# A General Public Application of Pedagogic and Linguistic Vocations of Speech Synthesis: Ordictée

Marc Guyomard
Université de Rennes I, IRISA/ENSSAT 6, rue de Kerampont
BP 447 F-22305 Lannion Cedex France
guyomard@enssat.fr

Jacques Siroux
IRISA/IUT Lannion France

Dominique Pernici and Christophe Royer
ENSSAT Lannion France

## Abstract

We present the Ordictée© software which allows the pupil to independently practice dictation exercises. Ordictée acts as a substitute for the teacher with regards to reading and dictation correction and allows the pupil to be almost completely in control in a non-stressful environment. Ordictée is made up of three modules which respectively allow the *pupil* to carry out the actual dictation, the *tutor* to administer a dictation database whilst adapting each text to the context of the dictation and the *designer* to generically specify the environment in which the dictations are carried out. Ordictée uses the *Proverb* software from the company "Elan Informatique".

## 1. General presentation

At the moment, we are witnessing an explosion in the software market of products aimed at assisting language learning. At the same time, research involving automatic speech processing (recognition and synthesis) which has been confined to the laboratories for a long time, is today ready for use. High quality products – especially in the domain of speech synthesis – are marketed. However, the merging of these trends remains often too superficial: the vocal terminals serve only as substitutes for the most classical means, without having a synergetic effect. On the other hand, some applications would be quite impossible without vocal based softwares. Such is the case of software designed to help with dictation. Nevertheless and especially for the French, such systems are rare (see however (Cotto 1990a, Cotto 1990b)) and those marketed are on this side of the state of the art.

In this context, we have decided to start the Ordictée project which enables a pupil to practice dictation exercises. Having selected a dictation which corresponds to their abilities, the pupil can ask Ordictée to read the dictation, to do the exercise (during which time they can type in the text being read to them). Following a second phase including further readings and subsequent corrections, the machine is asked to do the correcting.

In section 2 of the paper, we develop the structure of the software, demonstrate its various functions and explain its method of use. Section 3 presents the solution adopted to follow the typing activity. We then describe the algorithm employed for the correction

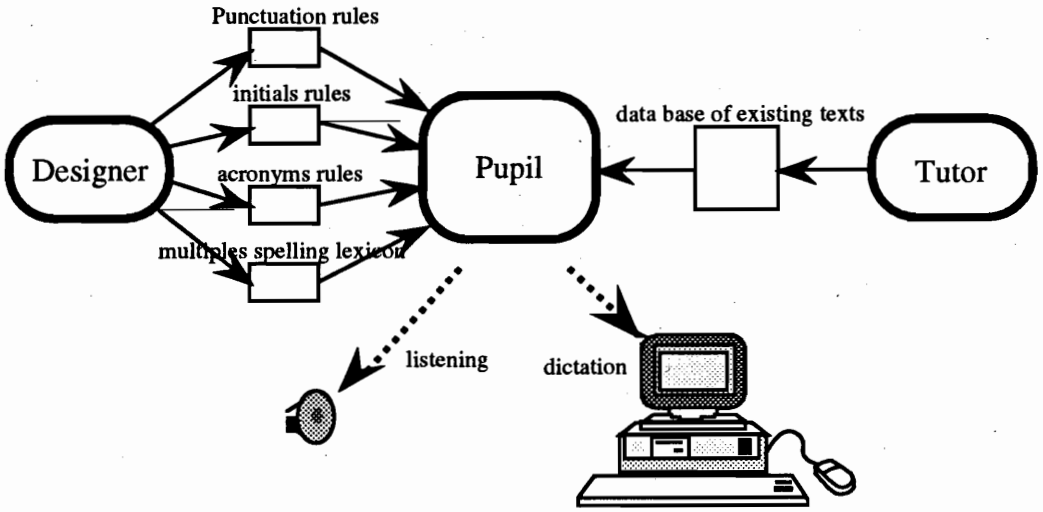


Figure 1: Architecture of the system

of dictations. An evaluation of the software, carried out with the pupils of a CM2[1] class, is then presented.

## 2. Structure and use of Ordictée

Ordictée is made up of three complementary modules: the *pupil* module, the *tutor* module and the *designer* module. The *pupil* module enables the pupil to carry out dictations. The *tutor* module aimed at the lecturer, allows them to form a collection of texts, adapting them in order to dictate them, whereas the *designer* module, aimed at the software manager, allows the management of lexicons and rule bases, giving rise to customised dictations and corrections (see figure 1).

**The pupil module.** The purpose of the pupil module is to act as a teacher for the different stages involved in dictation: reading, actual dictation and correction. Provided the tutor has designed the dictation in the correct way, Ordictée can (on request) offer linguistic explanations concerning the errors that are made. An average session with the pupil unfolds as follows: the pupil alters (should they wish to do so) the listening parameters (diction speed, pitch and volume of the voice).

They then select a dictation, listen to it at will, whilst looking at pictures illustrating the purpose of the text which reads.

They can then commence the exercise. Ordictée adapts itself to the pupil's rhythm, repeating the current phrase as much as necessary. The pupil can hear the dictation once more and carry out the necessary corrections. They then ask Ordictée to proceed with correction. Here, the various errors are revealed to the pupil.

The French vocabulary contains a group of about fifty words, which have, *out of context,* and for a given pronunciation and a given meaning, several correct spellings (in general, an advised spelling and one or several variants, such is the case with words like (clé, clef), (khôl, koheul, kohol)).

During the correction stage, and having spotted a fault, the pupil module consults

[1]Cours moyen de seconde année (the last year in French primary school).

a dictionary containing the various spellings for each word (see *designer* module), so as to check that the word thought to be incorrectly spelt does not have an alternative spelling.

**The tutor module.** The tutor module enables the lecturer to manage the dictations database. It also offers the option of going into the text in question in order to set markers informing the pupil module of particular pronunciations (initials, acronyms and heterophone words) or to section off the text into phrases to be repeated during the exercise.

**The designer module.** The designer module allows the management of a base of rules which are applied before the text is read. Three types of rules are permitted. They concern:

- initials,
- acronyms,
- punctuation.

During normal reading by an individual, or by a machine, punctuation is used to create prosody. Whilst reading the dictation, the punctuation must also be pronounced. The punctuation rules allow us to specify the pronunciation of each sign (?, !, "to the line", etc.).

The designer module also allows us to manage a dictionary of words, each of which has various spellings. This dictionary is used by the pupil module during correction.

## 3. Following the typing process

The objective of following the typing process is to assure synchronisation between the reading of a piece of dictation with the writing of a pupil. A second objective is to disrupt the pupil as little as possible, notably by avoiding pointless repetitions. Whereas in the traditional dictation, a range of information is available for the teacher to decide if the pupil has reached to the end of the typing, Ordictée is based uniquely on the information typed, on the speed the pupil types at and on the expected information. In a first version we founded a solution on an approximation of the number of characters given for deciding either to repeat the current piece of text or to proceed. This solution has rapidly shown its limits and we have implemented a solution of a heuristic nature which is based on hypotheses on the pupils' performance and on various measures performed during the typing. This solution is presented below.

### 3.1. Hypothesis on pupil performance

Following of the typing is based on the assumption that the pupils performance complies with the three following hypotheses:

1. The pupil stops typing
    - either because they think they have finished to type the current piece of text,
    - or because they do not know what they should write.

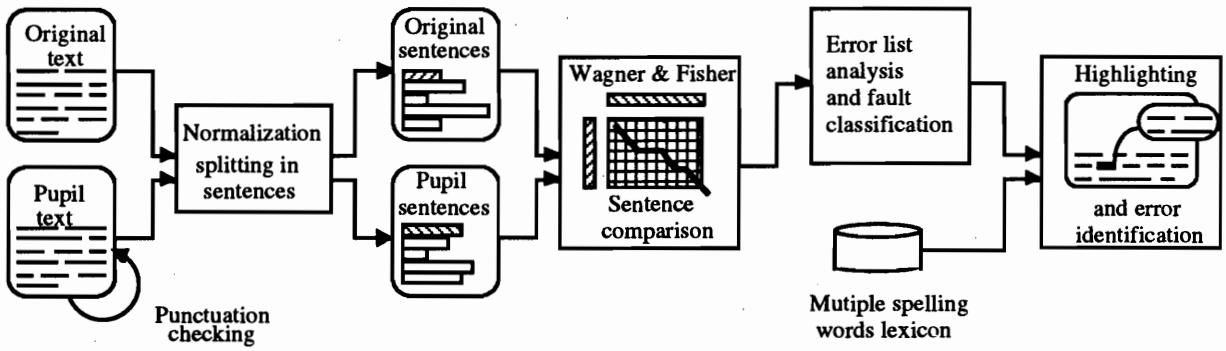


Figure 2: The various stages during the correction processing

2. The pupil who types knows what they should type. As the pupil types their text, it is therefore pointless to repeat the current piece.

3. A pupil can make a few mistakes but the insertion or forgetting of letters in a word is more probable than inserting or forgetting words. However it is not possible to found the following of typing uniquely on the counting of words.

### 3.2. The principal solution

Two decisions should be taken by Ordictée: "When should a piece of text be changed over ?" and "When should the current piece be repeated ?". The evaluation of these two conditions are based on the calculations of the following elements:

- the average speed of typing,
- the instantaneous speed of typing (which permits one to decide the inactivity of the pupil),
- where the pupil has not given up,
- the number of words remaining to be typed,
- where an audio output is still running.

The decision concerning the inactivity is based on the comparison between the speed of instantaneous typing and the average speed. The decision to withdraw is founded as regards to the inactivity of the pupil during several repetitions. The changing of the piece is then decided when the following condition occurs:

(enough words typed **and** inactive) **or** withdrawn)
**and not** audio output.

The decision of repetition is taken when the following occurs:

**not** enough words typed **and** inactive
**and** the audio output is finished
due to the timing (function of the average speed).

In a following version we envisage to enrich the solution presented below through using a comparison of typed text and of expected text through using a dynamic programming method (Gries 1988).

## 4. Spelling correction

When a dictation has been heard by the pupil, they can then ask the software to correct the mistakes. Ordictée has at his disposal the original text supplied by the tutor as well as the text issuing from the pupil. The correction procedure involves comparing both texts and thus finding the mistakes made by the pupil. This comparison is carried out by the Wagner and Fisher algorithm. We will go on to describe the principle of this algorithm and explain the different stages of correction.

### 4.1. The principal of the correcting algorithm

The principal difficulty of the correcting algorithm resides in the fact that the Wagner and Fisher algorithm produces a result in terms of comparing characters, whilst Ordictée should provide a word comparison. Beyond the simple application of the Wagner and Fisher algorithm, it is a question of having an algorithm at one's disposal which can be capable of interpreting an editing list of characters (deleting, adding and replacing of characters) in an editing list of words (misspelling, deleting, adding and replacing of words as well as merging or forbidding cutting of words).

### 4.2. The correction algorithm

The Wagner and Fisher dynamic programming algorithm (Du 1992, Stephen 1992, Wagner 1983) constructs a distance matrix between two strings of characters in order to determine the longest common sub-sequence with a minimum cost. The matrix is filled using the following recurrence formula:

$$D_{i,j} = \min\left(D_{i-1,j} + w(x_i, \varepsilon), D_{i,j-1} + w(\varepsilon, y_j), D_{i-1,j-1} + w(x_i, y_j)\right)$$

$D_{i,j}$ represents the distance between string $X$, with length $i$ and string $Y$, with length $j$. $w(a,b)$ is the cost of the substitution of character $'a'$, by character $'b'$. $w(a,\varepsilon)$ and $w(\varepsilon,a)$ are the respective costs associated with the deletion and insertion of character $'a'$.

The Wagner and Fisher algorithm seeks the shortest route (in terms of character substitution, insertion or deletion), which allows the string $X$ to be transformed into the string $Y$ and tells us which is the longest common sub-sequence. This sub-sequence holds little interest for us, but it is nevertheless interesting to scan the matrix along the optimal route, in order to locate faults. This enables us to make a list of index couples showing the differences between the pupil's text and the original. Having analysed this list, we are then able to distinguish between several types of mistakes, and to offer adequate explanations.

### 4.3. The correction stages

A dictation is broken down into a group of sentences which are delimited by the punctuation marks at the end ('.', '!', '?' or '...'). It is thus necessary, during correction, to compare the chunks which correspond to the same sentence so as to ensure that the explanations supplied are coherent.

The first stage involves checking the pupil's text for punctuation, so that their sentences complying with the original text. If the punctuation is found to differ in any

way from the original, the pupil has the opportunity to alter it. It is possible for them to hear the dictation once again, in order to correct their text.

During the second stage the pupil's text is normalised: this processing allows Ordictée to remove the leftover spacing characters and to adopt the same typographic conventions as the original version (for example, a comma is preceded and followed by a space). The pupil's text is split into sentences which will be individually compared to those in the original text.

The third stage makes it possible to find the eventual errors of the pupil. The complexity of the Wagner and Fisher algorithm is in $O(m.n)$, where $m$ and $n$ are the respective lengths of strings to be compared. The effective time for processing long sentences can be non-negligible. So, in order to limit it, the sentence is pre-processed: Ordictée retains the sub-sequence between the word preceding the first wrong word and the word following the last wrong word. Reducing the search space in that way generally brings about better results. These sub-sequences are placed in the comparison matrix and we obtain the list of the pairs of errors. Its analysis determines the type of error.

Finally, the last stage consists of informing the pupil of the fault locations and to provide them with an explanation. Thanks to the groupings of pairs, we can see easily which word is incorrect and its correct form. Highlighting the incorrect word in the pupil's text is a delicate matter, as the pupil has total freedom as far as the use of spacing of characters is concerned. However, thanks to the method employed to notice differences between the two texts, we can see exactly where the mistake is. At the same time, we can identify the kind of mistake. Figure 2 recaps on the stages that we have described above.

### 4.4. Categories of detected errores

The difficulty with the interpretation of results supplied by the Wagner and Fisher algorithm in the context of spelling correction is mostly due to the lack of semantic knowledge concerning the sentences. In actual fact, this algorithm is based on the notion of distance between two strings. If it helps to locate an error, it does not give enough information which permits the exact description of mistakes. In these conditions it is preferable to limit the number of categories and to find ways to make trustworthy decisions to class mistakes. At the moment the following categories are taken into account:

1. a word forgotten by the pupil,
2. a word which was not requested,
3. a word is badly spelt.

The detection of the first two categories ensures the synchronisation of the correction, even when the pupil has forgotten part of the sentence, or has typed a few words which have not been requested. The omission of a word is detected in view, within the correction matrix, of a series of character insertions in the pupil's sentence. These characters correspond to the word expected in the original sentence. Similarly, a word which was not requested is identified when a series of deletions occur.

The last category is very general, as it indicates just a spelling mistake. The error is highlighted and the correct spelling is shown to the pupil. When a word is classed as incorrectly spelt, a few simple tests are carried out to decide if the error is merely due

to the misuse of capitals and small letters, or if it is because we are facing a word which has several possible spellings.

## 5. Evaluation

In spite of the fact that over the past twenty years its importance has reduced, dictation remains a key exercise within the pedagogics of the French primary school education system. Moreover, it is increasingly common to find schools with a computer room, offering the pupils an opportunity to familiarise themselves with a computing environment, yet at the same time allowing them to do exercises which help them to learn the French language.

Given all these reasons, together with the fact that the computer scientist should not make a product commercially available without it being put to the test among its future users, we have decided to proceed with an evaluation of Ordictée based on the behaviour of pupils at a primary school. Below we will describe the protocol used, the main observations made and the synthesis of interviews resulting from the exercises.

### 5.1. Protocol

We have used a class of pupils aged 10-12 from Saint Roch school in Lannion. Prior to the experiment, each of the pupils from the class has had some experience with a computer during computing sessions (about half an hour per week on PC-DOS without a mouse). As far as Ordictée is concerned, several demonstrations, followed by tests carried out by some pupils in front of the whole class, took place a few weeks before the stated evaluation. Each session takes the following form: the location is in an isolated room, in front of an assistant and a secretary.

1. The assistant explains to the pupil how the program works and the role of each button.

2. There is a start up program, so that the pupil can get used to the buttons, the mouse and double function keys (some French accents) and the use of the keyboard.

3. A dictation (of about thirty words) is selected by the assistant and the pupil is left to complete the dictation at there own pace, whilst leaving the pupil the possibility of calling the assistant if necessary. However, for each error detected and presented by Ordictée, the assistant asks the pupil to explain their mistake.

4. A directed interview based on the evaluation of Ordictée ergonomics as well as on the quality of synthesis, diction and correction is finally carried out.

During the presentation of Ordictée to the whole class, we have noticed that unlike normal dictation, the pupil feels challenged by the machine: they try, using an iterative process of self-correction automatic correction, to produce a perfect text. Achieving a fault-free text is an objective embedded in point three of the protocol.

Let us note that, following the demonstrations, all the pupils in the class understand the concept of the software and volunteers are plentiful.

### 5.2. Observation during the sessions

Ten pupils, who according to the teacher, have unequal abilities, are chosen to participate in the experiment. Two types of observations deserve to be made here, concerning the handling of the software and the speech synthesis.

**Handling.** The main difficulties encountered by the pupils stemmed from their inexperience with the computer, especially editing the text: problems with placing the cursor accurately on the text, forgetting to click, etc. Nevertheless, we can see that they have rapidly familiarised themselves with theses procedures.

**Speech synthesis.** The excellent intelligibility of the speech synthesis results in it being at the root of few of the pupils' problems. However, when pupils comes across a word they are unfamiliar with, they do not appear able to ask the advice of the teacher, as would be the case in the classroom. This can result in misunderstandings and thus, mistakes.

### 5.3. Interviews

A "direct" interview is held, following the exercise. It concerns the general ergonomics of the software, speech synthesis, diction and correction.

**General ergonomics.** On the whole, the software is considered to be well constructed and practical to use. The pupils feel at ease.

**Speech synthesis.** The opinions here are mixed. They range from "good", to "an artificial voice". An exercise in order to let people get used to it could have been carried out prior to the experiment. On the other hand, research in speech synthesis is leaning towards a wider variety of voices (for example, woman's voice/man's voice, an advertising voice, or an airport voice). Maybe a teacher's voice would be a good idea.

**Correction.** The pupils were happy with the correction.

**Preferences.** One of the purposes of the interview is to evaluate the difference between Ordictée and traditional dictation. In answer to the question, "What difference do you note between this type of dictation compared to that of the teacher?" two different answers emerged: "With the teacher we can ask questions" (meaning that if we are stuck the teacher can help us). And on the other hand "We have a good control of the exercise" (meaning that we can manage our time without being dependant on the teacher).

## 6. Conclusion

Ordictée is a kind of software which allows pupils to practice dictation. The correction, carried out in one of three modules is based on an optimal comparison of strings using a dynamic programming algorithm. An evaluation of the pupil module has been carried out. Although at the moment only the French version is available, the technique used to carry out the correction of the dictation authorises practically immediate transport towards languages for which speech synthesis does exist. A version in Breton language is also planned.

## References

D. Cotto, M. De Calmes, I. Ferrané, J.F. Malet, J.-M. Pécatte, G. Pérennou, C. Santiago. Usage didactique de produits de l'Industrie de la Langue : Un Système interactif de

dictée/correction pour l'apprentissage autonome du français. *ERGO-IA,* Biarritz, 19-21 septembre 1990, p 121-129.

D. Cotto, M. De Calmes, I. Ferrané, J.F. Malet, J.-M. Pécatte, G. Pérennou, C. Santiago. An Automatic Interactive Dictation and Correction System. Application to Language Teaching/Learning. *Proceedings of COGNITIVA 90.* AFCET, Madrid, Spain, p. 419-426. November 20-23, 1990.

M.W. Du, S.C. Chang. A model and a fast algorithm for multiple errors spelling correction. *Acta Informatica 29, 281,* p. 281-302. Springer-Verlag 1992. D. Feneuille, J.-C. Fontaine. Synthèse vocale et lecture. EA007 Cap d'Agde, ADI1987 Paris.

J.P. Fournier. Correction automatisée dans les systèmes questions- réponses en langage naturel. *CIIAM86, 1-5 décembre 1986,* Marseille. Hermès. 1986.

D. Gries, B. Burkhardt. *Presenting an Algorithm to find the Minimum Edit Distance.* TR 88-903. Cornell University. March 1988.

G. A. Stephen. *String Search.* TR-92-gas-01. School of Electronic Engineering Science. University College of North Wales. 1992.

R.A. Wagner. *On the Complexity of the Extended String-to-String Correction Problem. Time Warps String Edits, and Macromolecules: the theory and Practice of Sequence Comparison.* Addison-Wesley, 1983.

**Acknowledgements.** The authors would like to thank C. Delessalle, M. Gaillard, S. Mironnet, L. Egan, Primary school inspecteur at Lannion, and G. Janin, director of Saint Roch school at Lannion for their contributions to the project. This project was partly funded by ANVAR BRETAGNE.