A Proposal of an Evaluation Framework for Writing Assistance Systems: Application to VITIPI

Philippe Boissière, Daniel Dours

IRIT – UPS – CNRS, 118 Route de Narbonne, 31 062 TOULOUSE Cedex France
{boissier, dours}@irit.fr

Abstract. We propose an evaluation framework for writing assistance systems. This framework is used to evaluate VITIPI and provide the first results.

Introduction

Writing is very important for anyone in everyday life. For some disabled, it may be the only way of communication. Unfortunately, it takes them a long time to write. Thanks to Information and Communication Technology, we can find a lot of writing assistance systems, using abbreviated, selected or predicted words. But it is difficult to make unbiased evaluation systems, and users have no objective criteria to select one of them. In the first part of the talk, we are going to list all main features of existing systems. Going further, we shall explain our frame evaluation. Finally, we will give VITIPI as an instance for the framework evaluation.

Evaluation Framework

In order to help the user to choose a writing assistance system, according to his or her intellectual and/or physical abilities, we propose objective criteria for comparisons. We are to notice that it is difficult to compare two systems in regards to their variability. First of all, several systems are still mainly laboratory prototypes and their characteristics are not known. Then, evaluation results are often not published or based on tries with non-identical corpora.

Rigorously, an evaluation is made up of five ergonomic criteria: utility, usability, efficiency, no dangerousness and user's satisfaction [1]. In the field of writing assistance some of these criteria are trivial. We just want to focus on the efficiency criteria which is the most important according to user's point of view. The evaluation framework is divided in two parts: The first one deals with "*system features*" The second one provides "*system performances*".

"*System features*" represents the features characterizing the assistance writing system (type of coding, interactivity, domain and lexicon size, language, etc.).

"*system performances*" is made up with various parameters. We have avoided qualitative results based on users' or therapists' judgments because they are too personalized.

System Features

- **Coding principles :** The user has to write a text shorter and easier than he should have written without assistance system. Two kinds of coding can be distinguished. *Static and explicit coding methods* : User knows a list of abbreviated words and rewriting rules, *Dynamic and implicit coding methods* : User doesn't know coding rules. When he types one letter, system displays others letters, it is an *interactive coding*. Depending on the interaction mode, two sorts of systems can be found: *Displaying list of words* or *not displaying list of words*: System displays a part of word.
- **Running:** Systems dealt with *isolated words* only. Nowadays, systems try to take into account *previous word* by introducing grammatical rules or others patterns (Markov models, N-grams, ….).
- **Typing adaptability** : User can type *new entities* i.e. new words or sentences that do not belong to the initial database, or misspelled words (typing or orthographic mistakes). System can accept it.
- **Database update**: When new entities are detected by the system, they could be: *Integrated on line* during the session, (enjoying typing adaptability) or *Integrated off line* after the session, requiring intervention of target (sometimes with specialized knowledge).
- **Automatic database building**: User can build (him/her) self its database, with its own parameters, without standard lexicon, only for *one language* or for *every languages*, without knowing grammatical rules.
- **Separated windows editor**: Window text editor could be an ordinary word processor.

System Performances

The main value evaluated is the ratio time saving with the evaluation of the following parameter: The ratio of input keys typed by the user compared to the ratio of output characters provided by the system output. To evaluate it, we have to compute the number of keys typed by the user with system assistance :

$$Input\_symbols + key\_functions \quad (1)$$

With:

- *Input_symbols*: Number of letters typed by user. It includes also digits or numbers contained in the text and spacing (just one for each word)
- *key_functions*: Number of keys functions typed to select word in list or to correct word prediction.

Without system assistance, user should have to type all the entire text .

$$Input\_symbols + Output\_symbols \quad (2)$$

With :

- *Output_symbols*: Number of symbols automatically provided by the system.

Finely, the ratio time saving (*R.T.S)* is obtained by the following formula (**3**) which is similar to Zagler keystroke saving [2].

$$\boldsymbol{R.T.S} = \left(1 - \frac{(1)}{(2)}\right) \qquad (\mathbf{3})$$

VITIPI Evaluation

VITIPI is a writing assistance system. We are not going to explain its running yet exposed in various publications [3], [4], we just want to project it through the Frame Work evaluation, as an instance of assistant system.

System Features: VITIPI is a *dynamic and implicit coding method* that doesn't display list of words. It runs with *previous words* with a very huge *typing adaptability*. Database is automatically *updated online* and could be built in *every language*. VITIPI can be used with *a separate window editor*.

System Performances: Various simulation tests were done with several corpora. Using a weather forecasting French corpus (439 sentences, 7,830 words 985 vocabulary size) it has been found that with unknown sentences corpus and 3 previous words, R.T.S. is 45 % when a standard lexicon (5,930 French words) is used, 47 % without French lexicon. When all sentences belong to corpus, R.T.S. climbs up to 78.9 % without French lexicon and 10 previous words. Another testing has been made with e-mail exchanges concerning a workshop organization (159 sentences, 2,539 words 912 vocabulary size). It has been shown that with unknown sentences corpus and 3 previous words, R.T.S is 30 % with lexicon versus 43% without. When all sentences belong to corpus R.T.S. reaches 82.7 % with 10 previous words without lexicon.

Conclusion

We point out some features to evaluate a writing assistance system. Thanks to this framework we hope we reach for two objectives: The first one is to help users by providing unbiased criteria to help them select their system. The second one is to help designers by optimizing parameters in order to near the maximum efficiency. This frame is available to evaluate several writing assistance systems.

References

1. Sperandio, J.C., *Critères Ergonomiques d'Assistance Technologique aux Opérateurs*, JIM'2001 : Interaction Homme-Machine & Assistance pour Grand Public in Systèmes Complexes / Handicap, p. 30-37 Metz, 4-6 juillet 2001.

2. Zagler, W.L Seisenbacher, G. *German Language Predictive Typing : Results from a Feasibilty Investigation*. in Proceedings of the ICCHP'2000, pp.771-779, Karlsruhe, 17-21 July 2000.
3. Boissière, Ph., Dours, D., VITIPI : Comment un Système d'Assistance à l'écriture pour les Personnes Handicapées Peut Offrir des Propriétés Intéressantes pour le TALN ? in : TALN 2001 Tours 5 – 7 juillet 2001, pp. 183 – 192.
4. Boissière, Ph., Dours, D., *From a Specialised Writing Interface Created for Disabled, to a Predictive Interface for all: The VITIPI System*. in 1st International UAHCI 2001 Conference, News-Orleans. 5 - 10 août 2001, pp. 895-899.