An Investigation of Code-Switching Attitude Dependent Language Modeling

Ngoc Thang Vu, Heike Adel, Tanja Schultz

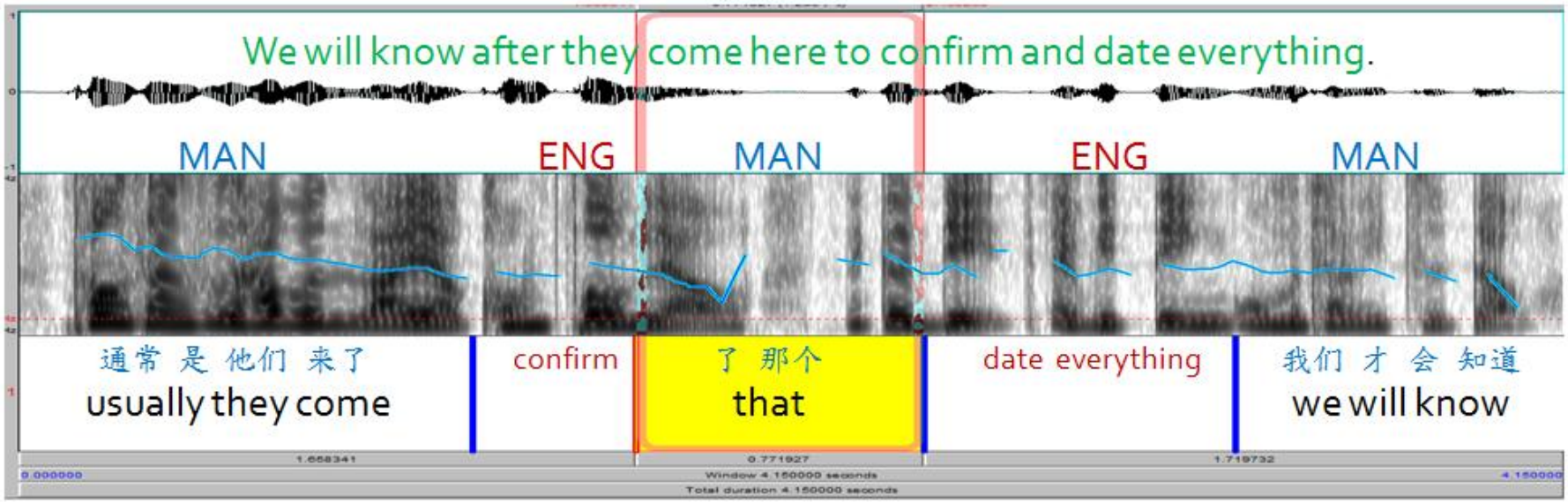
SLSP 2013 in Tarragona, Spain
July 30th 2013

Outline

- Introduction
- Corpus: SEAME
- Code-Switching Prediction using Part-of-Speech Tags
  - Speaker Independent Analysis
  - Speaker Dependent Analysis
- Code-Switching Attitude Dependent Language Modeling
  - Code-Switching Attitude Clustering
  - N-Gram Model
  - Recurrent Neural Network Language Model
  - Decoding Experiments
- Conclusion

Introduction

- Code-Switching = speech containing more than one language
- Exists in multilingual communities or among immigrants
- Example:



Introduction

- Language Modeling for Code-Switching speech: challenging due to the multilingualism and few training data
- Decision whether and when a speaker changes may be individual
- Definition:

  Code-Switching attitude: Code-Switching behavior of a speaker
- => adapted models may lead to better results

Corpus: SEAME

- SEAME = South East Asia Mandarin-English
- Recorded from Singaporean and Malaysian speakers by [1]
- Originally used for the research project ‘Code-Switch‘ (Nanyang Technological University (NTU) and Karlsruhe Institute of Technology (KIT))
- 63 hours of transcribed audio data

- [1] Lyu, D.C. et al.: An analysis of Mandarin-English Codeswitching speech corpus: SEAME , ICSLP 2010

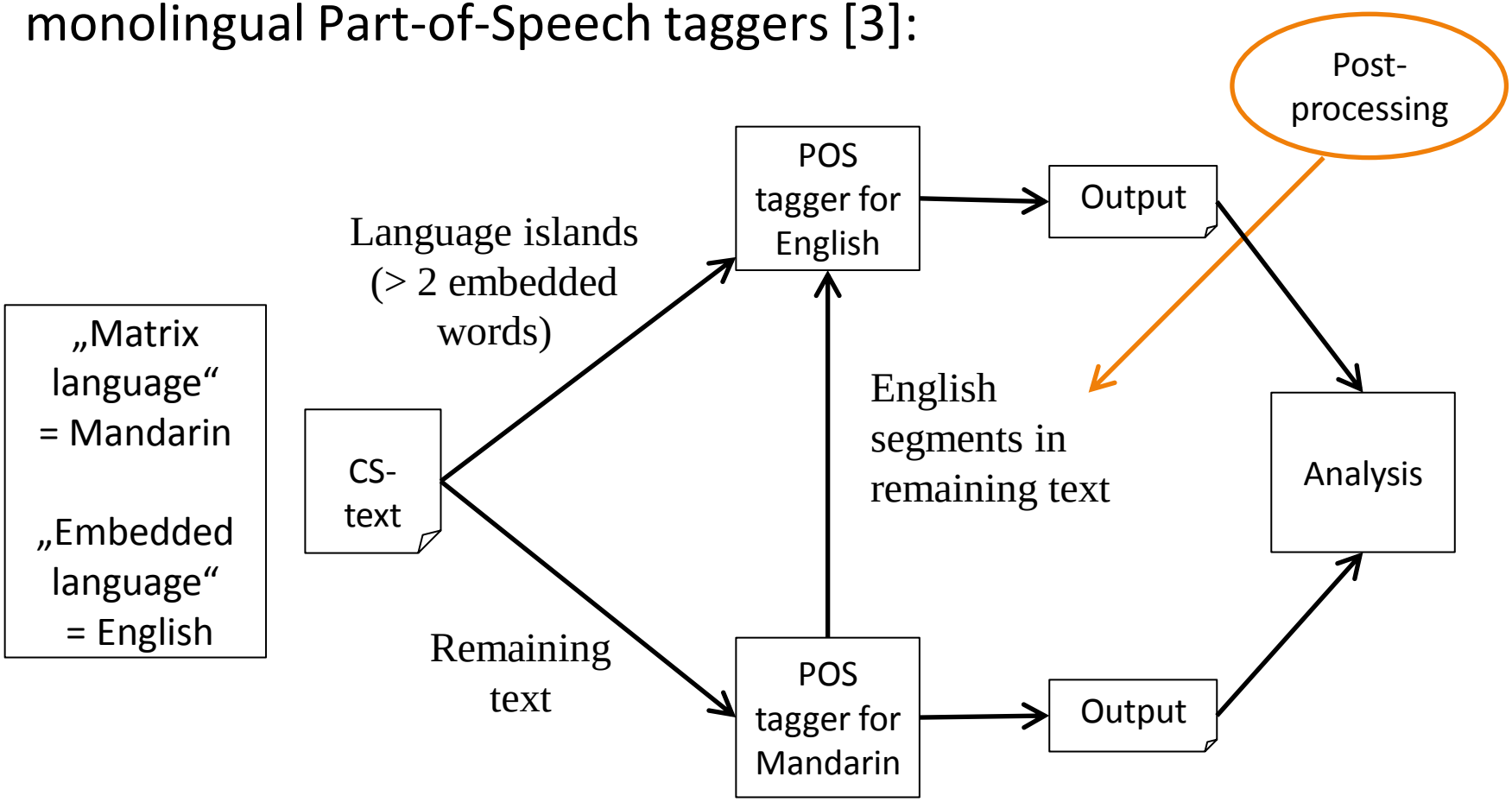
Analysis of the Corpus

- High Code-Switching rate:
  - Average number of code switches: 2.6 per utterance
  - Short monolingual segments: 0.67 seconds (English), 0.81 seconds (Mandarin)
- Important trigger events for Code-Switching [2][3]:
  - Words
  - Part-of-speech (POS) tags
- => Due to few training data, we concentrate on POS tags

- [2] Poplack, S.: Sometimes i'll start a sentence in spanish y termino en español: toward a typology of code-switching, 1980.
- [3] Schultz, T. et al.: Detecting code-switch events based on textual features, 2010.

Code-Switching Prediction using POS tags

- Part-of-Speech tagger: combination of English and Mandarin monolingual Part-of-Speech taggers [3]:



- [3] Schultz, T. et al.: Detecting code-switch events based on textual features, 2010.

Speaker Independent Analysis

- Top 5 Part-of-speech tags sorted by their Code-Switching rate

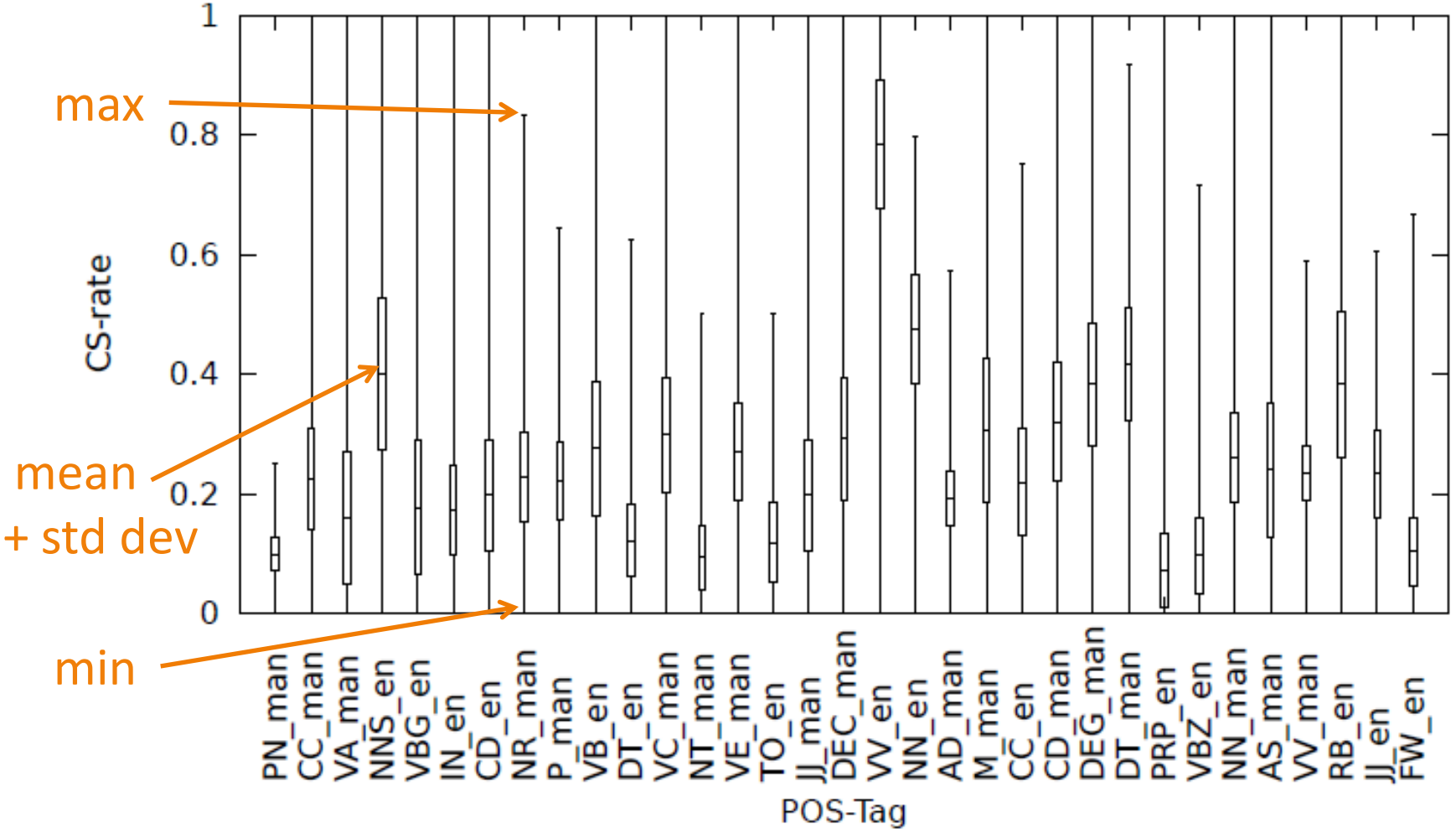
$$\text{CS}-\text{rate}\,(\text{i}) = \frac{\text{frequency}\,(\text{i CS}-\text{point})}{\text{frequency}\,(\text{i})}$$

Tag	meaning	frequency	CS-rate	
DT	determiner	11276	40.44%	From Mandarin to English
DEG	associative 的	4395	36.91%	
MSP	other particle	507	32-74%	
VC	是	6183	25.85%	
DEC	的 in a relative-clause	5763	23.86%	
NN	noun	49060	49.07%	From English to Mandarin
NNS	noun (plural)	4613	40.82%	
RP	particle	330	36.06%	
RB	adverb	21096	31.84%	
JJ	adjective	10856	26.48%	

- POS tags: important trigger events, but CS-rates less than 50%
  => motivation for speaker dependent analysis!

Speaker Dependent Analysis

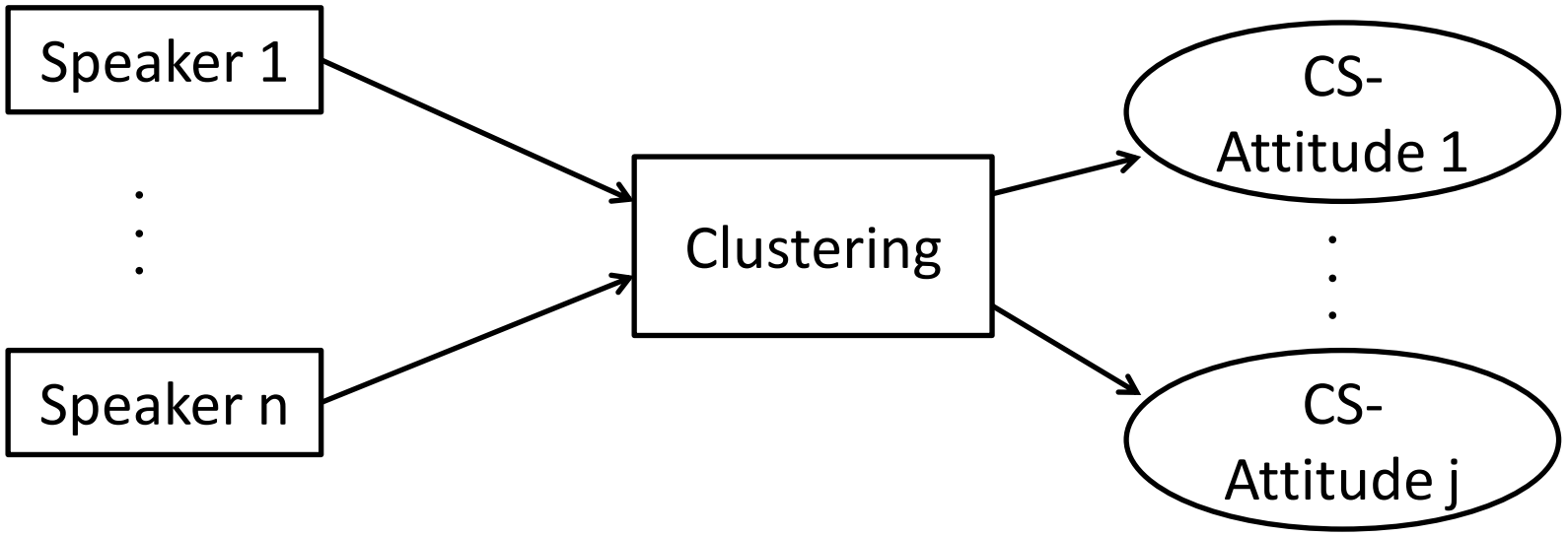
- Calculation of CS-rates per speaker



- => high spread between min and max values

Code-Switching Attitude Clustering

- Goal: Cluster training transcriptions into different CS attitudes



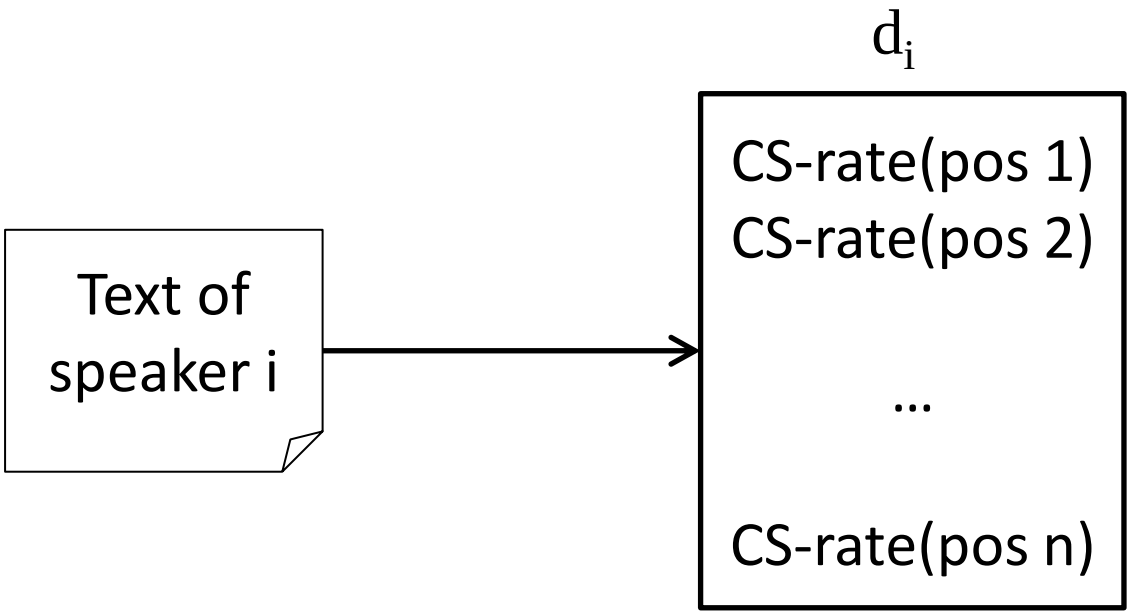
Code-Switching Attitude Clustering

- Text-based clustering using k-means and cosine-distance:

$$Dist(d_1, d_2) = \frac{d_1 \cdot d_2}{\| d_1 \| \cdot \| d_2 \|}$$

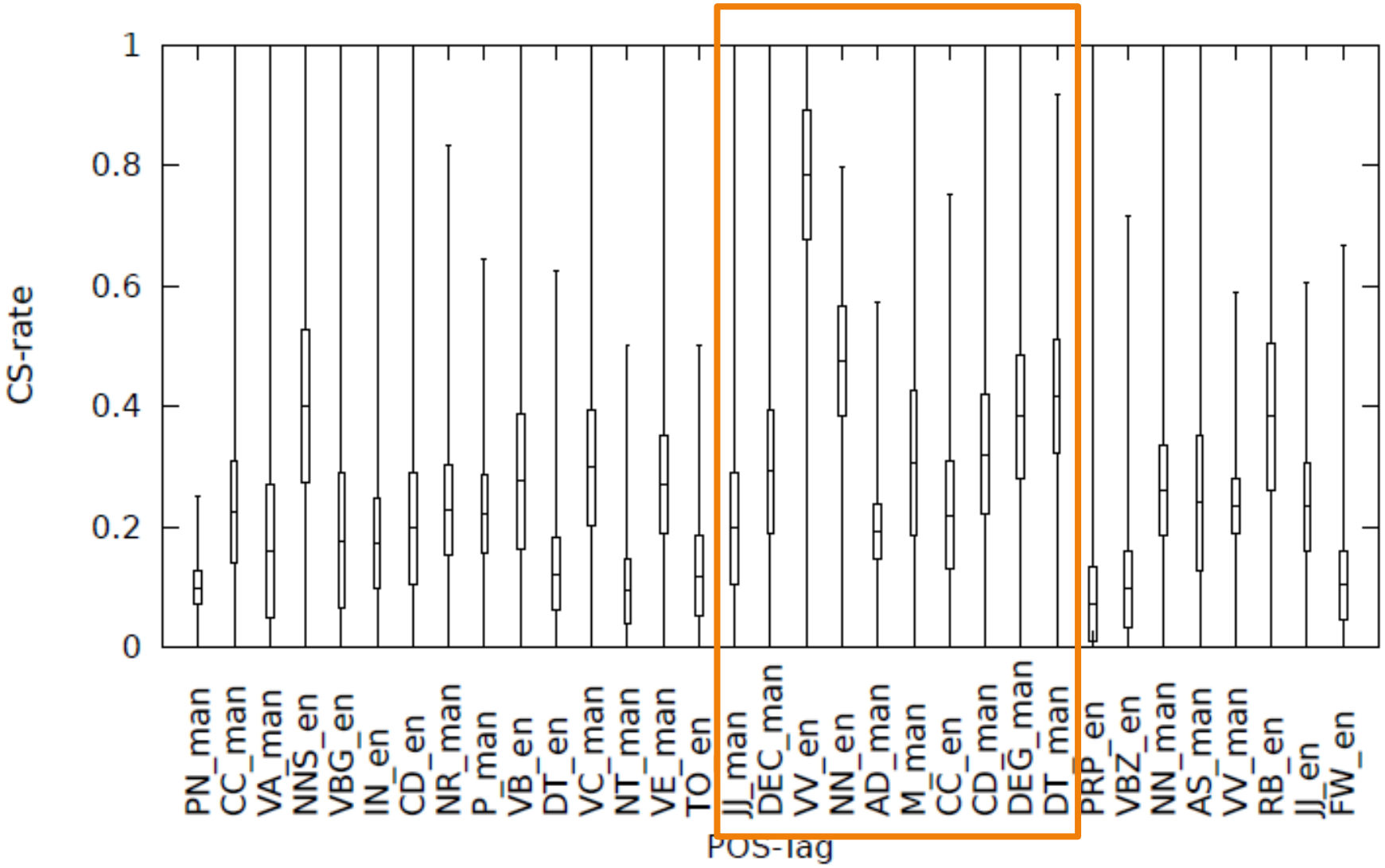
(d_i: vector of speaker i)

Since the CS attitude should be modelled, d_i is defined as follows:



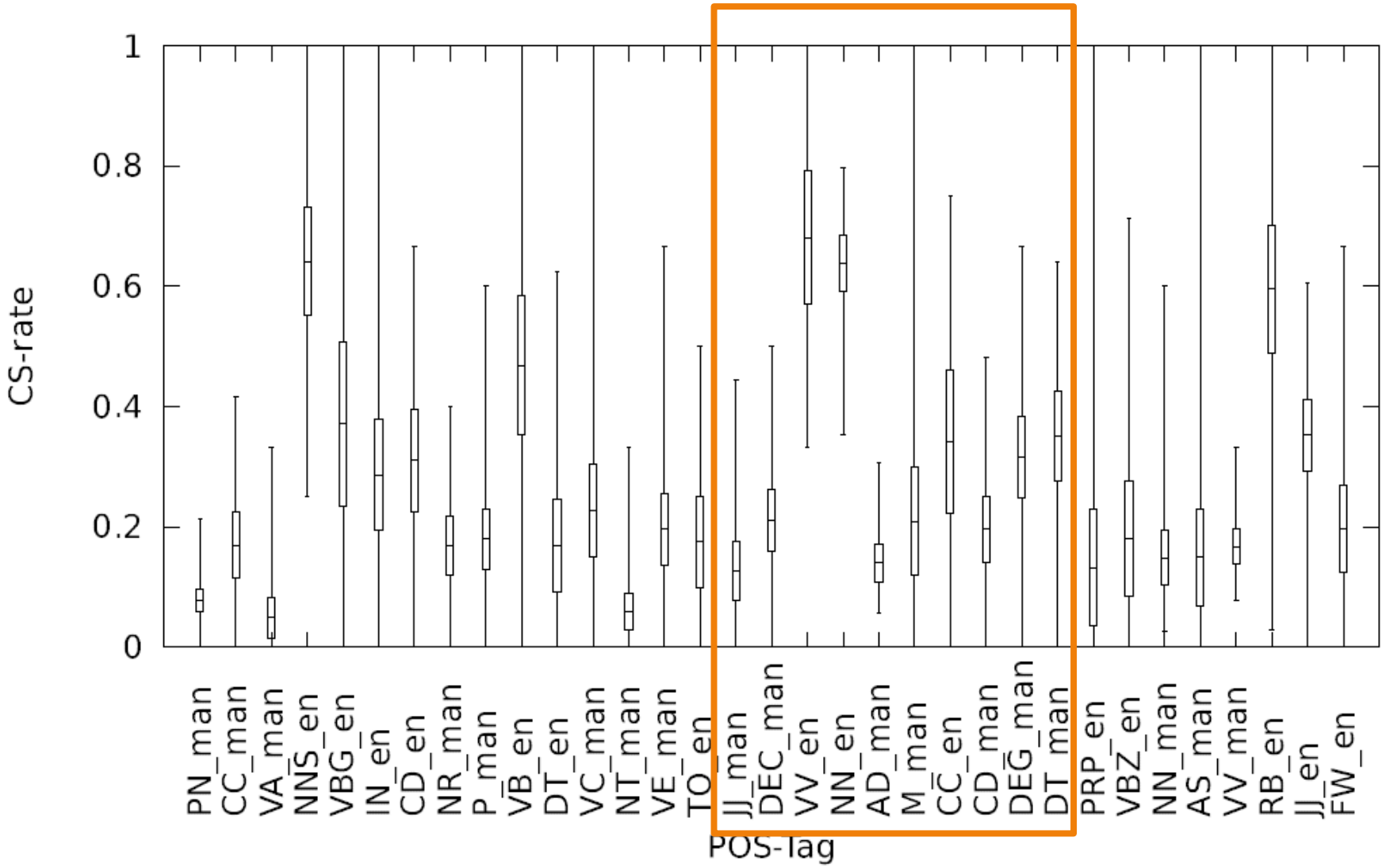
Text clustering: Result

- Speaker dependent analysis before the clustering process:



Text clustering: Result

- Speaker dependent analysis after the clustering process:



N-Gram Language Modeling

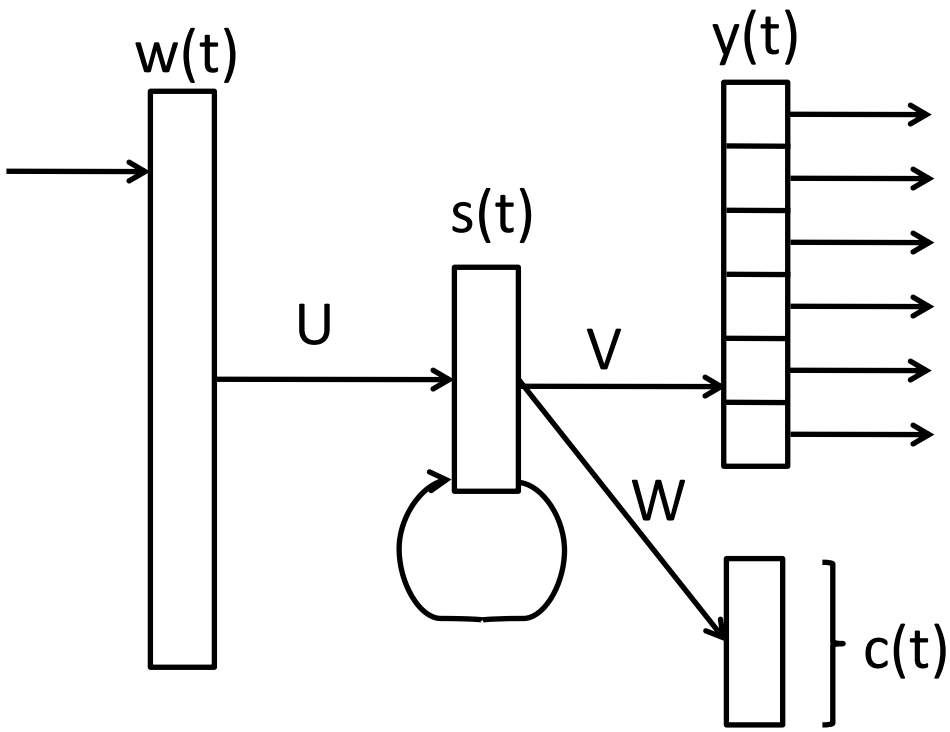
- Trigram model, built with the SRILM toolkit [4]
- Data: SEAME training transcriptions
- Adaptation: assigning higher weights to the texts of the different classes
- Perplexity results on the evaluation set:

	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8
Baseline	257.67	236.62	228.64	197.40	382.64	330.20	**358.22**	298.77
Adapted	**246.37**	**228.08**	**220.43**	**187.72**	**356.18**	**307.99**	358.97	**280.71**
Relative gain	4.39%	3.61%	3.59%	4.90%	6.92%	6.73%	-0.21%	6.04%

- [4] Stolcke, A. et al.: SRILM – an extensible language modeling toolkit, Interspeech 2002.

Recurrent Neural Network Language Modeling

- Traditional Model [5,6]



Input:
w(t): word vector

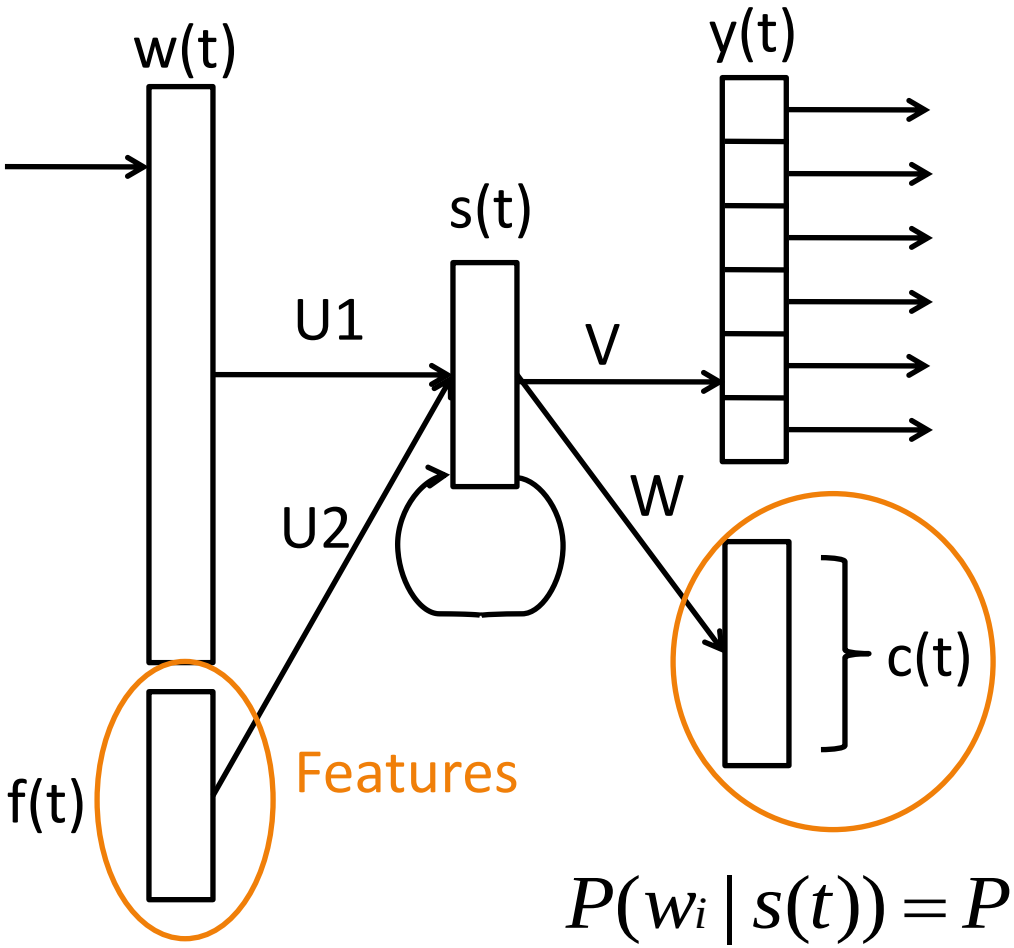
Hidden layer:
s(t)

Output:
y(t): word probabilities
c(t): class probabilities

- [5] Mikolov, T.: Recurrent neural network based language model, Interspeech 2010.
- [6] Mikolov, T.: Extensions of recurrent neural network language model, ICASSP 2011.

Recurrent Neural Network Language Modeling

- Model as proposed at ICASSP 2013 [7]:



Input:
w(t): word vector
f(t): feature vector

Hidden layer:
s(t)

Output:
y(t): word probabilities
c(t): class probabilities
classes = languages

$$P(w_i \mid s(t)) = P(c_i \mid s(t)) \cdot P(w_i \mid c_i, s(t))$$

- [7] Adel, Vu et al.: Recurrent neural network language modeling for Code-Switching conversational speech

Recurrent Neural Network Language Modeling

- Adaptation: one-iteration retraining using the texts of the different classes

- Perplexity results on the evaluation set

	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8
Baseline	200.66	181.60	187.04	174.13	364.59	275.89	286.31	256.99
Adapted	**197.74**	**175.85**	**170.92**	**160.58**	**327.33**	**253.67**	**286.30**	**241.69**
Relative gain	1.46%	3.17%	8.62%	7.78%	10.22%	8.05%	0.003%	5.95%

Comparison of the Language Models

- Perplexity results on the evaluation set

	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8
(1) N Gram	257.67	236.62	228.64	197.40	382.64	330.20	358.22	298.77
(2) Adapted N Gram	246.37	228.08	220.43	187.72	356.18	307.99	358.97	280.71
Relative gain (1-2)	**4.39%**	**3.61%**	**3.59%**	**4.90%**	**6.92%**	**6.73%**	-0.21%	**6.04%**

Comparison of the Language Models

- Perplexity results on the evaluation set

	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8
(1) N Gram	257.67	236.62	228.64	197.40	382.64	330.20	358.22	298.77
(2) Adapted N Gram	246.37	228.08	220.43	187.72	356.18	307.99	358.97	280.71
Relative gain (1-2)	**4.39%**	**3.61%**	**3.59%**	**4.90%**	**6.92%**	**6.73%**	-0.21%	**6.04%**
(3) RNNLM trad.	235.71	214.47	224.35	217.63	496.61	366.03	320.68	294.36
(4) RNNLM + feat + LID	200.66	181.60	187.04	174.13	364.59	275.89	286.31	256.99
Relative gain (1-4)	**22.13%**	**23.25%**	**18.19%**	**11.79%**	**4.72%**	**16.45%**	**20.07%**	**13.98%**

Comparison of the Language Models

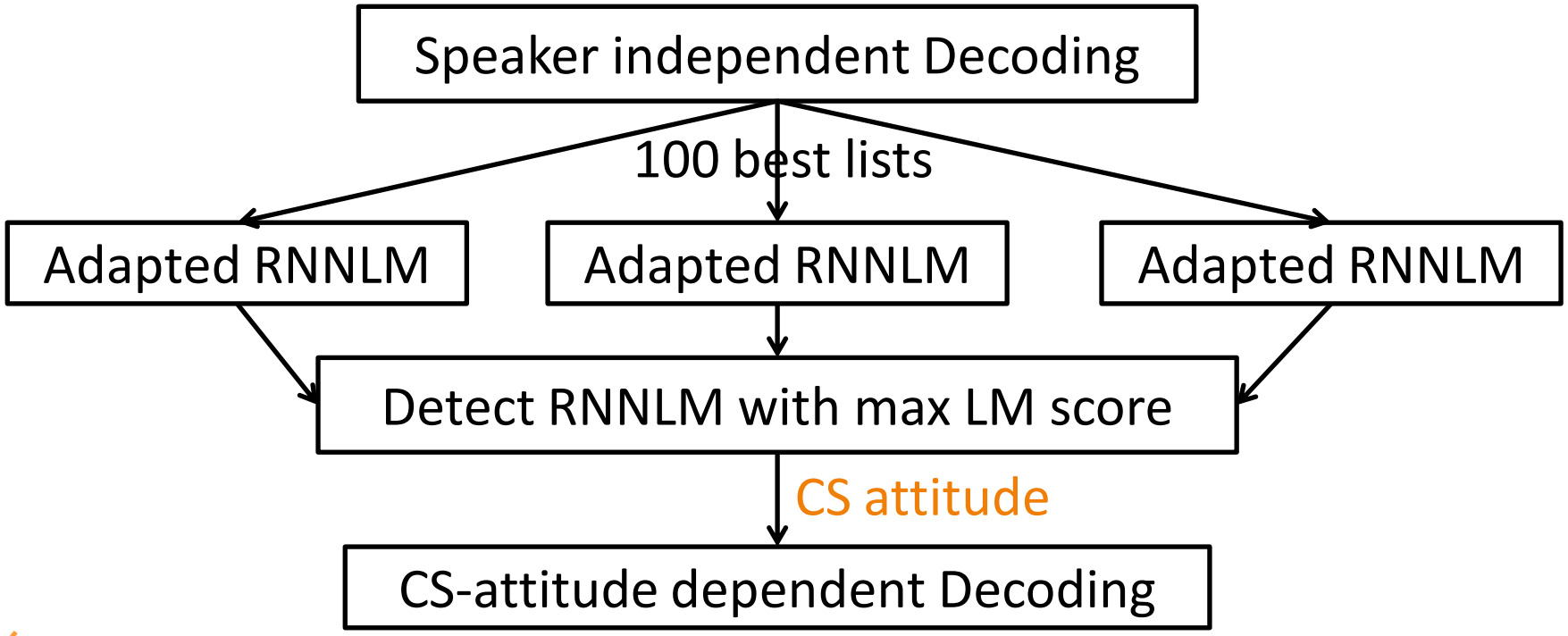
- Perplexity results on the evaluation set

	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8
(1) N Gram	257.67	236.62	228.64	197.40	382.64	330.20	358.22	298.77
(2) Adapted N Gram	246.37	228.08	220.43	187.72	356.18	307.99	358.97	280.71
Relative gain (1-2)	**4.39%**	**3.61%**	**3.59%**	**4.90%**	**6.92%**	**6.73%**	-0.21%	**6.04%**
(3) RNNLM trad.	235.71	214.47	224.35	217.63	496.61	366.03	320.68	294.36
(4) RNNLM + feat + LID	200.66	181.60	187.04	174.13	364.59	275.89	286.31	256.99
Relative gain (1-4)	**22.13%**	**23.25%**	**18.19%**	**11.79%**	**4.72%**	**16.45%**	**20.07%**	**13.98%**
(5) Adapted RNNLM	197.74	175.85	170.92	160.58	327.33	253.67	286.30	241.69
Relative gain (4-5)	**1.46%**	**3.17%**	**8.62%**	**7.78%**	**10.22%**	**8.05%**	**0.003%**	**5.95%**

Decoding Experiments

- Decoding of CS speech using the N-Gram model
- Rescoring of 100 best lists using the RNNLM

- Decoding Process with the adapted models without prior knowledge about the speakers:



Decoding Experiments

- Performance measure: Mixed Error Rate:
  - Word error rate for English segments
  - Character error rate for Mandarin segments

- Mixed Error rate results:

Model	Development set	Evaluation set
(1) SI N-Gram model (Baseline)	35.5 %	30.0 %
(2) SI N-Gram model + SI RNNLM	34.74 %	29.23 %
(3) Adapted N-Gram model + adapted RNNLM	**34.47 %**	**28.89 %**
Relative gain (1-3)	2.9 %	3.7 %

Conclusion

- Code-Switching is a speaker dependent phenomenon
  => adapted Language Models outperform speaker independent Language Models
- Clustered similar Code-Switching attitudes using k-means and cosine-distances
- Trained N-Gram and recurrent neural network language models and adapted them to Code-Switching attitudes
- Improvements in terms of perplexity and mixed error rate

Thank you for your attention!