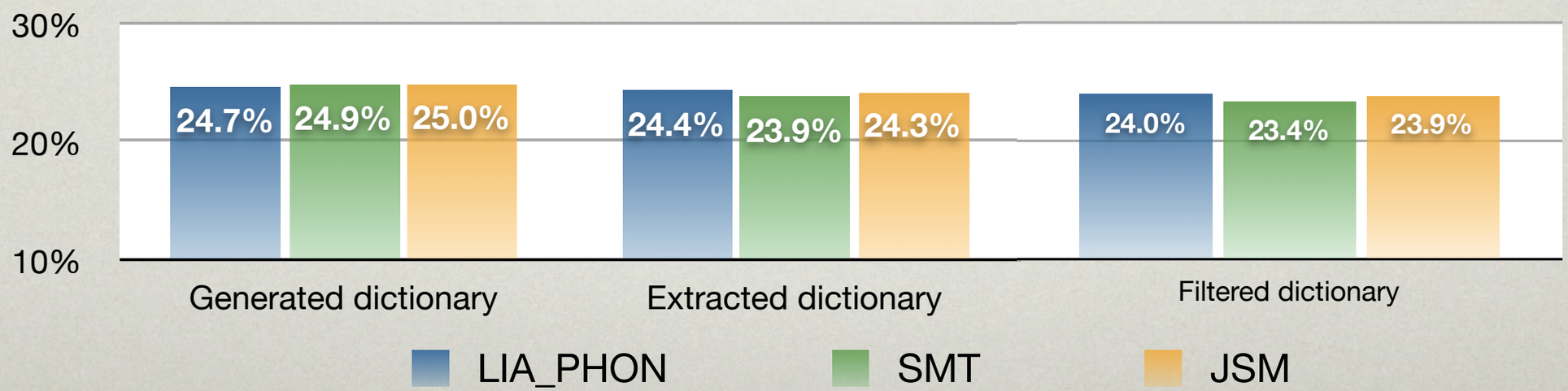


- TimeAlign is done by using sphinx3_align
- Acoustic Phonetic Decoding using Sphinx 3
- Each phonetic transcription gets validated by an iterative filtering

Phonetic transcription of proper nouns

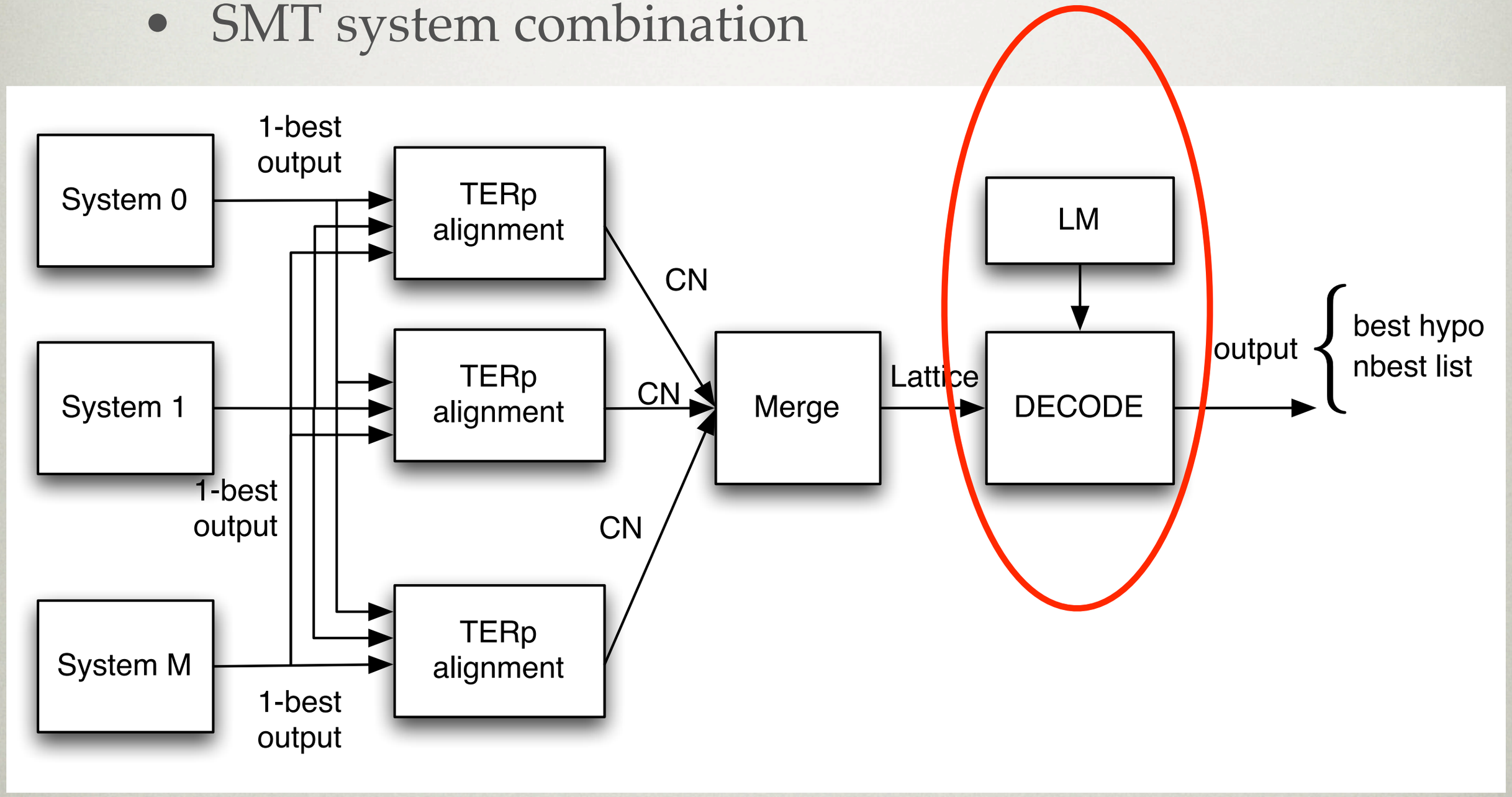
- *Results :*





Statistical Machine Translation

- SMT system combination



Statistical Machine Translation

- Decoding
  - Token pass decoding algorithm
  - Based on the Sphinx 4 library
  - Using a language model hosted on a lm-server (no restriction of n-gram size)
  - Probabilities computed by the decoder:

$$log(P_W) = \sum_{n=0}^{Len(W)} [log(P_{ws}(n)) + \alpha P_{lm}(n)] + Len_{pen}(W) + Null_{pen}(W)$$

Statistical Machine Translation

- System successfully used in the IWSLT'09 evaluation campaign
- Open-source and available here:
  - http://www-lium.univ-lemans.fr/~barrault/MANY

Improving French ASR by targeting specific errors

- Context
  - Errors, which do not prevent understanding, are often neglected
  - Example: agreement in number and gender
  - But could be important for some applications (subtitling, assisted-transcription...)

Improving French ASR by targeting specific errors

- Approach
  - Repair some errors by post-processing the ASR output obtained with the Sphinx decoder
  - Build a specific correction solution for each specific error
  - Complex grammatical rules can not be modeled with a n-gram language model

Improving French ASR by targeting specific errors

Confusion pairs in ASR outputs

110	→	à ➟ a
91	→	est ➟ ces
70	→	cent ➟ cents
69	→	est ➟ et
45	→	des ➟ les
9	→	vingts ➟ vingt
7	→	chargée ➟ chargé
4	→	organisée ➟ organisé
4	→	force ➟ forces

etc.

Analysis of most frequent errors and generalization

Find agreement errors that seem correctable

Automatic error detection

Use of a specific detector for each kind of error

Potential errors

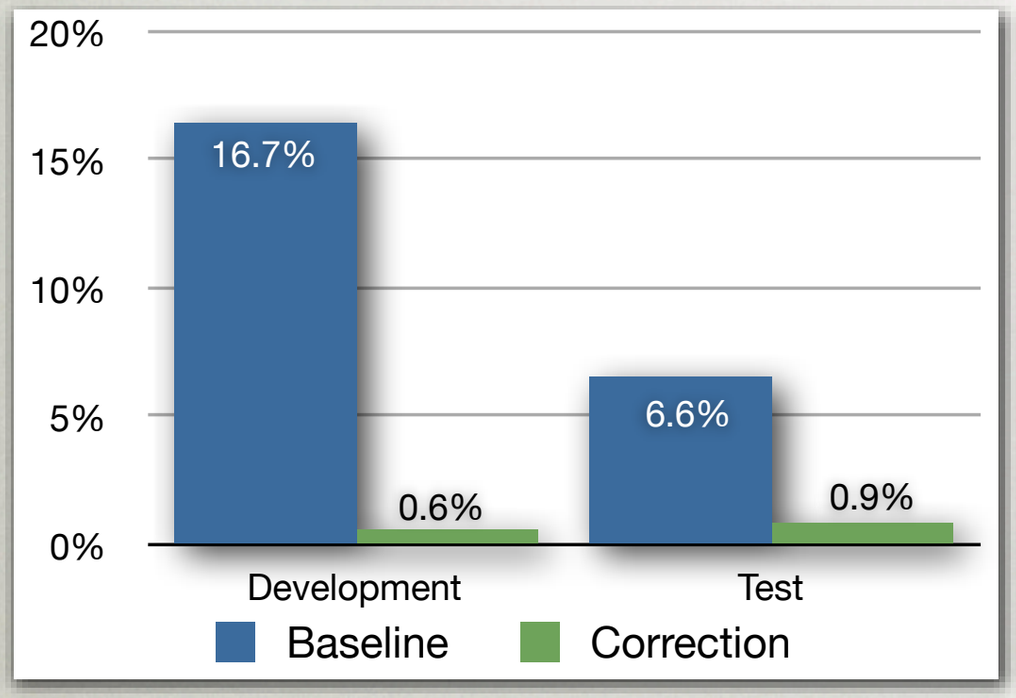
Automatic error recovery

Statistical Method

Formal Rule

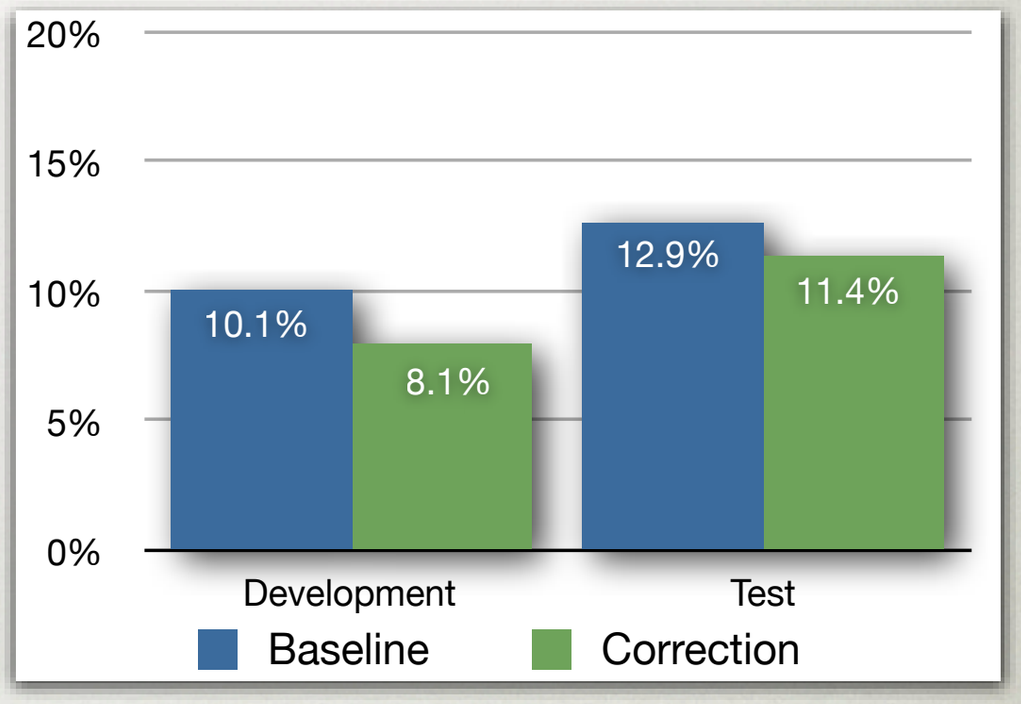
Improving French ASR by targeting specific errors

Formal rule



Rate of agreement errors on word ‘cent (hundred)’ and ‘vingt (twenty)’

Statistical method



Rate of past participles having an agreement error

Spontaneous speech characterization and detection

- Approach
  - Spontaneous speech
    - Disfluencies, ungrammaticality...
    - More difficult to transcribe than prepared speech
  - Study specific features of spontaneous speech in opposition to prepared speech

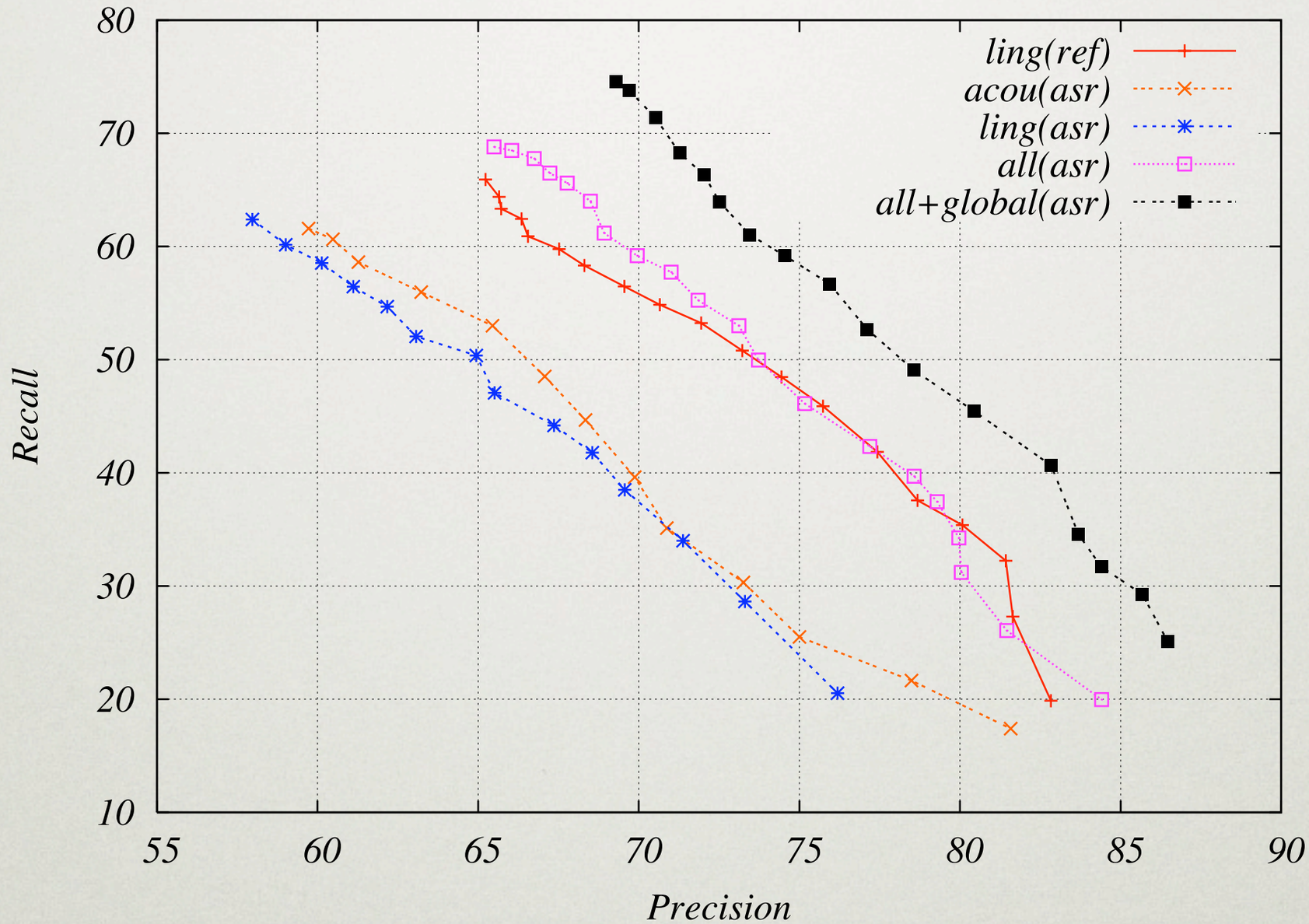
Spontaneous speech characterization and detection

- Corpus
  - 11 hour corpus (French Broadcast News)
    - Manually labeled by two human judges
    - 3 classes of spontaneity (prepared, low sponta., high sponta.)
    - Cut into segments (automatic segmentation and diarization)
    - Transcribed (by the Sphinx decoder)

Spontaneous speech characterization and detection

- Automatic detection (two levels)
  - At segment level
    - 3 sets of features used (acoustic, linguistic and confidence measures given by ASR)
    - Classification process
  - At audio file level
    - Taking into consideration information about surrounding segments after segment classification
    - Statistical method

Spontaneous speech characterization and detection



Detection performance on high spontaneous segments

Discussion (1)

- At this time, it is hard for us to merge our code with the current Sphinx code
  - we have developed the major part of our current tools from
    - Sphinx4-beta1 (2004)
    - Sphinx3.5 (2005)
    - SphinxTrain (2003)
  - In the mean time, the official code evolved a lot
    - it's a good thing for Sphinx, but it's hard for us to merge against that code

Discussion (2)

- Is it possible to build a common roadmap to anticipate future changes (and to help for future collaboration between all the Sphinx developers)?
- Can we work together around a common project (demonstrator, evaluation campaign, or other)?
  - in order to federate our efforts
  - to valorize the CMU Sphinx project

Contacts

- General contact: yannick.esteve@lium.univ-lemans.fr
- Sphinx Development
  - yannick.esteve@lium.univ-lemans.fr
  - sylvain.meignier@lium.univ-lemans.fr
  - paul.deleglise@lium.univ-lemans.fr
- Acoustic-Based phonetic transcription method for proper nouns
  - antoine.laurent@lium.univ-lemans.fr
- Combining outputs of machine translation
  - loic.barrault@lium.univ-lemans.fr
- Specific corrections for specific errors in French
  - richard.dufour@lium.univ-lemans.fr
- Spontaneous speech detection in large audio database
  - richard.dufour@lium.univ-lemans.fr