

**ICFHR–2010 Tutorial:**  
**Multimodal Computer Assisted Transcription of**  
**Handwriting Images**  
**Ilp – CATTI in practice**

Alejandro H. Toselli & Moisés Pastor & Verónica Romero

{ahector,moises,vromero}@iti.upv.es



*Pattern Recognition and Human Language Technology Group*

Instituto Tecnológico de Informática – Universidad Politécnica de Valencia



Spain

November 2010

A BLANK PAGE

## Tutorial Contents and Schedule

- I Introduction
  - Multimodal Interaction in Pattern Recognition
  - Quick Survey of Handwritten Text Recognition (HTR) concepts and techniques
  - Interactive-Predictive Pattern Recognition and Document Image Analysis
- I-p Off-line HTR in practice
  - HTR Preprocessing
  - Training HMMs using the "Hidden Markov Model Toolkit" (HTK)
  - Training Language Models and Dictionaries for HTR
  - HTR Experiments
- II Computer-Assisted Transcription of Text Images (CATTI)
  - Human interaction in HTR
  - A CATTI formal framework
  - Feedback-derived dynamic language modelling and search
  - Performance measures and results achieved in typical applications
- II-p **CATTI in practice**
  - Adapting Language Models and Search for CATTI
  - CATTI Experiments
  - Analyzing quantitatively the CATTI performance
- III Multimodality in CATTI (MM-CATTI)
  - Touchscreen based multimodal user correction
  - A MM-CATTI formal framework
  - Multimodal language modelling and search
  - Performance measures and results achieved in typical applications
- III-p Demonstration of a complete MM-CATTI System in a real HTR task

## Index

- 1 CATTI Experiments ▷ 3
- 2 Analyzing quantitatively the CATTI performance ▷ 6
- 3 Bibliography ▷ 13

## Color Notation used for CATTI Output

References	<b>Green text</b>	denotes the reference text of the sample being transcribed.
	<b>Red text:</b>	denotes the user introduced/corrected words (by keyboard).
	<b>Blue text:</b>	denotes the error-free prefix validated by user before introducing/correcting the following word.
	<b>Cyan text:</b>	denotes the whole prefix (the validated part plus the corrected word) used by CATTI to suggest the new suffix continuation.
	<b>Black text:</b>	denotes the output hypothesis/suffix from CATTI.

## CATTI Experiments: “Spanish Number Corpus”

- Multi-writer corpus.
- Character HMMs: 6 states, 2 Gaussian densities per state
- Language models: Bi-grams

	Train	Test	Total	Lexicon
# writers	18	11	29	–
# sentences	297	187	484	–
# words	1 300	827	2 127	52

LM training data: 20 920 sentences (91 760 **running words**)

WER (%)	WSR (%)	Effort-Reduction (%)
18.74	17.53	6.45

## CATTI Experiments: IAMDB

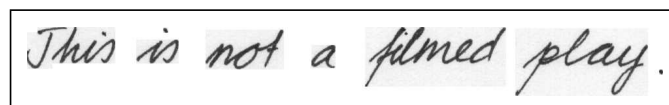
- Multi-writer corpus.
- No case distinction or diacritics; no punctuation marks.
- Character HMMs: 6 states, 64 Gaussian densities per state
- Language models: Bi-grams

Number of:	Training	Test	Total	Lexicon	OOV	Tr. Ratio
writers	448	100	548	–	–	–
sentences	2 124	200	2 324	–	–	–
words	42 832	3 957	46 789	<b>8 017</b>	921	<b>128</b>
characters	216 774	20 726	237 500	78	0	<b>2 779</b>

LM training data: approx.  $10^6$  **running words** from the LOB corpus

WER (%)	WSR (%)	Effort-Reduction (%)
25.8	21.8	16

## CATTI on IAM DB: Example 1 - ID: c03-003d-s00



Reference: this is not a filmed play

Hypothesis: this is not [] allowed play

Errors: D=1 S=1 I=0

Prefix 1: this is not a

Hypothesis 1: this is not a filmed play

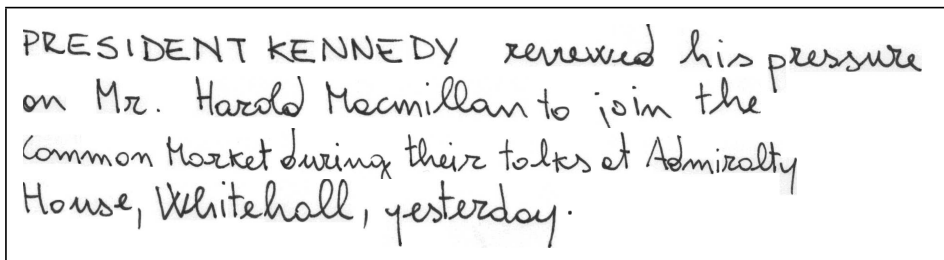
Final Result: this is not a filmed play

Post-editing WER: 2/6 (33%)

Interactive WSR: 1/6 (17%)

Estimated effort reduction:  $1 - 17/33$  (48%).

## CATTI on IAM DB: Example 2 - ID: a04-000-s00



PRESIDENT KENNEDY renewed his pressure on Mr. Harold Macmillan to join the Common Market during their talks at Admiralty House, Whitehall, yesterday.

Reference: president kennedy renewed his pressure on mr harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

Hypothesis: [this] [per] cent clay renewed his pressman [] is harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

Errors: D=1 S=4 I=2

## CATTI on IAM DB: Example 2 - ID: a04-000-s00 (cont)

Prefix 1: president

Hypothesis 1: president kennedy renewed his pressman is harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

Prefix 2: president kennedy renewed his pressure

Hypothesis 2: president kennedy renewed his pressure on the harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

Prefix 3: president kennedy renewed his pressure on mr

Hypothesis 3: president kennedy renewed his pressure on mr harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

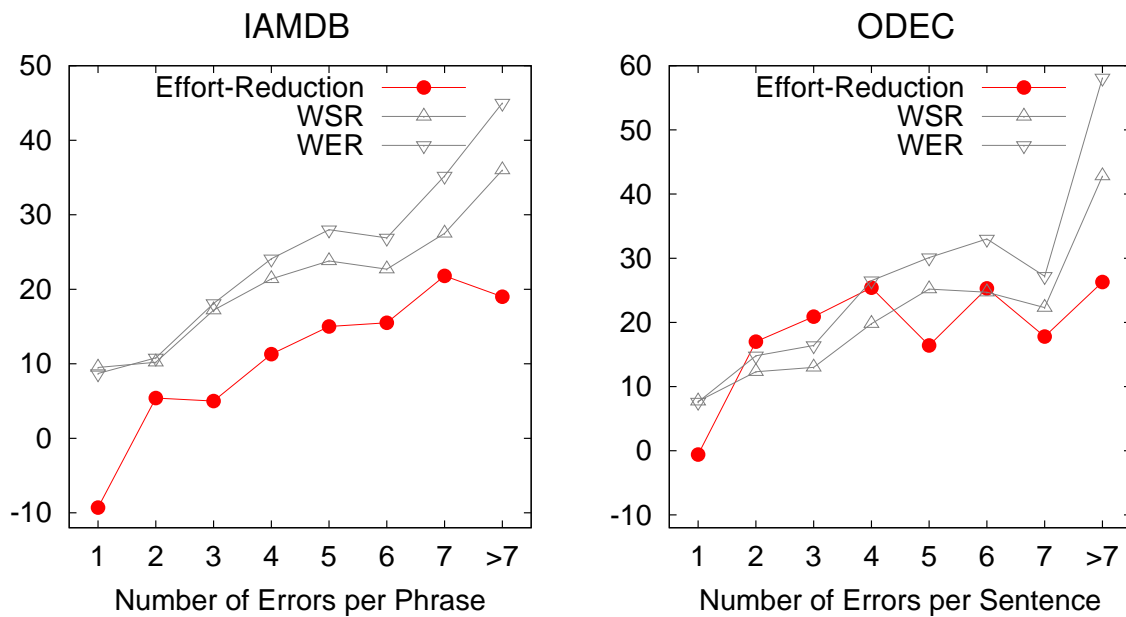
Final Result: president kennedy renewed his pressure on mr harold macmillan to join the common market during their talks at admiralty house whitehall yesterday

Post-editing WER: 7/22 (32%)

Interactive WSR: 3/22 (14%)

Estimated effort reduction: 1 – 14/32 (56%).

## CATTI Results: impact of number of errors per sentence



CATTI performance for sentences with different number of errors

## CATTI Results: impact of number of errors per sentence

- **Estimated effort reduction** increase significantly with the number of errors per sentence.
- CATTI efficacy correcting more than one error per interaction step in sentences with several initially misrecognized words.
- For the special case of sentences with only one initial error, post-editing the single error does not risk unwished changes of other correct words.

## Statistics of Final Results for “Spanish Number Corpus”

#Er/Phr	#Samp	#Err	#lter	Long-Ref	WER	WSR	EFR
*0	97	0	0	426	0.00	0.00	0.00
*1	55	55	58	232	23.71	25.00	-5.45
2	18	36	38	84	42.86	45.24	-5.56
3	11	33	26	57	57.89	45.61	21.21
4	3	12	7	14	85.71	50.00	41.67
5	1	5	3	6	83.33	50.00	40.00
6	2	12	4	8	150.00	50.00	66.67

GLOBAL: WER=18.50 WSR=16.44 EFR=11.11 (D=7 S=127 I=26)

GLOBAL: WER=57.99 WSR=46.15 EFR=20.41 (D=4 S=73 I=25) without considering \*

## Statistics of Final Results for IAMDB

#Er/Phr	#Samp	#Err	#lter	Long-Ref	WER	WSR	EFR
*0	22	0	0	211	0.00	0.00	0.00
*1	32	32	35	369	8.67	9.49	-9.38
2	28	56	53	518	10.81	10.23	5.36
3	20	60	57	332	18.07	17.17	5.00
4	20	80	71	332	24.10	21.39	11.25
5	12	60	51	215	27.91	23.72	15.00
6	14	84	71	313	26.84	22.68	15.48
7	17	119	93	338	35.21	27.51	21.85
8	5	40	33	146	27.40	22.60	17.50
9	9	81	65	176	46.02	36.93	19.75
10	2	20	21	39	51.28	53.85	-5.00
11	3	33	29	73	45.21	39.73	12.12
12	7	84	65	183	45.90	35.52	22.62
13	1	13	8	37	35.14	21.62	38.46
14	3	42	33	106	39.62	31.13	21.43
15	1	15	12	30	50.00	40.00	20.00
18	1	18	13	26	69.23	50.00	27.78
20	2	40	30	76	52.63	39.47	25.00
22	1	22	16	32	68.75	50.00	27.27
24	1	24	22	27	88.89	81.48	8.33

GLOBAL: WER=25.79 WSR=21.74 EFR=15.71 (D=118 S=695 I=110)

GLOBAL: WER=29.71 WSR=24.77 EFR=16.61 (D=116 S=665 I=110) without considering \*

## Bibliography

- E. Vidal, F. Casacuberta, L. Rodríguez, J. Civera and C. Martínez. “Computer-assisted translation using speech recognition”. IEEE Trans. on Audio, Speech and Language Processing, 14(3):941-951, 2006.
- E. Vidal, L. Rodríguez, F. Casacuberta and I. García-Varea: “Interactive Pattern Recognition”. 4th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI-07), Volume 4892 of LNCS, pp.60-71. Brno, Czech Republic, June 2007.
- L. Rodríguez, F. Casacuberta, and E. Vidal. “Computer Assisted Transcription of Speech” Proc. of the third Iberian Conference on Pattern Recognition and Image Analysis, Volume 4477 of LNCS, pp.241-248, Girona (Spain), June 2007.
- A.H. Toselli, V. Romero, L. Rodríguez and E. Vidal. “Computer Assisted Transcription of Handwritten Text”. 9th Int. Conference on Document Analysis and Recognition (ICDAR 2007), pp.944-948. IEEE Computer Society, Curitiba, Paraná (Brazil), September 2007.
- V. Romero, A.H. Toselli, L. Rodríguez and E. Vidal. “Computer Assisted Transcription for Ancient Text Images”. Int. Conference on Image Analysis and Recognition (ICIAR 2007), volume 4633 of LNCS, pp.1182-1193. Montreal (Canada), August 2007.