Word Embeddings combination and neural network for robustness in ASR error detection

Sahar Ghannay, Yannick Estève, Nathalie Camelin

LIUM, University of Le Mans, France

EUSIPCO 2015 Nice, The 2015 European Signal Processing Conference

03/09/2015

Introduction

MGB 2015 challenge results for ASR task on BBC data

	Best Sys	CRIM/ LIUM	Sys1	Sys2	Sys3	LIUM	Sys4	Sys5	Sys6	Sys7	Sys8	Sys9
Overall WER(%)	**23.7**	26.6	27.5	27.8	28.8	30.4	30.9	31.2	35.5	38.0	38.7	40.8

Introduction

MGB 2015 challenge result

Detailed performance of the best system

Show	CU
Daily Politics	10.4
Magnetic North	11.6
Dragons'Den	11.5
Eggheads	14.1
Athletics London	14.7
Point of View	13.5
Syd Barrett	21.3
Top Gear	21.8
Blue Peter	24.6
Legend of the Dragon	21.7
The North West 200	27.7
Holby City	32.1
The Wall	33.7
One Life Special Mum	35.3
Goodness Gracious ME	37.2
Oliver Twist	**41.4**
Overall WER(%)	***23.7***

Introduction

ASR errors have impact on applications:

- ✤ Information retrieval
- ✤ Speech to speech translation
- ✤ Spoken language understanding
- ✤ etc.

Introduction

ASR errors have impact on applications:

- Information retrieval
- Speech to speech translation
- Spoken language understanding
- etc.

⇨ ASR error detection can help

Introduction

✓ Related work

- Approaches based on Conditional Random Field (CRF)
  - OOV detection [C. Parada *et al.* 2010]
    - contextual informations
  - Errors detection [F. Béchet & B. Favre 2013]
    - ASR based, lexical and syntactic informations
- Approach based on neural network
  - Errors detection [T. Yik-Cheung *et al.* 2014]
    - complementary ASR systems

Introduction

✓ Related work

- Approaches based on Conditional Random Field (CRF)
  - OOV detection [C. Parada *et al.* 2010]
    - contextual informations
  - Errors detection [F. Béchet & B. Favre 2013]
    - ASR based, lexical and syntactic informations
- Approach based on neural network
  - Errors detection [T. Yik-Cheung *et al.* 2014]
    - complementary ASR systems

✓ Contributions

- Neural approach
  - Effective word embeddings combination
  - New neural architectures

Set of features

The features are inspired by [F. Béchet and B. Favre 2013]

- ✤ Posterior probabilities
- ✤ Lexical features
  - word length
  - existence 3-gram
- ✤ Syntactic features
  - POS tag
  - dependency labels
  - word governors
- ✤ Word

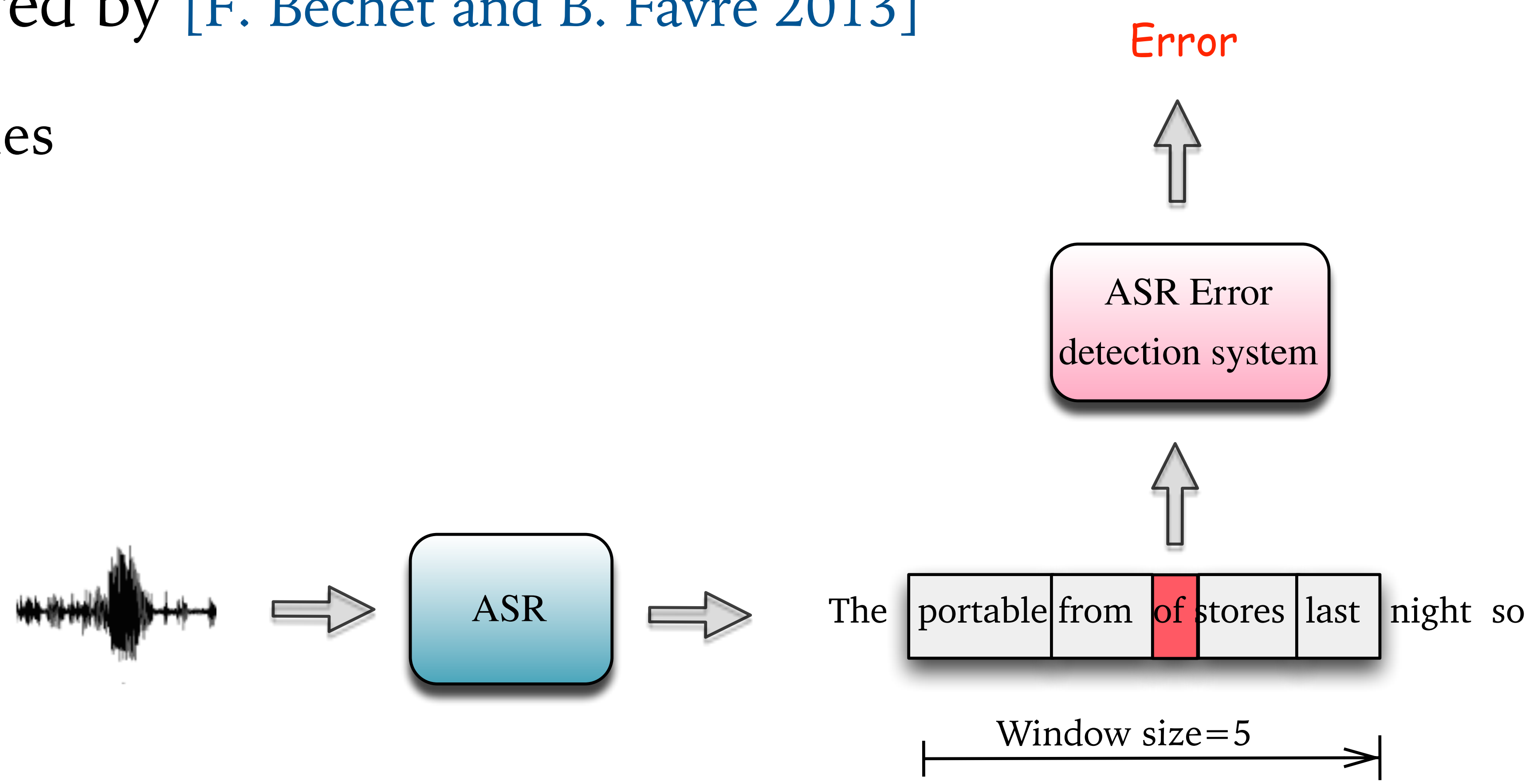


Figure 1: ASR error detection system

Set of features

The features are inspired by [F. Béchet and B. Favre 2013]

- Posterior probabilities
- Lexical features
  - word length
  - existence 3-gram
- Syntactic features
  - POS tag
  - dependency labels
  - word governors
- Word → Word embeddings

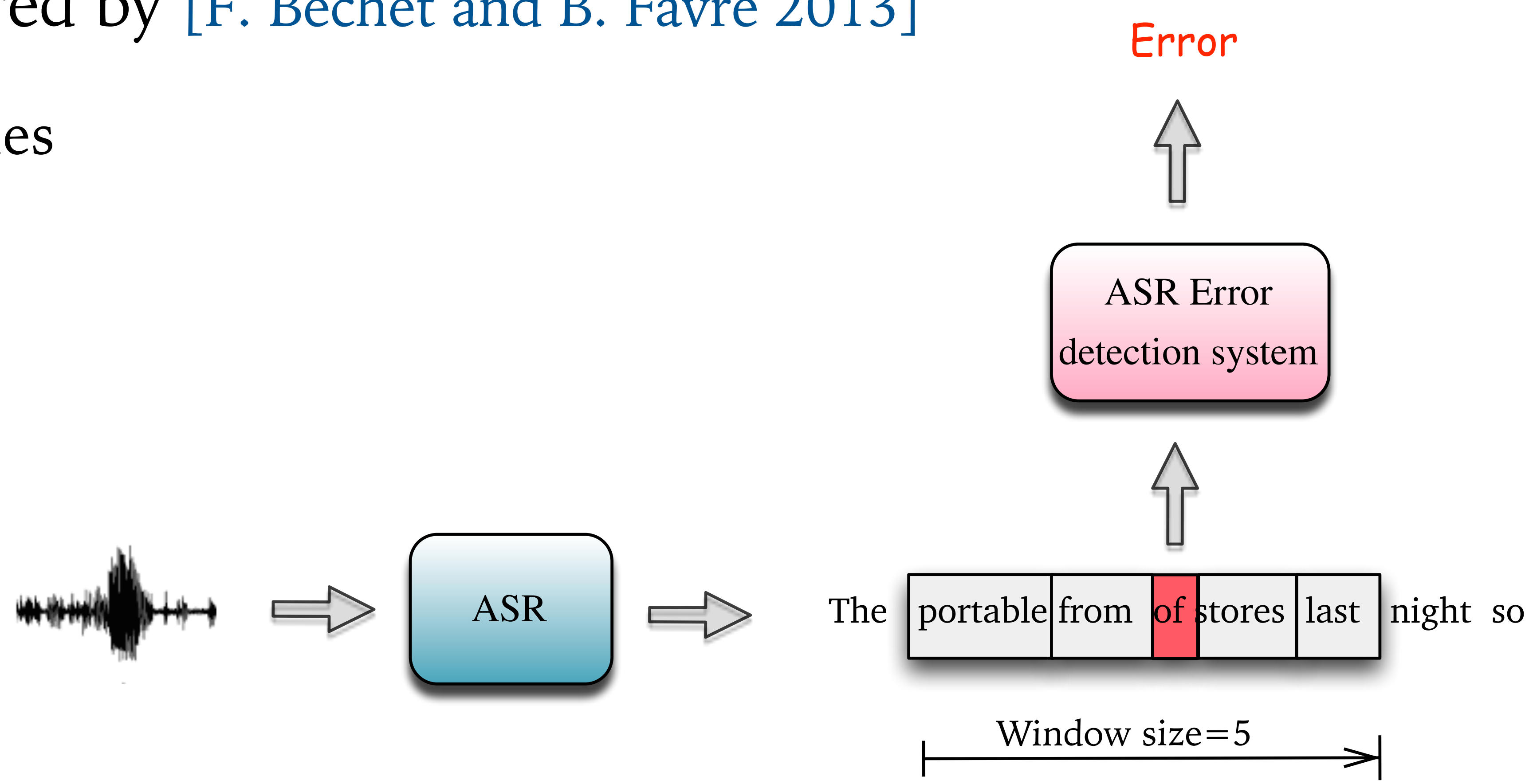


Figure 1: ASR error detection system

Word embeddings

Mapping words to high-dimensional vectors (e.g. 200 dimensions)

$$R : Words = \{W_1, ..., W_n\} \rightarrow Vectors = \{R(W_1), ..., R(W_n)\} \subset R^d$$

Distance between vectors indicates the relation between words

$$R(W_1) \approx R(W_n) \rightarrow W_1 \approx W_n$$

Word embeddings

Mapping words to high-dimensional vectors (e.g. 200 dimensions)

$$R : Words = \{W_1, ..., W_n\} \rightarrow Vectors = \{R(W_1), ..., R(W_n)\} \subset R^d$$

Distance between vectors indicates the relation between words

$$R(W_1) \approx R(W_n) \rightarrow W_1 \approx W_n$$

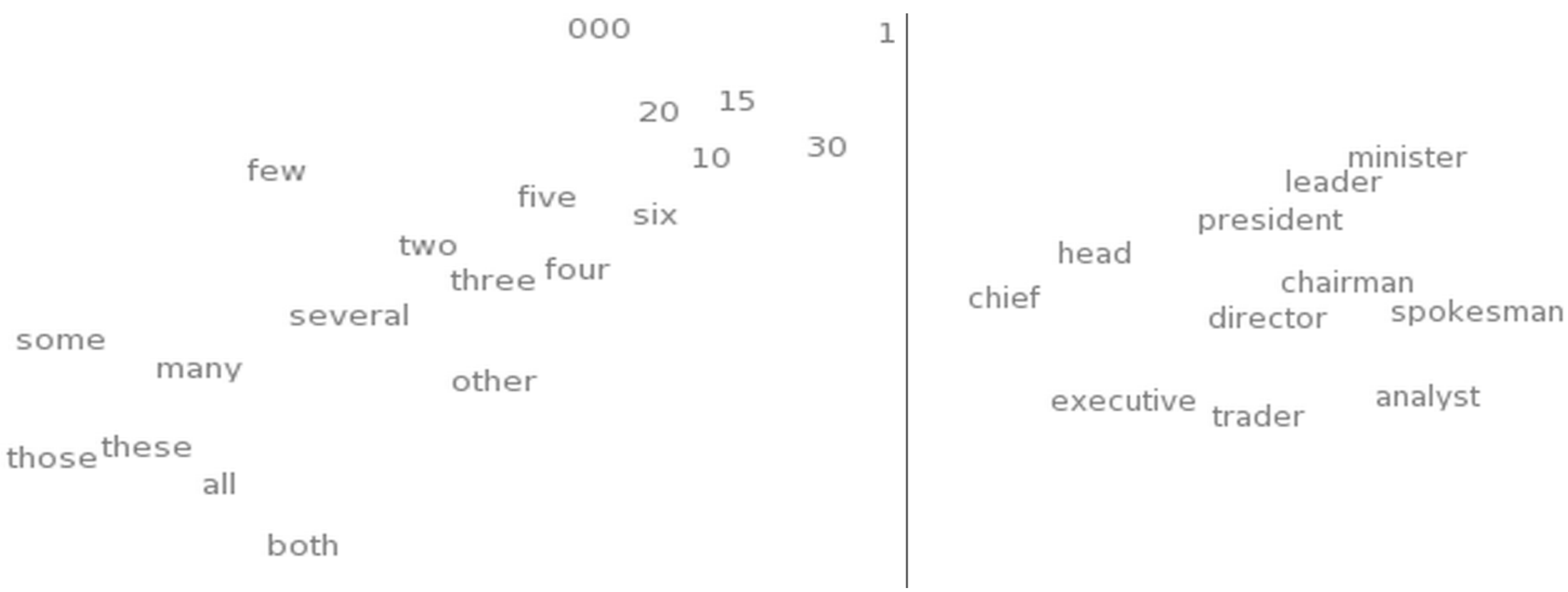


Figure 2: 2D t-SNE visualizations of word embeddings.
Left: Number Region; Right: Jobs Region [J.Turian *et al* . 2010]

Word embeddings

Mapping words to high-dimensional vectors (e.g. 200 dimensions)

$$R : Words = \{W_1, ..., W_n\} \rightarrow Vectors = \{R(W_1), ..., R(W_n)\} \subset R^d$$

Distance between vectors indicates the relation between words

$$R(W_1) \approx R(W_n) \rightarrow W_1 \approx W_n$$

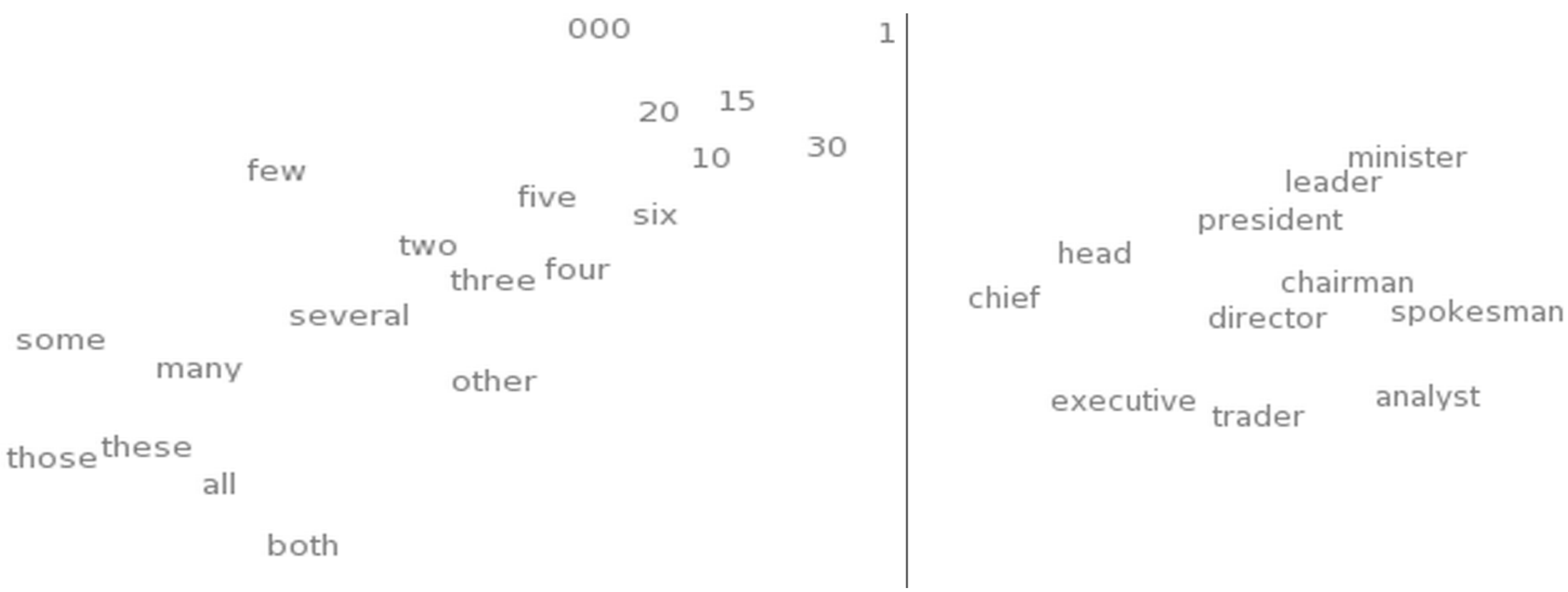


Figure 2: 2D t-SNE visualizations of word embeddings.
Left: Number Region; Right: Jobs Region [J.Turian *et al* . 2010]

FRANCE	JESUS	XBOX
AUSTRIA	GOD	AMIGA
BELGIUM	SATI	PLAYSTATION
GERMANY	CHRIST	MSX
ITALY	SATAN	IPOD
GREECE	KALI	SEGA
SWEDEN	INDRA	PSNUMBER
NORWAY	VISHNU	HD
EUROPE	ANANDA	DREAMCAST
HUNGARY	PARVATI	GEFORCE
SWITZERLAND	GRACE	CAPCOM

Figure 3: What words have embeddings closest to a given word? [R.Collobert *et al* . 2011]

Word embeddings approaches(1/3)

1. Tur: Collobert and Weston embeddings revised by Joseph Turian [J.Turian *et al.* 2010]

- ✤ Existence n-gram
- ✤ Training criterion: score (n-gram) > score (corrupted n-gram) + some margin

➡ Morpho-syntactic similarities

Word embedding of «**the**»

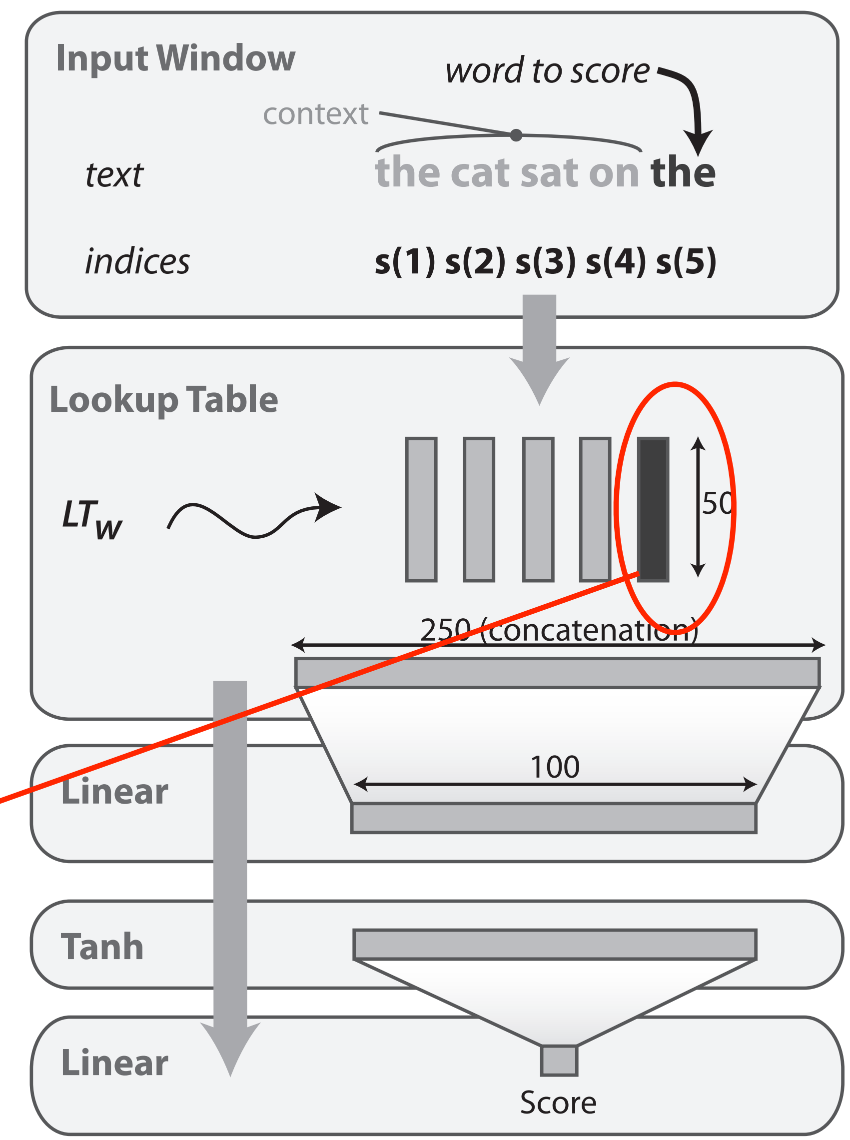


Figure 4: Neural architecture to compute 50 dimensional word embeddings

Word embeddings approaches(2/3)

2. Word2vec [T.Micolov *et al.* 2013]

- Continuous bag of words (CBOW)
  - predicting the current word based on its context

➡ Syntactic modeling

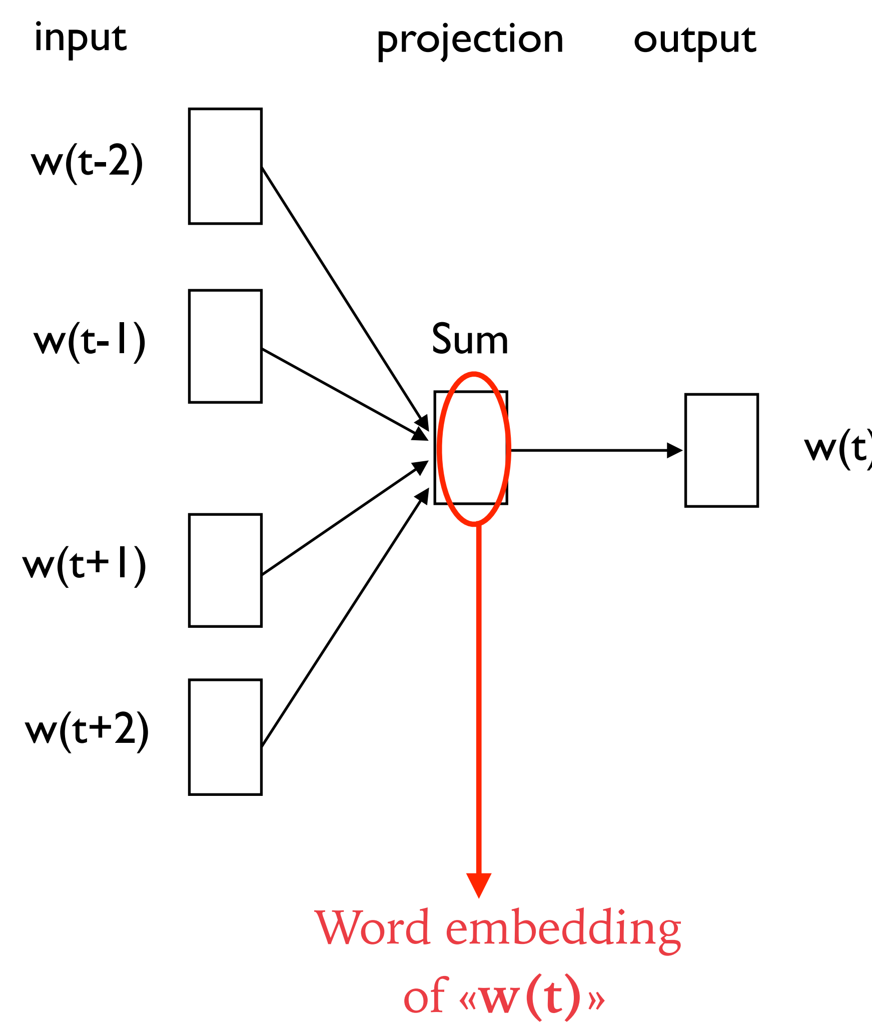


Figure 5: CBOW architecture

Word embeddings approaches(3/3)

3. GloVe: global vector for word representation [J.Pennington *et al.* 2014]

- ✤ Analysis of co-occurrences of words in a window
  - ✦ building a co-occurrence matrix
  - ✦ estimating continuous representations of the words

➡ Semantic similarities

Word embeddings combination

Combine word embeddings using denoising auto-encoder

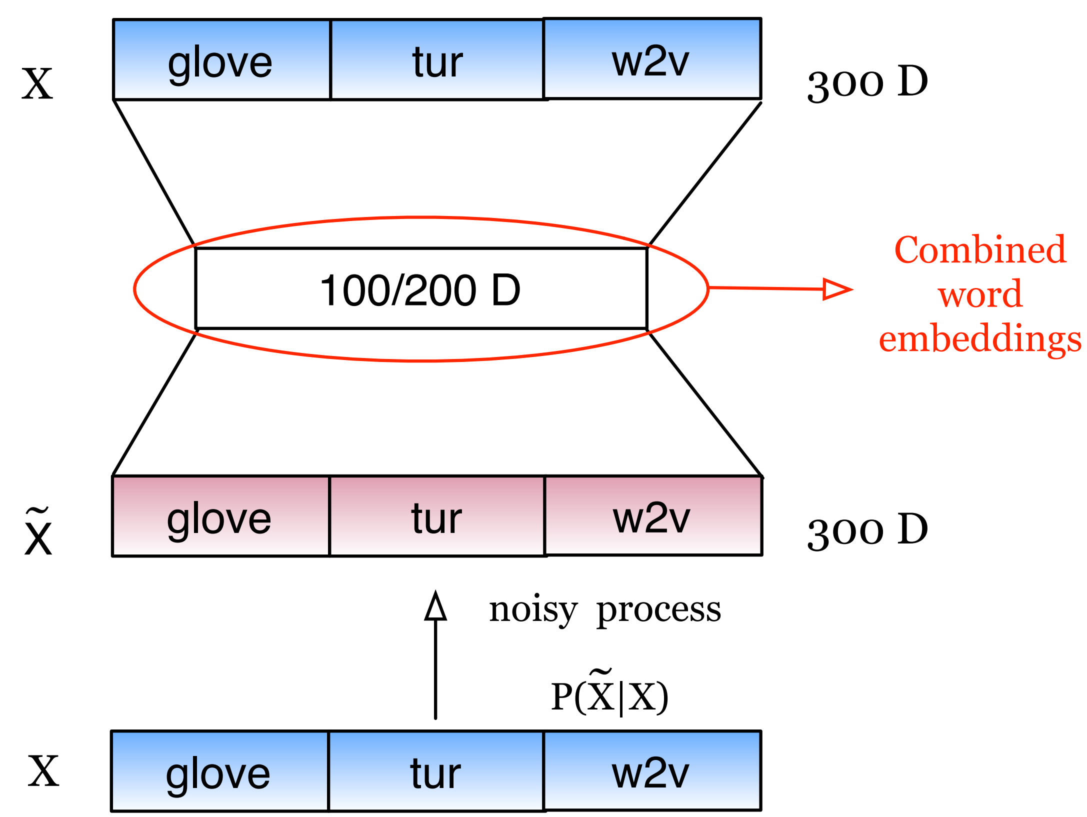


Figure 6: Using denoising auto-encoder to combine word embeddings

Neural architecture 1: Classical MLP

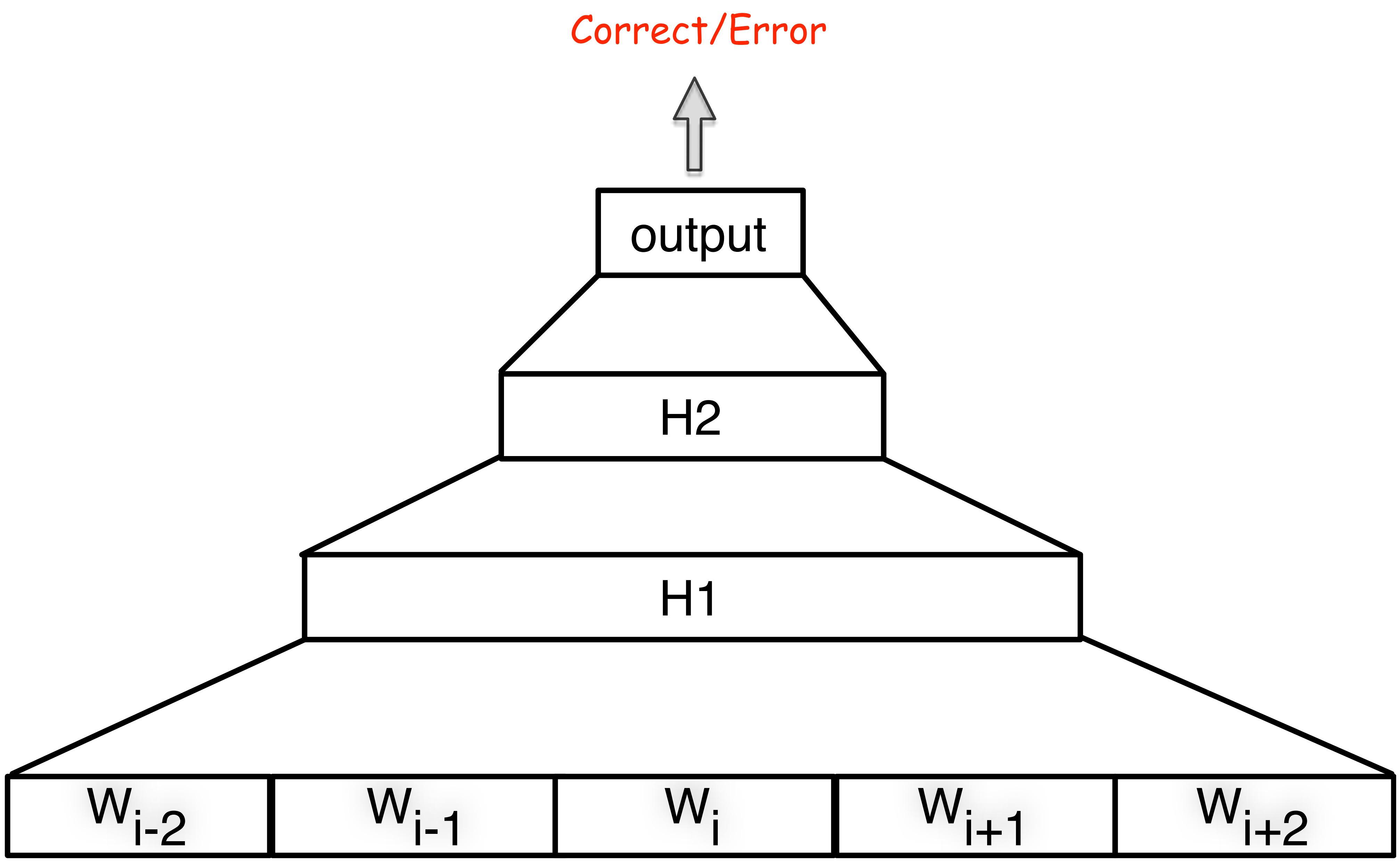


Figure 7: MLP architecture for ASR error detection task

Neural architecture 2: MLP-Multi-Stream

Inspired by [Y. Estève *et al.* 2015]

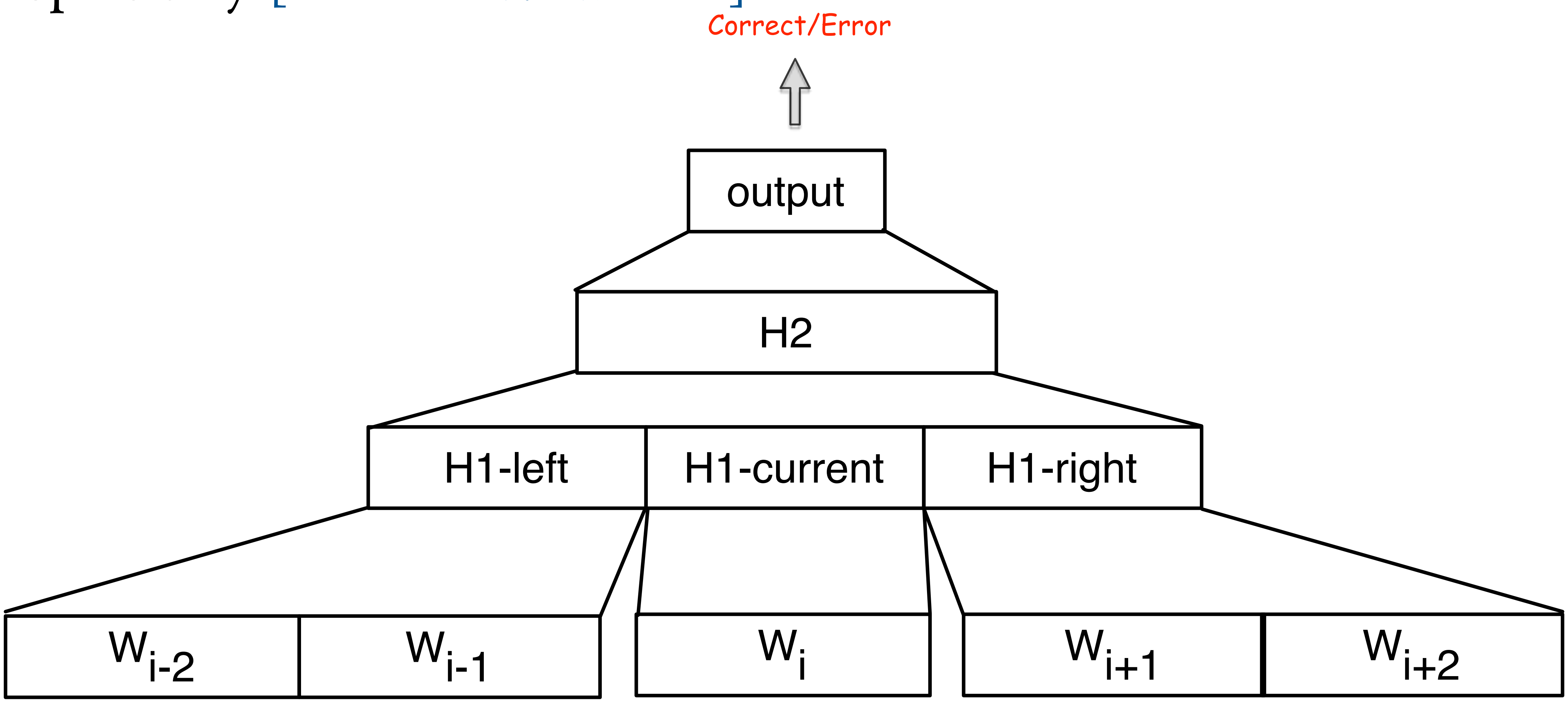


Figure 8: MLP-MS architecture for ASR error detection task

Neural architecture 3: MLP-Multi-Stream-i

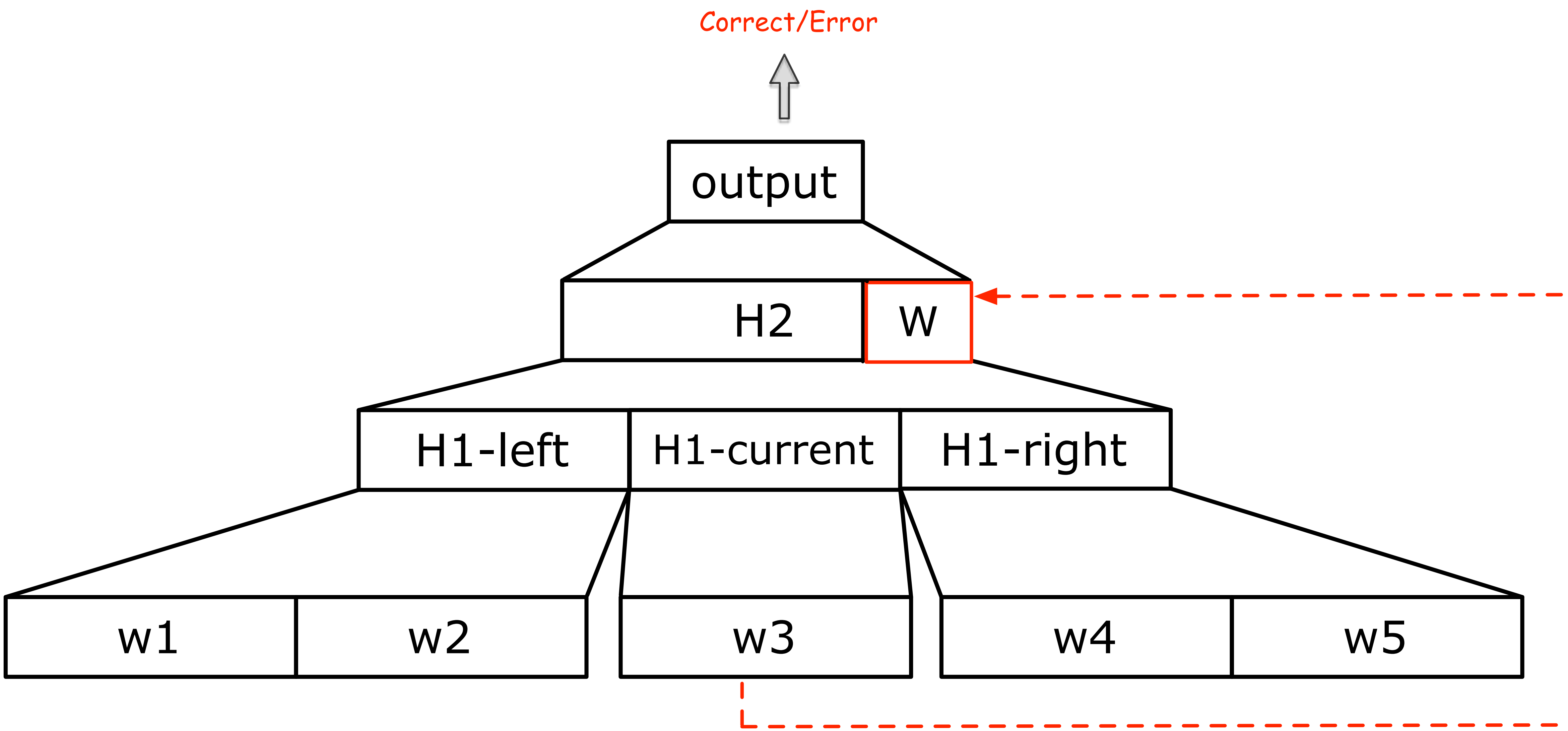


Figure 9: MLP-MS-i architecture for ASR error detection task

ASR error detection process

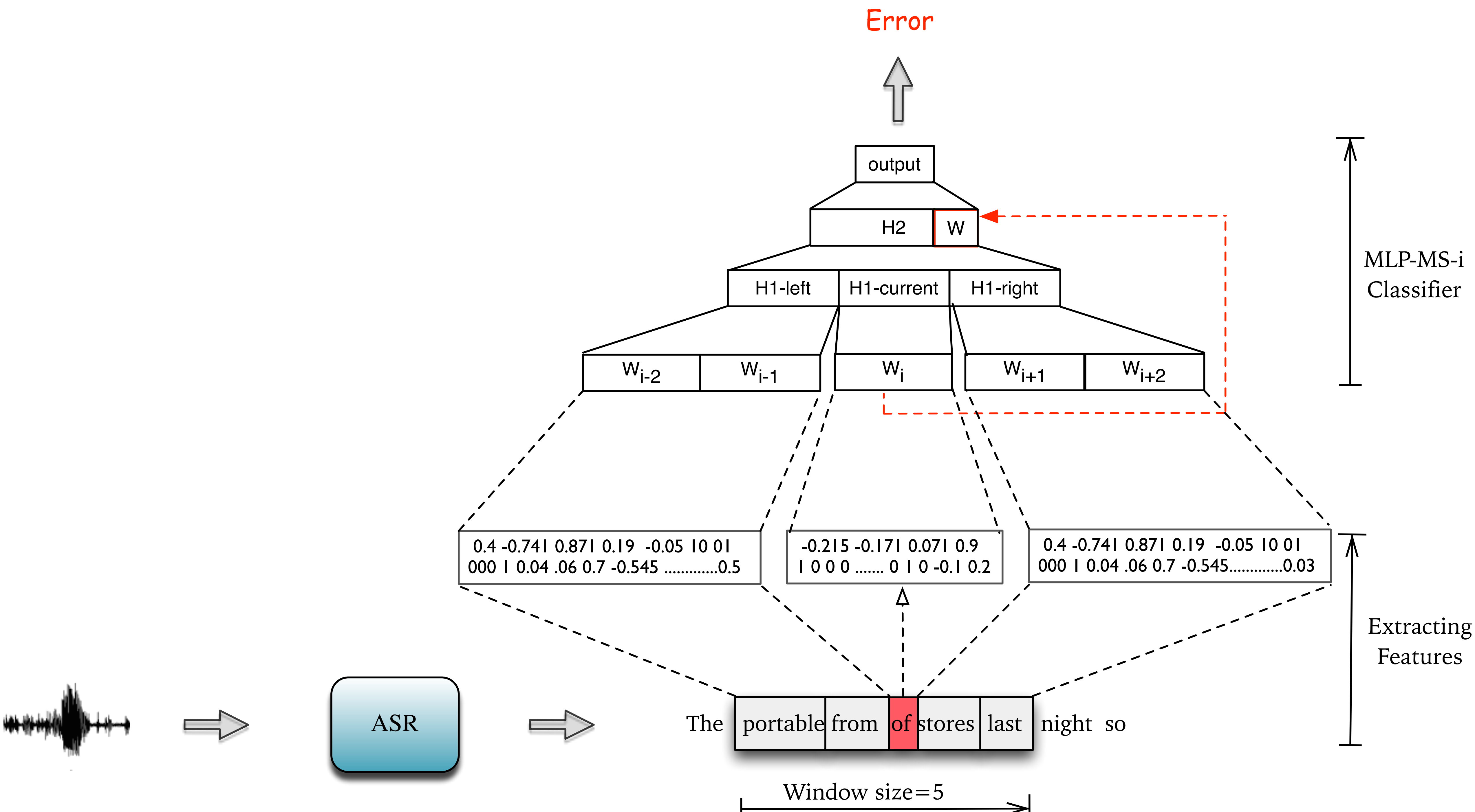


Figure 10: ASR error detection process

Experimental data

Training of the neural systems:

Automatic transcriptions of the ETAPE Corpus, generated by:

- ASR 1: CMU Sphinx decoder
  - acoustic models: GMM/HMM
- ASR 2: Kaldi decoder
  - acoustic models: DNN/HMM

ASR	Name	#words REF	#words HYP	WER
Sphinx GMM	Train	349K	316K	25.9
	Dev Sphinx	54K	50K	25.2
	Test Sphinx	58K	53K	22.5
Kaldi DNN	Dev Kaldi	54K	50K	23.1
	Test Kaldi	58K	53K	20.4

Table 1: Composition of the experimental corpus

Experimental data

Training data of the word embeddings:

Corpus composed of 2 billions of words:

- Articles of the French newspaper ”Le Monde”,
- French Gigaword corpus,
- Articles provided by Google News,
- Manual transcriptions: 400 hours of French broadcast news.

Evaluation results

- Neural architectures vs. CRF
- Evaluation metrics:
  - Error label: F-measure
  - Overall classification: CER

Comparison of different word embeddings

		Label error	Global
Neural architecture	Embeddings	F-measure	CER
MLP	GloVe	59.9	10.56
	w2v	61.1	10.36
	tur	60.4	10.32
	Auto-encoder-100	61.8	10.18
	Auto-encoder-200	**62.5**	**10.07**

Table 2: Comparison on Dev-sphinx of different types of word embeddings used as additional features in MLP error detection system.

Comparison and Robustness of different neural architectures

Train	Test	Approaches	Label error F-measure	Global CER
Train Sphinx	Test Sphinx	*CRF(baseline)*	*57.6*	*8.78*
		MLP	61.5	8.52
		MLP-MS	61.4	**8.43**
		MLP-MS-i	**62.1**	8.49
Train Sphinx	Test kaldi	*CRF(baseline)*	*51.3*	*8.59*
		MLP	50.4	8.34
		MLP-MS	49.4	8.29
		MLP-MS-i	**52.7**	**8.15**

Table 3: Error detection results on Test Sphinx and Test kaldi transcriptions.

Comparison and Robustness of different neural architectures

			Label error	Global
Train	Test	Approaches	F-measure	CER
Train Sphinx	Test Sphinx	*CRF(baseline)*	*57.6*	*8.78*
		MLP	61.5	8.52
		MLP-MS	61.4	**8.43**
		MLP-MS-i	**62.1**	8.49
Train Sphinx	Test kaldi	*CRF(baseline)*	*51.3*	*8.59*
		MLP	50.4	8.34
		MLP-MS	49.4	8.29
		MLP-MS-i	**52.7**	**8.15**

Table 3: Error detection results on Test Sphinx and Test kaldi transcriptions.

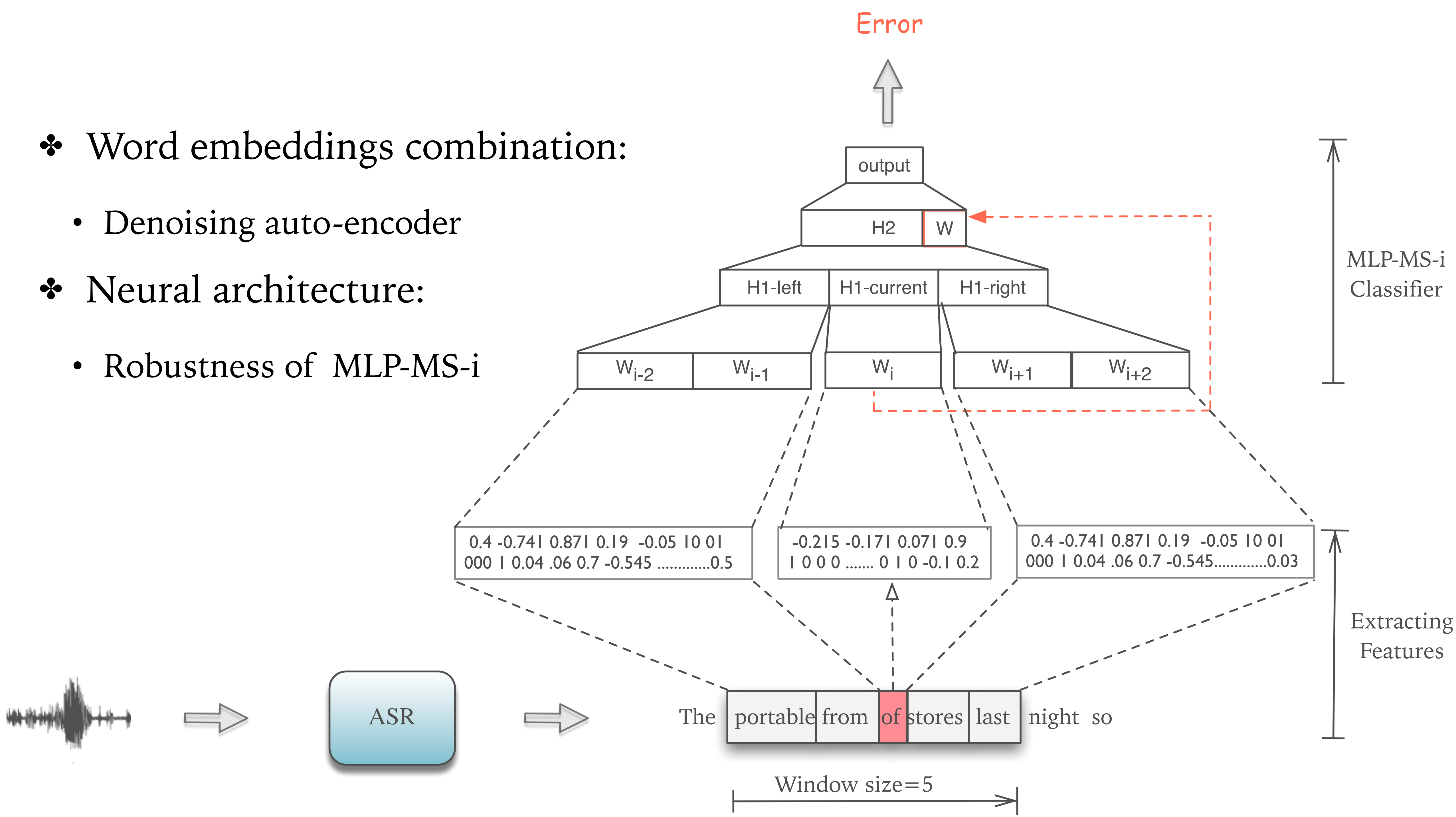
Comparison and Robustness of different neural architectures

Train	Test	Approaches	Label error F-measure	Global CER
Train Sphinx	Test Sphinx	*CRF(baseline)*	*57.6*	*8.78*
		MLP	61.5	8.52
		MLP-MS	61.4	**8.43**
		MLP-MS-i	**62.1**	8.49
Train Sphinx	Test kaldi	*CRF(baseline)*	*51.3*	*8.59*
		MLP	50.4	8.34
		MLP-MS	49.4	8.29
		MLP-MS-i	**52.7**	**8.15**

Table 3: Error detection results on Test Sphinx and Test kaldi transcriptions.

Conclusions

- Word embeddings combination:
  - Denoising auto-encoder
- Neural architecture:
  - Robustness of MLP-MS-i



Conclusions

Perspectives:

- Analysis of ASR error detection system outputs → Which ASR error are hard to detect, [S.Ghannay *et al.* ERRARE 2015]
- Exploiting new features:
  - Prosodic features → Combining continuous word representation and prosodic features for ASR error prediction [S.Ghannay *et al.* SLSP 2015]
  - Global semantic information
- Recurrent neural network → sequence prediction

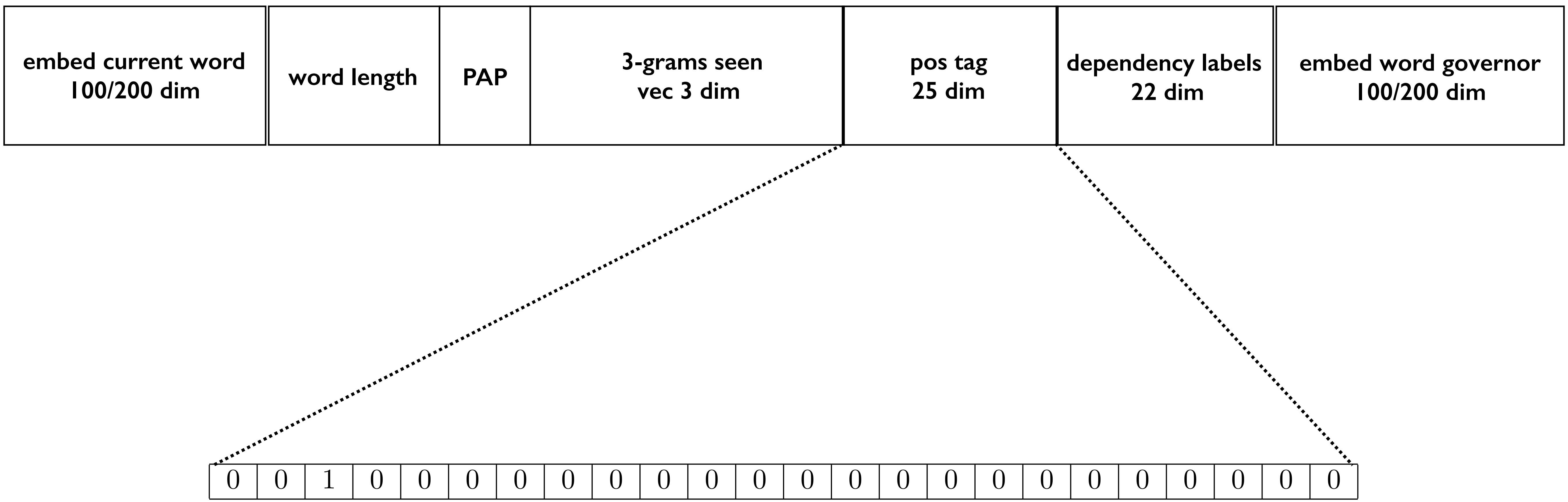
Thank you

Comparison and Robustness of different neural architectures

			Label error	Global
Train	Test	Approaches	F-measure	CER
Train Sphinx	Test Sphinx	*CRF(baseline)*	*57.6*	*8.78*
		MLP	61.5	8.52
		MLP-i	59.77	8.56
		MLP-MS	61.4	**8.43**
		MLP-MS-i	**62.1**	8.49
Train Sphinx	Test kaldi	*CRF(baseline)*	*51.3*	*8.59*
		MLP	50.4	8.34
		MLP-i	48.80	8.30
		MLP-MS	49.4	8.29
		MLP-MS-i	**52.7**	**8.15**

Table 3: Error detection results on Test Sphinx and Test kaldi transcriptions.

Neural network input feature vector format



Example: 25 POS tags, 3^{rd} POS tag

Figure 2 : Neural network input feature vector format

Comparison and Robustness of different neural architectures

		Label error			Global
corpus	**approach**	**P**	**R**	**F**	**CER**
Dev-sphinx	Naive	66.9	48.8	56.4	11.21
	CRF	70.8	50.6	59.0	10.44
	MLP-1	71.0	53.3	60.9	10.17
	MLP-2	70.0	56.4	62.5	10.07
	MLP-MS	70.7	55.9	62.5	**9.99**
	MLP-MS-i	68.8	58.0	**63.0**	10.15
Test-sphinx	Naive	65.3	47.1	54.7	9.42
	CRF	69.2	49.3	57.6	8.78
	MLP-1	69.3	53.3	60.3	8.50
	MLP-2	67.8	56.3	61.5	8.52
	MLP-MS	68.8	55.5	61.4	**8.43**
	MLP-MS-i	67.5	57.4	**62.1**	8.49

Table 3 : Error detection results on ASR Sphinx transcriptions.

Comparison and Robustness of different neural architectures

		Label error			Global
corpus	**approach**	**P**	**R**	**F**	**CER**
Dev-kaldi	Naive	68.6	31.0	42.7	10.95
	CRF	63.6	40.5	49.5	10.88
	MLP-1	70.3	35.9	47.6	10.43
	MLP-2	68.0	38.4	49.1	10.49
	MLP-MS	69.8	36.5	47.9	10.44
	MLP-MS-i	68.3	41.3	**51.5**	**10.25**
Test-kaldi	Naive	69.3	32.2	43.9	8.70
	CRF	64.3	42.6	51.3	8.59
	MLP-1	69.4	37.0	48.3	8.41
	MLP-2	68.2	40.0	50.4	8.34
	MLP-MS	70.0	38.2	49.4	8.29
	MLP-MS-i	68.5	42.9	**52.7**	**8.15**

Table 3 : Error detection results on ASR kaldi transcriptions