

Reshaping Automatic Speech Transcripts for Robust High-level Document Analysis

Julien Fayolle, Fabienne Moreau,
Christian Raymond, Guillaume Gravier

INRIA & IRISA, Rennes, France

AND 2010

Fourth workshop on Analytics for Noisy Unstructured Data

Our text data

- Speech transcripts of TV broadcast news



“Le suisse **Cancellara** reste maillot jaune.”

Speech Recognition

ERROR!

le suisse quand c' est lara reste maillot jaune

- Unstructured
 - ✓ no sentence, no punctuation, no capitalization
- Noisy
 - ✓ from 10% to 60% of words are misrecognized

High-level document analysis

- Spoken language processing applications
 - ✓ spoken document retrieval, summarization, machine translation, topic threading, named entity recognition, ...
 - ✓ harder on noisy text than on clean text

→ Requirements

- ✓ Is this word reliable ? (confidence measure)

le	suisse	quand	c'	est	lara	reste	maillot	jaune
yes	yes	no	no	no	no	yes	yes	yes

- ✓ Is this word meaningful ? (named entity)

le	suisse	quand	c'	est	lara	reste	maillot	jaune
	person	mis-recognized person						

Problem

- Provide reliable and rich information from noisy transcripts
- We need a reliable confidence measure (CM)
 - ✓ value in $[0, 1]$ for each word
 - ✓ computable from any speech transcripts
 - ✓ focusing on rich information (named entities)

State-of-the-art on confidence measure

System	Small Voca (6k words)	Large Voca (~60k words) Continuous Speech	
		Specific errors	All words
ASR CM	✓	✓	✓
Phonetic	✓		✓
Syntactic	✓		✓
Linguistic	✓	✓	✓
Semantic	✓	✓	NE
Context	✓	✓	✓
Machine learning combination	TBL SVM Adaboost	Adaboost	CRF

ASR CM

ASR: Automatic Speech Recognition

CM: Confidence Measure [0,1]

- From any ASR system
- Baseline
- **Goal: Enrich this CM with other sources of information**
 - ✓ **Standard features (S)**
 - ✓ **Named entity features (NE)**

Standard features (S)

Well-known and easy to obtain

→ Morpho-syntactic

- ✓ part-of-speech

feminine plural noun,
verb, adjective, ...

→ Linguistic

- ✓ language model
back-off behavior

largest ngram in the
LM from 1 to 4

→ Word length

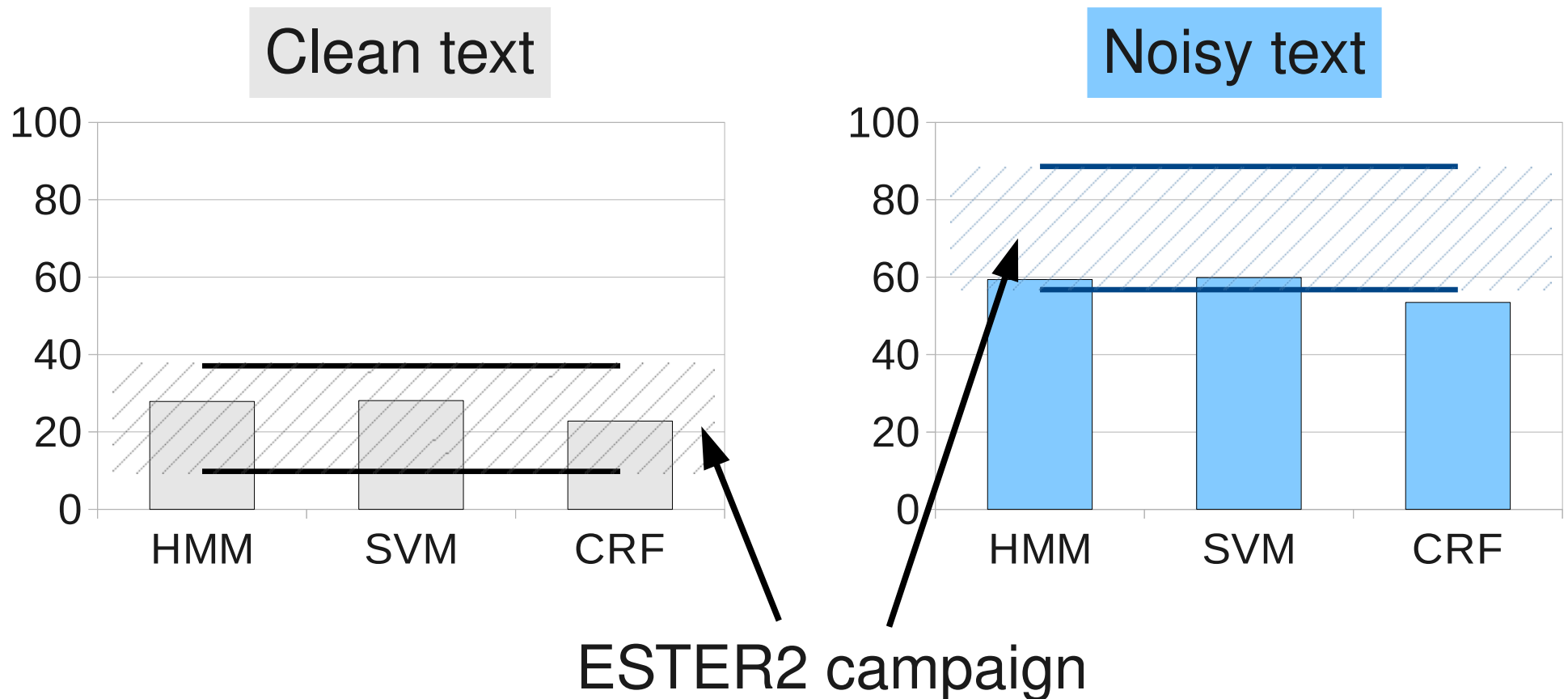
- ✓ duration
- ✓ number of
phonemes

rather short words vs.
rather long words

Named entity features (NE)

→ 3 robust NER systems [Raymond TALN'10]

Performance on ESTER2 (slot error rate)



Named entity features (NE)

→ 5 NE features

- ✓ class NE-HMM
- ✓ class NE-SVM
- ✓ class NE-CRF

✓ $p(\text{NE-CRF})$ in $[0,1]$

✓ agreement class {'yes', 'no', 'maybe'}

agreement
NE detected
by the 3 systems

agreement
NE not detected

disagreement =
ambiguous cases

previous context

current position

next context

-3 -2 -1 0 +1 +2 +3

le tour de france troisième étape remportée

ASR
CM

0.99 0.99 0.99 0.99 0.99 0.99 0.99

S
feat

_le NCMS _de NPSIG ADJFS NCFS VPARPFS

1,1 2,2 3,3 1,4 1,3 1,2 1,3

0.13 0.23 0.03 0.31 0.39 0.27 0.44

2 3 1 4 8 4 7

NE
feat

no yes yes yes no no no

no org org loc no no no

no org org loc no no no

no org org loc no no no

0.99 0.99 0.99 0.99 0.99 0.99 0.99



CRF-based combination



CM₋₃

CM₋₂

CM₋₁

CM₀

CM₊₁

CM₊₂

CM₊₃

Experiments

- Setup
 - ✓ Large vocabulary system [Huet'10]
 - A posteriori CM based on N-best lists = baseline
 - ✓ Corpus ESTER2
 - 12h French broadcast news
 - word error rate = 26.1%
 - ✓ Evaluation
 - 5-fold cross validation (80% train, 20% test)

- CM for error detection
 - ✓ on all words
 - ✓ on recognized NE

- CM for NE recognition

CM for error detection

CM	EER	Improvement
ASR	29.21	.
ASR+NE	28.17	-3.6%
ASR+S	24.18	-17.2%
ASR+S+NE	23.94	-18.0%

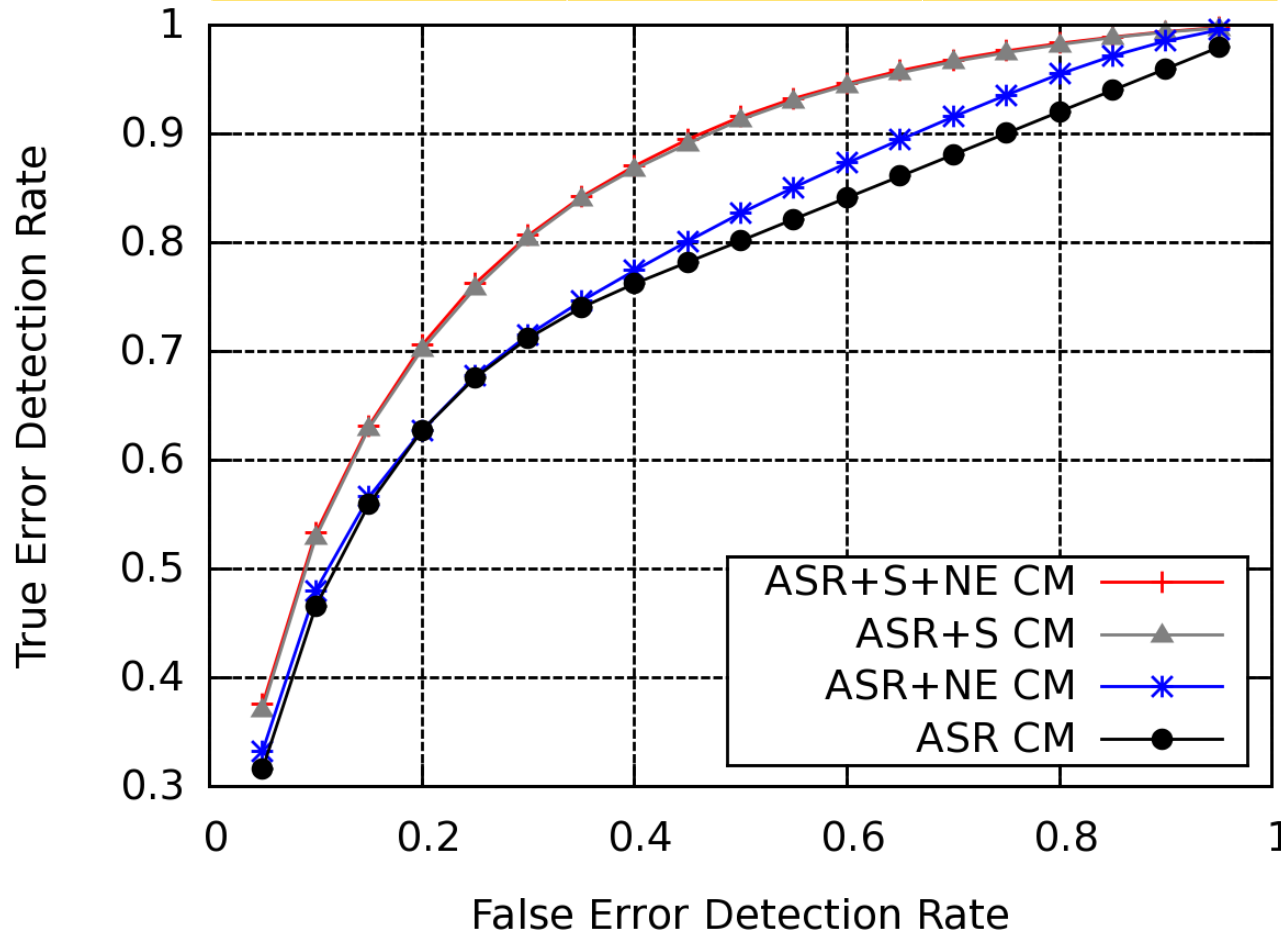
→ All features useful in combination

→ Significant improv. 18%

→ NE

✓ slight improv.

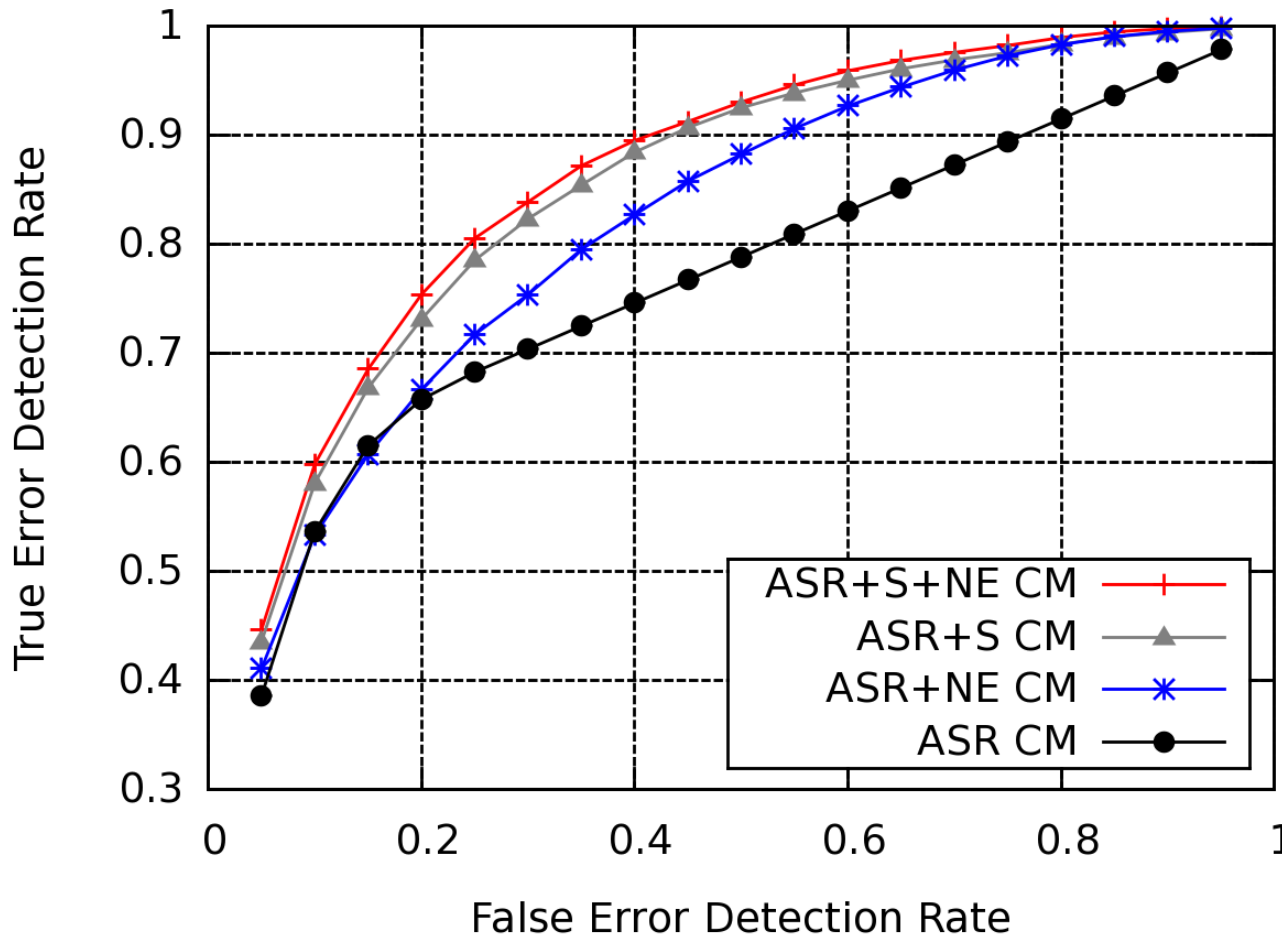
✓ recognized NE = only 17.6% of all words



Zoom on recognized named entities

CM	EER	Improvement
ASR	29.69	.
ASR+NE	24.89	-16.2%
ASR+S	22.27	-22.5%
ASR+S+NE	21.48	-27.7%

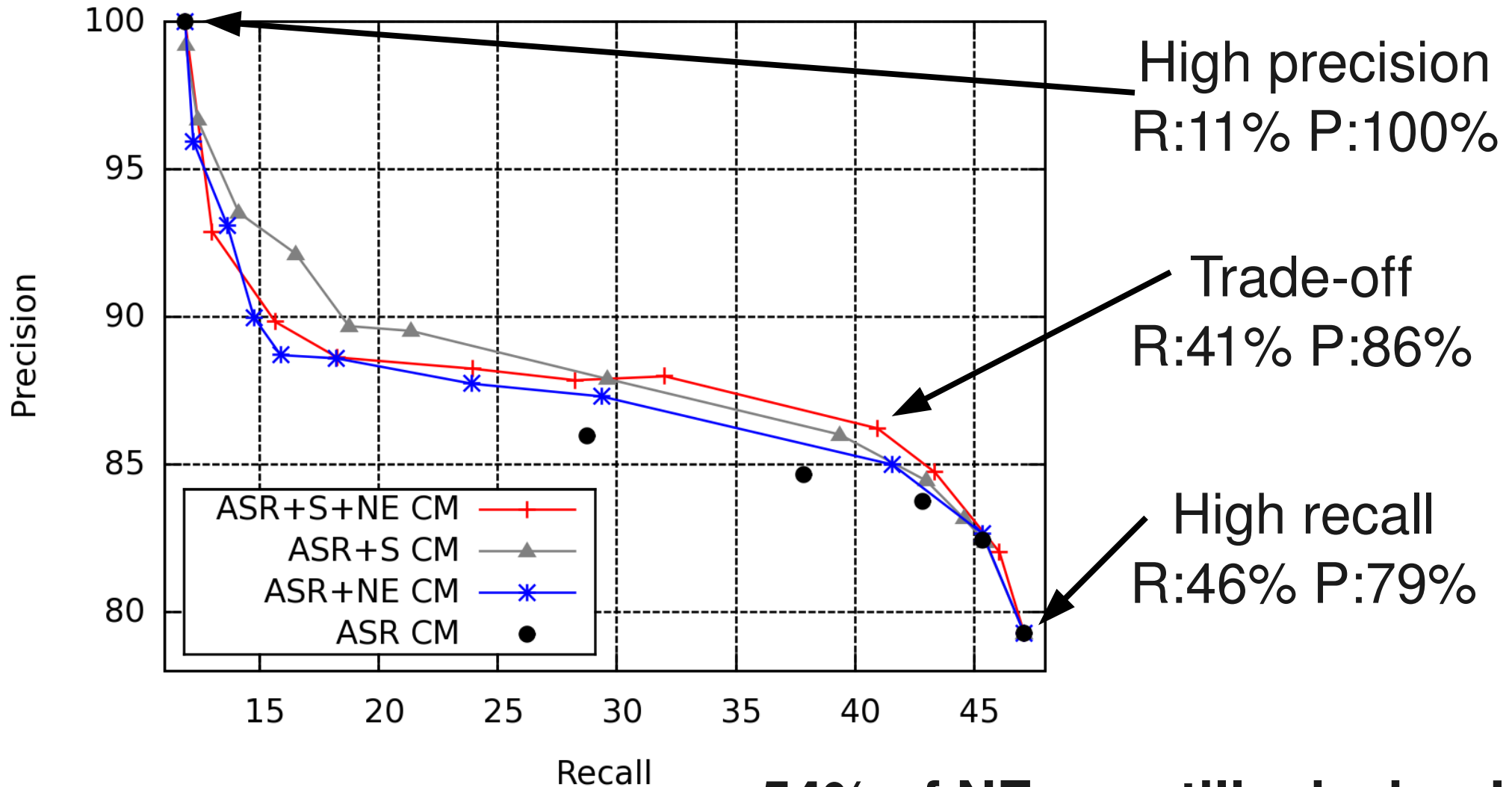
- Better CM estimation
- Significant improv. 28%



- S & NE feat
 - ✓ **redundant** on agreement cases (class 'yes') = 14% all words
 - ✓ **complementary** on disagreement cases ('maybe') = 3.6% all words !

CM for NE recognition

Score(rec NE) = mean(CMs)



54% of NE are still missing !

Conclusions

- An enriched CM for error detection and NER
 - ✓ Trade-off between noise and rich information depending on the application
 - ✓ Improvement in EER
 - **18%** on all words
 - **28%** on recognized NEs

→ Limitations

- ✓ NE feat useful for a small subset of words
- ✓ 54% of NE are still missing!

→ Future work

- ✓ process graphs for more information
- ✓ lexico-phonetic transcript

Thank you for your attention !
Questions ?

Future work



“le suisse **cancellara** reste maillot jaune”

Speech Recognition

→ Lexical

ERROR!

le suisse **quand c' est lara** reste maillot jaune

→ Phonetic

ka~t s E laRa

→ Reliable confidence measure

1 1 | 0 0 0 0 | 1 1 1