

# CRF-based Combination of Contextual Features to Improve A Posteriori Word-level Confidence Measures

**Julien Fayolle**, Fabienne Moreau,  
Christian Raymond, Guillaume Gravier, Patrick Gros

INRIA & IRISA, Rennes, France

INTERSPEECH 2010

# Motivation



“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

# A reliable confidence measure (CM)

## → Context

- ✓ Retrieving information in TV broadcast news
- ✓ Large Vocabulary Continuous Speech

## → A reliable CM

- ✓ at the word-level
  - correct / erroneous
- ✓ for any LVCSR system
  - no specific decoder needed
- ✓ on all transcribed words
  - no specific focus (e.g. OOV)

# State-of-the-art

System	Small Vocabulary	Large Vocabulary Continuous Speech	
		OOV	All words
Decoder CM	A posteriori CM N-best / graph	A posteriori CM N-best / graph	A posteriori CM N-best / graph
Phonetic	#syllables	.	dur / #ph
Syntactic	POS	.	POS
Linguistic	LMBB	LMBB	LMBB
Semantic	discourse	LSA	NE
Context	prev / next	prev / next	adapted
Machine learning combination	TBL SVM Adaboost	Adaboost	CRF

# Our approach

- Input: any decoder output (+decoder CM)
- Goal: (more) reliable CM
- 1 Multi-level features
  - ✓ obtainable from any decoder
- 2 Context
  - ✓ errors impact surrounded words
- 3 Machine-learning-based combination
  - ✓ dealing with sequential data

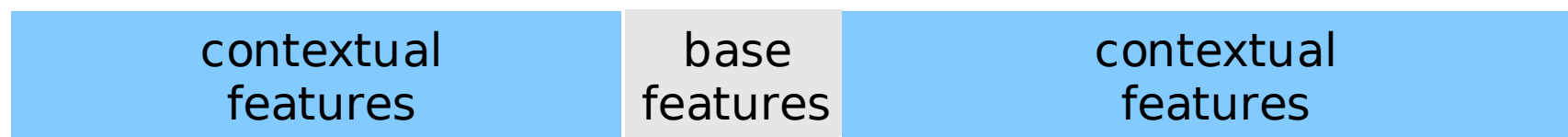
# Multi-level features

- Any decoder CM
  - ✓ **Conf** : a posteriori [Huet'10]
- Syntactic
  - ✓ **POS** : part-of-speech
- Linguistic
  - ✓ **LMBB** : language model back-off behavior
- Phonetic
  - ✓ **dur** : duration
  - ✓ **#ph** : number of phonemes
- Semantic
  - ✓ **NE** : named entities [Raymond'10]

# Context

- Errors impact surrounded words
- State-of-the-art
  - ✓ prev / next words systematically
  - ✓ specific features
- Our approach
  - ✓ Larger context
  - ✓ For all features
  - ✓ Well-suited for each feature

	previous context			current position	next context		
	-3	-2	-1	0	+1	+2	+3
	le	tour	de	france	troisième	étape	remportée
conf:	0.99	0.99	0.99	0.99	0.99	0.99	0.99
POS:	_le	NCMS	_de	NPSIG	ADJFS	NCFS	VPARPFS
LMBB:	1,1	2,2	3,3	1,4	1,3	1,2	1,3
dur:	0.13	0.23	0.03	0.31	0.39	0.27	0.44
#ph:	2	3	1	4	8	4	7
NE:	no	yes	yes	yes	no	no	no



CM<sub>-3</sub>

CM<sub>-2</sub>

CM<sub>-1</sub>

CM<sub>0</sub>

CM<sub>+1</sub>

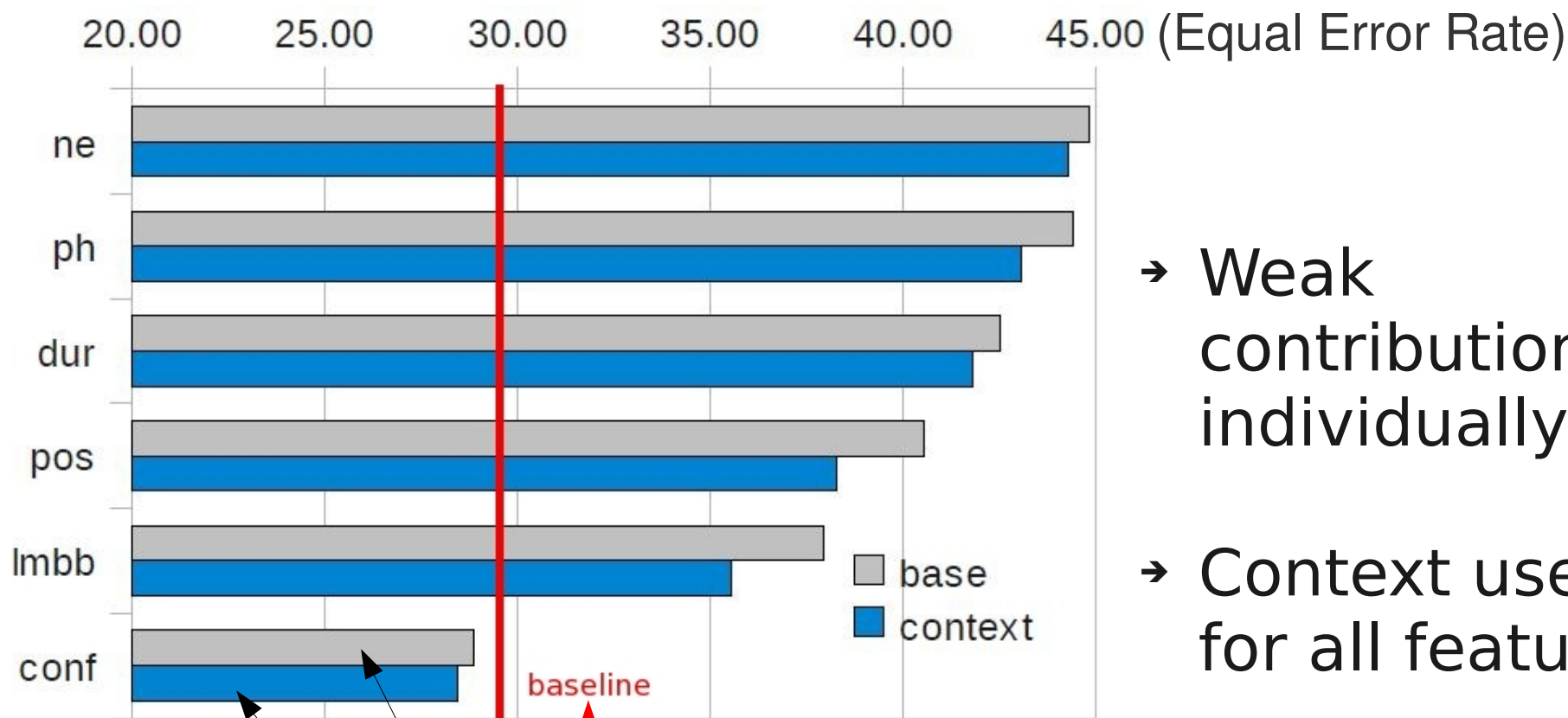
CM<sub>+2</sub>

CM<sub>+3</sub>

# Experimental setup

- LVCSR system [Huet'10]
  - ✓ A posteriori CM based on N-best lists
  - ✓ Baseline
- Corpus ESTER2 [Galliano'09]
  - ✓ 12h french broadcast news (dev+test)
  - ✓ 125k words
  - ✓ WER = 26.1%
- Evaluation
  - ✓ 5-fold cross-validation
    - 80% for training
    - 20% for testing
  - ✓ Equal Error Rate
- Experiments
  - ✓ Individual features
  - ✓ Combination
  - ✓ Model analysis

# Individual features



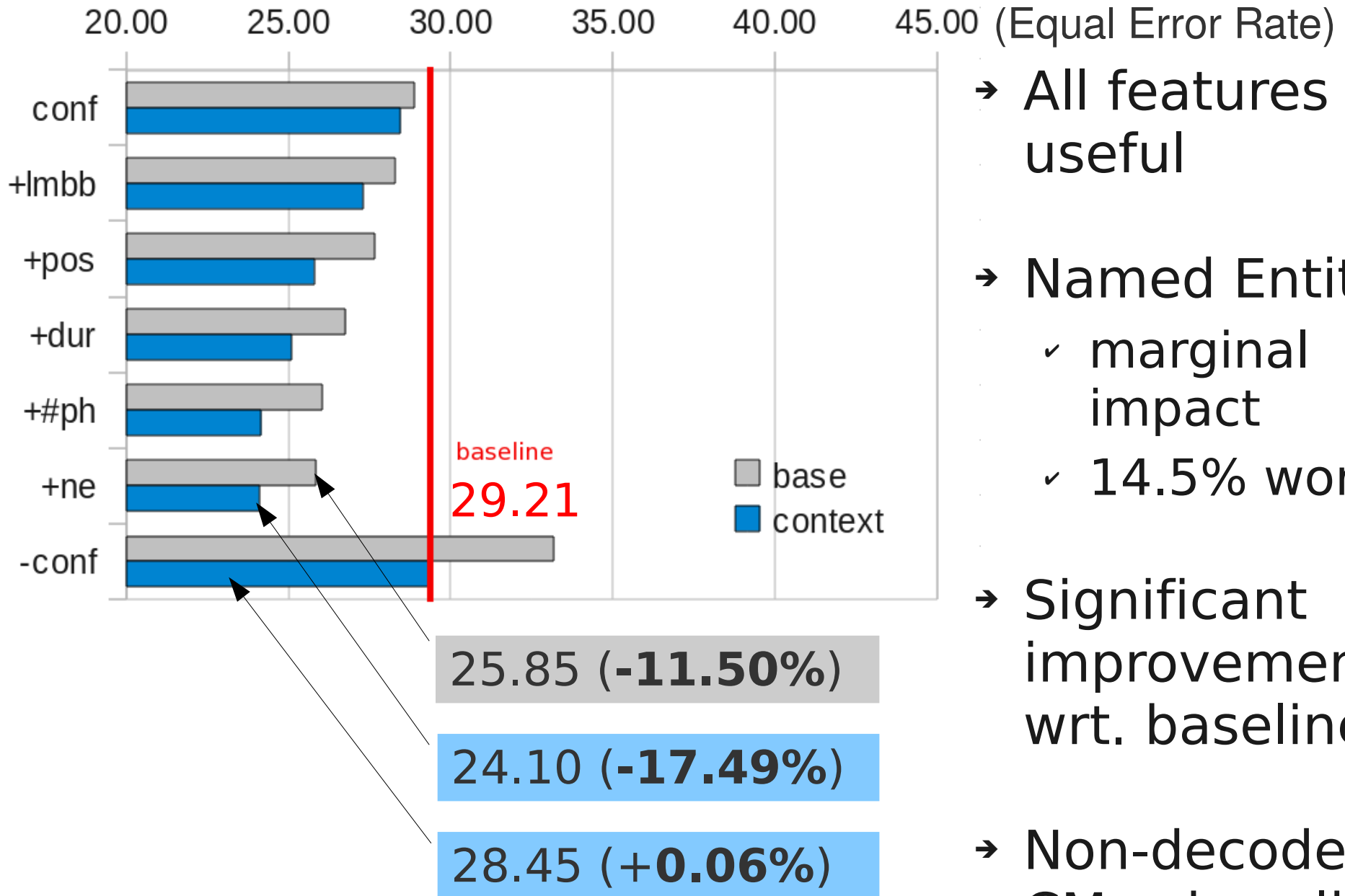
baseline  
29.21

28.87 (-1.16%)

28.45 (-2.67%)

- Weak contribution individually
- Context useful for all features
- Decoder CM as a feature ↗

# Combination



- All features useful
- Named Entities
  - ✓ marginal impact
  - ✓ 14.5% words
- Significant improvement wrt. baseline
- Non-decoder CM = baseline

# Model analysis

useful context = best weighted positions

	previous context			current position	next context		
	-3	-2	-1	0	+1	+2	+3
	le	tour	de	france	troisième	étape	remportée
conf:	0.99	0.99	0.99	0.99	0.99	0.99	0.99
POS:	_le	NCMS	_de	NPSIG	ADJFS	NCFS	VPARPFS
LMBB:	1,1	2,2	3,3	1,4	1,3	1,2	1,3
dur:	0.13	0.23	0.03	0.31	0.39	0.27	0.44
#ph:	2	3	1	4	8	4	7
NE:	no	yes	yes	yes	no	no	no

→ Useful context different for each feature

# Model analysis

## → Typical errors

- ✓ Low conf
- ✓ Specific POS classes
- ✓ Low LMBB (1-gram)
- ✓ Short words (OOV)
  - duration
  - #ph
- ✓ Disagreement between the 3 NE taggers

## → Typical correct words

- ✓ High conf
- ✓ Specific POS classes
- ✓ High LMBB (4-gram)
- ✓ Long words
  - duration
  - #ph
- ✓ Agreement between the 3 NE taggers

# Conclusions

## → Combination

- ✓ syntactic, linguistic, phonetic, semantic

## → If available decoder CM

- ✓ improv of **17.5%** EER

## → Else

- ✓ non-decoder CM  
= decoder CM

## → Context

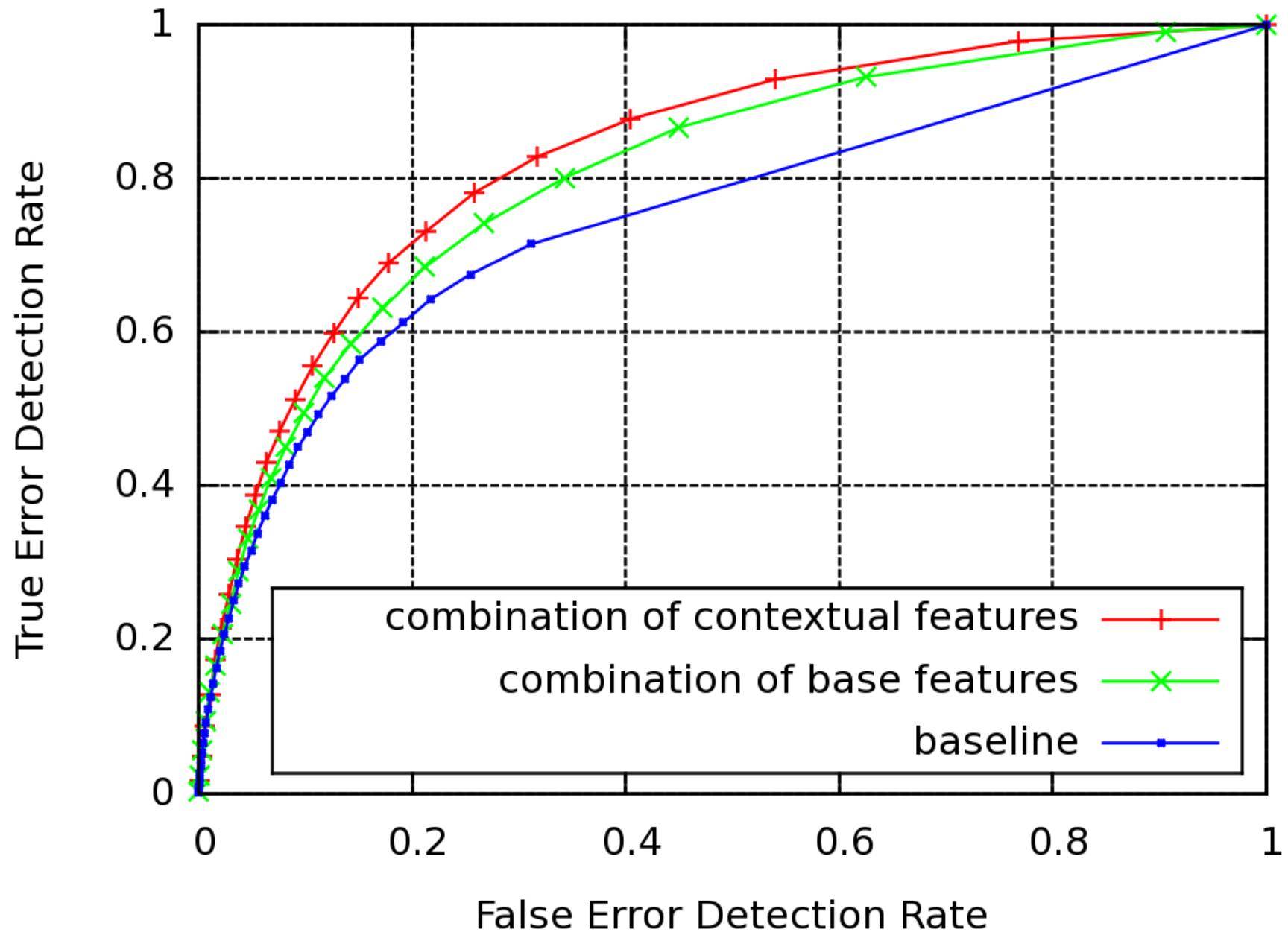
- ✓ useful for all features
- ✓ different for each feature

## → Future works

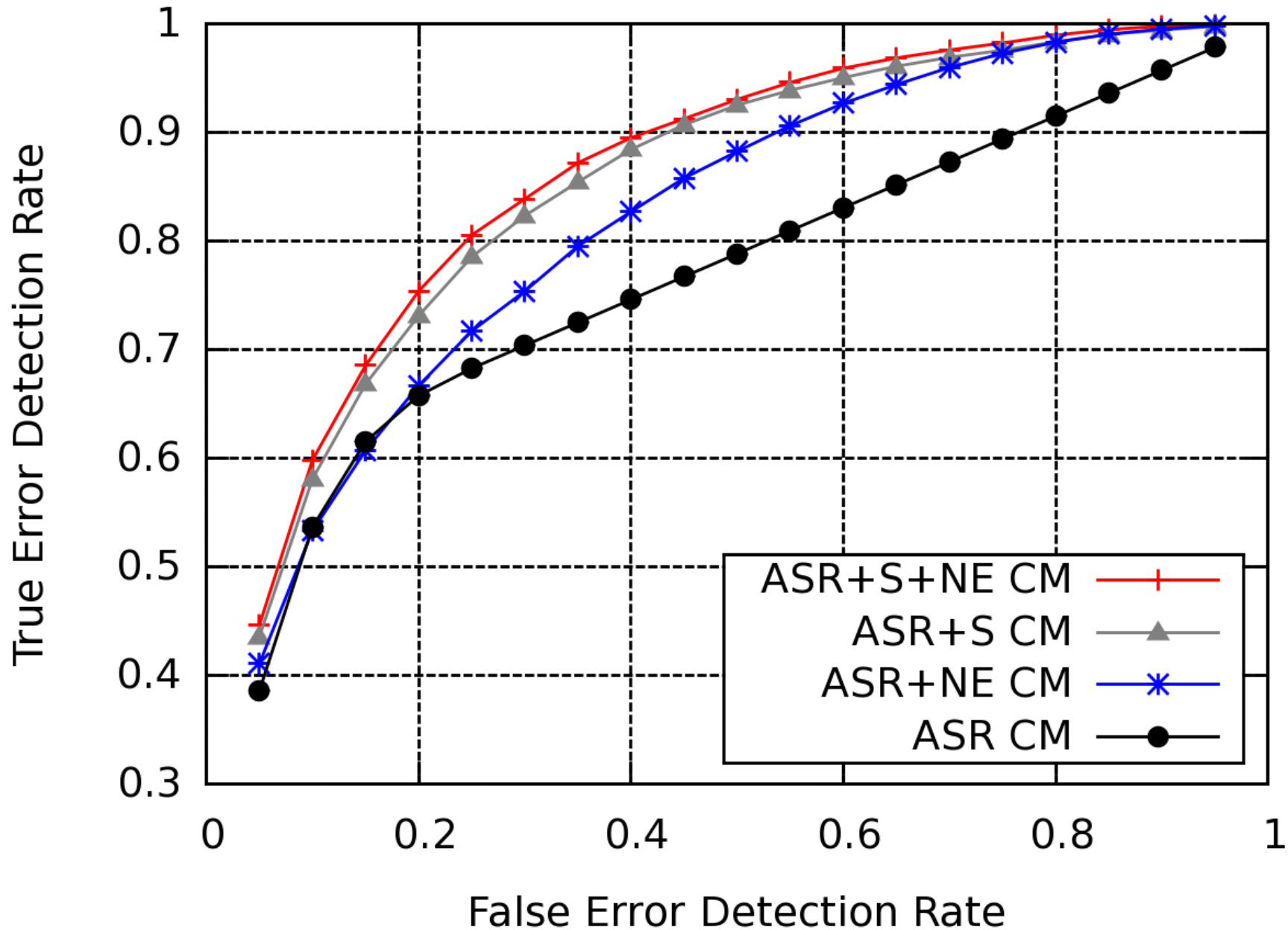
- ✓ semantic features
- ✓ lexico-phonetic hybrid transcription

**Thank you for your attention !**  
**Questions ?**

# Combination



# ROC curve on recognized named entities



# 3 CRF-based combination

- Conditional Random Fields [Lafferty'01]
  - ✓ Dedicated to manage sequential data
  - ✓ Binary classification : 'correct' or 'erroneous'
  - ✓ **Y = sequence of labels** to be predicted
  - ✓ **X = sequence of features** observed
  - ✓ at position  $i$  in the sequence,

$$CM_i = p(Y_i = 'correct' / X)$$

- Machine learning model
  - ✓ each feature is weighted
  - ✓ => explainable model