# CAT-API Framework Prototypes

V. Alabau, J. M. Benedí, F. Casacuberta, L. A. Leiva, D. Ortiz-Martínez,
V. Romero, J. A. Sánchez, R. Sánchez-Saez, A. H. Toselli, E. Vidal
*ITI – Instituto Tecnológico de Informática*
*Universidad Politécnica de Valencia*
*{valabau,jbenedi,fcn,luileito,dortiz,vromero,jandreu,rsanchez,ahector,evidal}@iti.upv.es*

***Abstract*—Traditionally, Pattern Recognition applications have focused on fully automatic systems. However, since their performance is far from being perfect, such automatic systems cannot replace the human expertise. Typically, experts use a Pattern Recognition system in a two-step operation: first, the application generates an output in a fully automatic way; and second, the user revises that output in order to achieve high-quality results. This post-edition approach is rather inefficient and uncomfortable for the user.**

**An alternative, yet effective approach to traditional Pattern Recognition systems is the interactive-predictive paradigm in which both the system is guided by the user and the user is assisted by the system to complete their tasks as efficiently as possible. We present three protypes of Computer Assisted Tools: transcription, translation, and syntactic parsing, respectively. Such prototypes combine the efficiency of the traditional Pattern Recognition systems with the accuracy of the human expertise, enforcing all a multimodal, interactive strategy, and fully integrating the user's knowledge into the Pattern Recognition process. User feedback directly allows to improve system accuracy, while multimodality increases both system ergonomy and user acceptability.**

***Keywords*-interactive; multimodal; transcription; translation; parsing; CAT; IHT; IMT; IPP;**

## I. Introduction

Typically, human experts use a Pattern Recognition (PR) system in a two-step operation: first, the application generates an output in a fully automatic way; and second, the user revises that output in order to achieve high-quality results. This post-edition approach is rather inefficient and uncomfortable for the user. An alternative, yet effective approach to traditional PR systems is the interactive-predictive paradigm in which both the system is guided by the user and the user is assisted by the system to complete their tasks as efficiently as possible.

We introduce three prototypes [2], [3], [4] developed with the Computer Assisted Tools API (CAT-API) library [5] under the MIPRCV[1] framework (Figure 3) — Interactive Handwritting Transcription (IHT, Figure 1a), Interactive Machine Translation (IMT, Figure 1b), and Interactive Predictive Parsing (IPP, Figure 1c), respectively. Such prototypes combine the efficiency of the traditional Pattern Recognition systems with the accuracy of the human expertise, enforcing all a multimodal, interactive strategy, and fully integrating the user's knowledge into the PR process.

User feedback directly allows to improve system accuracy, while multimodality increases both system ergonomy and user acceptability. Such feedback is of invaluable help to produce high quality, 100% error-free results.

The proposed techniques may reduce dramatically the typing effort, regarding to the effort needed to simply post-process the output automatically generated by a PR system. These demos work over the Web, so they can be implemented on a myriad of devices: e.g., laptops, netbooks, mobile gadgets, or embedded systems. One can try the CAT-API prototypes at `http://cat.iti.upv.es/`.

## II. User Interaction Protocol

Users first choose a task from a list of available corpora — for example, an IHT user may choose a corpus to transcribe, while an IMT user can upload a document and translate it. Then, the system and the user will interact each other until a perfect output is achieved.

Our CAT-API prototypes all share the common user interaction protocol illustrated in the MIPRCV framework [1], which is described as follows:

- To begin, the system proposes an automatic output to the user.
- She then validates the longest prefix of that output which is error-free, by sending some feedback signal to correct the first error found.
- At this point the system suggests a suitable, extended consolidated suffix based on the previous validated prefix, the decoded human feedback, and other possible user amendments.
- This process is iterated until the user considers that the system output is wholly correct.

Thereby a perfect, high-quality output is guaranteed, since the user is tightly embodied on the sytem's inner process. This way of interacting with the system is very natural and efficient, at the expense of the system being less deterministic.

On the other hand, all user interactions are stored on the web server as plain text logs, so three great advantages are derivated:

[1]Multimodal Interactive Pattern Recognition and Computer Vision


(a) Handwritting Transcription (b) Machine Translation (c) Predictive Parsing

Figure 1: Screenshots of interactive prototypes in keyboard input modality.


(a) Delete (b) Insert (c) Reject (d) Accept

Figure 2: Some CAT-API operations that can be achieved in the e-pen modality.

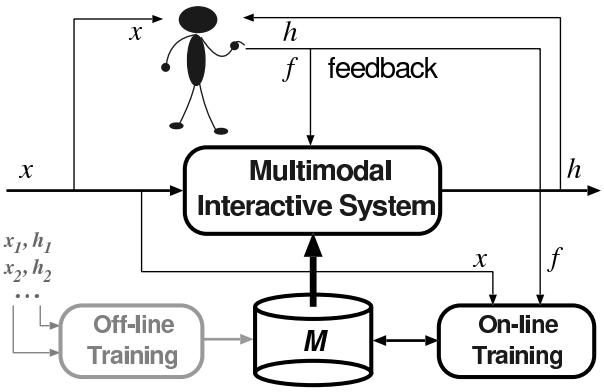


Figure 3: CAT-API implements the MIPRCV framework [1].

1) one can retake the tasks from the last image/sentence/tree she left;
2) other users can help to translate/transcribe/annotate the full document, exploiting thus the collaborative nature of the Web; and
3) detailed statistics can be generated about UI and prototype usage, without burdening the user.

## III. Conclusion and Future Work

These demo systems may entail a drastic reduction of user effort without sacrifying usability — quite the opposite, since, the Human-Computer Interaction paradigm is considered the main actor in front of a PR scenario. What is more, further promising figures are able to be used for adapting PR techniques to the dynamic and changing environments of interactive systems.

## Acknowledgments

This work has been supported by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018), the Regional research programmes PROMETEO/2009/014 and MITTRAL (TIN2009-14633-C03-01), the EC (FEDER), the Spanish MEC grant TIN2006-15694-CO2-01, the FPU fellowship AP2006-01363 and the FPI fellowship 2006-04.

## References

[1] E. Vidal, L. Rodríguez, F. Casacuberta, and I. García-Varea, "Interactive pattern recognition," pp. 60–71, 2007.

[2] V. Romero, L. A. Leiva, A. H. Toselli, and E. Vidal, "Interactive multimodal transcription of text imagse using a web-based demo system," in *Proceedings of the International Conference on Intelligent User Interfaces*, 2009, pp. 477–478.

[3] D. Ortiz-Martínez, L. A. Leiva, V. Alabau, and F. Casacuberta, "Interactive machine translation using a web-based architecture," in *Procedings of the International Conference on Intelligent User Interfaces*, 2010, pp. 423–424.

[4] R. Sánchez-Sáez, L. A. Leiva, J. A. Sánchez, and J. M. Benedí, "Interactive predictive parsing using a web-based architecture," in *Proceedings of NAACL*, 2010.

[5] V. Alabau, V. Romero, D. Ortiz-Martínez, and J. Ocampo, "A multimodal predictive-interactive application for computer assisted transcription and translation," in *Proceedings of International Conference on Multimodal Interfaces (ICMI)*, 2009, pp. 227–228.