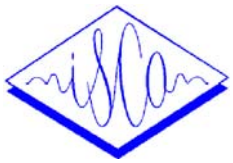

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

VOCAL SYNTHESIS IN A COMPUTERIZED DICTATION EXERCISE

Conception Santiago-Oriola
IRIT, Université Paul Sabatier
118, route de Narbonne, 31062 Toulouse Cedex 4, France
coriola@irit.fr
http://www.irit.fr

ABSTRACT

Dictation can be considered as a transcription exercise of an utterance. The main difficulty for the elaboration of a dictation system consists in modeling the errors and the associated explanations provided to the learner. On these bases, an experimental DICTOR system is being developed as an assistant tool to learn the French language spelling. DICTOR includes an automatic checking tool based on a stochastic alignment algorithm and French written linguistic knowledge. This paper focus on both aspects. We discuss the spelling learning issues due to the no one-to-one correspondence between the utterance and the written text. Then the principles of the spelling difficulty groups (sdg) and associated explanation rules are presented. Finally, we report an observation of the DICTOR system in a school environment for the French language spelling. We then analyze how the different vocal synthesis influence the exercise and are perceived by the pupils in this new generation of language learning systems.

Keywords : language learning, dictation, spelling learning, speech synthesis.

1. INTRODUCTION

Research on spoken and written language engineering offers nowadays reusable tools and linguistic knowledge (such as lexicons, alignment tools, etc.) for the development of new interactive applications about language. At the same time, multimedia tools integrating vocal synthesis systems emerge. This context and the linguistic knowledge maturity have led our research team to develop an interactive vocal dictation system in the framework of language learning.

The main difficulty for the elaboration of such a system consists in modeling the errors and the associated explanations provided to the learner.

On these bases, an experimental DICTOR system is being developed at IRIT as an assistant tool to learn French language spelling. DICTOR includes an automatic checking tool based on a stochastic alignment algorithm and French written linguistic knowledge. This paper focus on both aspects.

The spelling learning problems due to the no one-to-one correspondence between the utterance and the written text are discussed. Then the principles of the spelling difficulty groups (sdg) and associated explanation rules modelization are presented. Finally, we report an observation of the system with two different speech synthesis systems in a school environment.

2. THE DICTATION EXERCISE

In language learning, one of the basic exercises is dictation. Dictation can be considered as a transcription exercise of an utterance. This task rises several multidisciplinary problems from aural signal perception to its interpretation.

Firstly, the learner must recognize the spoken units such as words, phrases or sentences as part of the language being learned.

Then, the sounds heard must compose a series of words. This word segmentation can be a problem for learners because the spoken language has no one-to-one and clear acoustic parameter indicating the beginning or the end of words. Moreover, some liaison phenomena, which appear at the adjunction of a phoneme on a lexical frontier, or assimilation ones between phonemes due to adjacent words intervene. The oral segment can then correspond to several homophonic non homographic phrases. For example, [ləpətitami] can lead to ***le petit ami*** (the boyfriend) or ***le petit tamis*** (the small sieve). Another penalizing factor is the presence in the utterance of words or expressions not belonging to the learner's lexicon. The adaptation of the dictation text to the learner's cognitive and learning levels is essential for this phase.

Thirdly, once the series of words is determined, it must be meaningful. The terms activated into memory must be in a near relation with the theme of the dictation. In French, words such as ***saule*** (willow), ***sole*** (sole), ***sol*** (ground) can be pronounced [sɔl] depending on the speaker's region or the language acquisition level. This phase can lead to confusions in the following terms. To distinguish between homophonic words, the context is fundamental.

The latest phase consists in the effective transcription according to the letters provided by the language being learned and to the learner's conceptual model of the spelling. The spoken language can be very influent: the transcription can be biased by the learner's pronunciation related to the spoken dialect.

Those various problems show the need in terms of linguistic knowledge to efficiently model the transformation of an utterance in a written phrase.

3. AN INTERACTIVE DICTATION SYSTEM

To specify an interactive dictation system using vocal restitution systems, we differentiate two principal phases in the dictation exercise. The first one consists in the learner's typing of the spoken dictation text, through the keyboard of the computer, followed up by the diagnosis phase (Figure 1). Interactivity, explanations pertinence and diagnosis adaptation to the learner's knowledge level are the strong criteria for the

design of this interactive system.

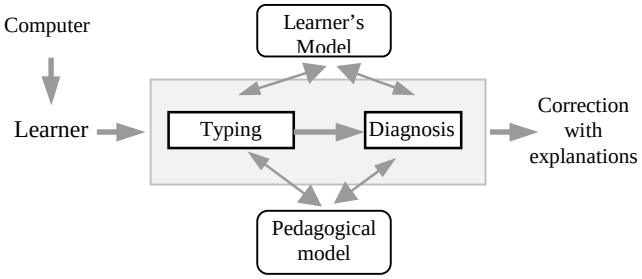

Figure 1. Synoptic of an interactive dictation system

As shown in Figure 1, the methods, used by the teacher, to design the exercise to do for example, and the learner's model are taken into account to permit a differentiate pedagogy. Here, the first step of our system (called ***typing***) is a dictation from the computer to the learner. The dictation text is enunciated by a speech synthesis system and the learner transcribes it via the computer keyboard.

After this phase, the second step consists in evaluating the learner's input for proposing him some explanations depending on the errors committed and the learner's spelling knowledge level. This evaluation is based on a string alignment algorithm between the correct dictation and the learner's input and on different corrective strategies (see Figure 2).

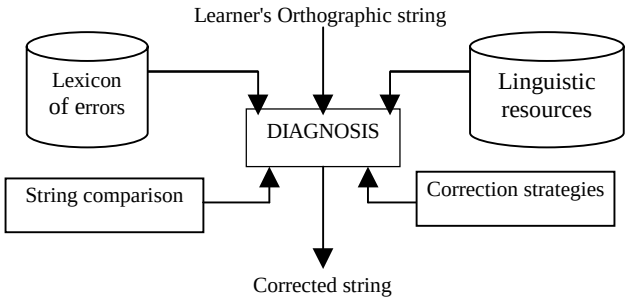

Figure 2. Second step of the dictation.

The diagnosis module generates an annotated string corresponding to the corrected string, completed with the explanations. To implement it, a lexicon of the learner's possible errors and a set of linguistic resources, containing among other data the explanations, are necessary.

3.1. String comparison algorithm

Our goal is to compare the transcript made by the learner, on an orthographic format, and the dictation correct text to induce the eventual errors with string alignment methods.

An alignment algorithm VERITEXT [2], derived from the VERIPHONE system (which realizes a phonetic alignment), based on rules has been developed. It relies on a stochastic model described in [3]. In our case, VERITEXT aligns the correct dictation with the learner's input. The spelling coder takes into account the spelling errors while as the typographic channel models the typographic ones such as insertion, deletion or the typing errors due to the keyboard.

Nina Catach has proved that the basic unit for the French spelling learning is the grapheme and particularly the phonogram, i.e. the grapheme transcribing a phoneme [1]. In the text correction framework, a new entity based upon phonograms has been defined: the **s**pelling **d**ifficulty **g**roups (noted ***sdg***) [3]. Without detailing the sdg, the six sdg rewriting rules can process the erroneous text ****les anfant*** written instead of ***les enfants*** (the children). Each rule can be decomposed, as indicated before, in:

sdg : correct rewriting probability{, erroneous rewriting probability}*

where the number of erroneous rewritings depends on the learner's model and learning stage.

sdg	Correct rewriting		Erroneous rewriting		Erroneous rewriting		Erroneous rewriting	
les	les	0.95	le	0.05				
#	_[1]	1						
en[ã]	en	0.95	an	0.05				
f	f	0.95	ph	0.05				
ant]	ant	0.85	ent	0.05	an	0.05	en	0.05
sPL	s	0.95	Λ	0.05				

VERITEXT makes an alignment and points out if an erroneous rewriting has been used.

Correct dictation	les	#	en[ã]	f	ant]	SPL
Learner's input	les		an	f	ant	
Error (Y/N)	N	N	Y	N	N	Y

VERITEXT must be associated with pedagogical strategies to be able to provide adapted explanations to the learner. On the example above, we may precise, for instance, the bad spelling of the term ***enfant*** and the absence of the plural mark.

3.2. Correction strategies

The correction strategies allow the system to choose the kind of presentation of the alignment results and linguistic knowledge. They may fit to the learner's knowledge level and cognitive abilities and to the teacher's pedagogical methods. All this information cooperates with the learner's model and the pedagogical module.

4. MODELIZATION OF THE ERROR DIAGNOSIS KNOWLEDGE

We focus here on the second step of this exercise, i.e. the diagnosis of the correctness of the learner's input. Our concern is multiple: how to determine the more precisely the learner's errors and how to present the results of our diagnosis.

To answer satisfactorily to our first concern, we must analyze the linguistic knowledge of the French language to settle down its potential difficulties. A modelization of the French language into **g**roup of letters with **s**pelling **p**roblems (sdg) and associated rewriting rules is here presented.

The second concern relies on corrective and explanation presentation strategies which may be adapted to the learner's capacities. Some explanations are related closely to the rewritings and therefore to the sdg.

4.1. Graphic vs. Phonetic material

As defined above, dictation is a "graphic" transcription process of utterance. In this paper we won't consider the comprehension problems in the neurolinguistics sense.

So dictation is only considered as the establishment of a correspondence between speech and text. To do it, the orthographic alphabet must be faced to the French phonetic alphabet. If we consider those two alphabets, some facts appear:

- the orthographic one has only six vowels (***a, e, i, o, u, y***) while the phonetic one has generally sixteen [a], [ɑ], [e], [ɛ], [ø], [ə], [œ], [i], [o], [ɔ], [y], [u], [ã], [ɛ̃], [ɔ̃] and [œ̃].
- There are twenty consonants in the graphic one but ***c, k*** and ***q*** correspond to the same phoneme [k].
- The letters ***c*** and ***g*** each have two different phonic values, [k] and [s] –for example ***café*** [kafe] (café) et ***cerise*** [səʀizə] (cherry)– [g] and [ʒ] –as in ***garçon*** [gaʀsɔ̃] (boy) and ***girafe*** [ʒiʀafə] (giraf)–.

Catach has demonstrated that the foundations of the French written language are phonogramic, that is its basic unit is the grapheme and particularly the phonogram, i.e. the grapheme transcribing a phoneme [1]: 80 to 85% of the characters of a text correspond to sounds.

So we can see that the French spelling has many difficulties to overcome and therefore is complex to learn and master.

4.2. Identification of the sdg

During the dictation exercise, the learner must write down graphemes of the language being learned from the meaningful phonemes he can hear.

Errors can have varied origins such as for example:

- He can make bad phonogram transcriptions because either he doesn't know them (incomplete lexicon) or his spelling knowledge is incomplete (for example, [faʀmasi] ****farmassie*** instead of ***pharmacie*** (drugstore)),
- The relations between grammatical categories aren't respected (*ils passe instead of ils passent (they pass)),
- The learner doesn't know some words or segments the text in an unexpected way in relation with the original text [œ̃nevje] can lead to *un névier instead of un évier (a sink)),

So, the sdg alphabet must allow to model all the possible errors generated by the phoneme-to-grapheme shift and to model the phonetic and phonologic errors due to an incomplete knowledge of the lexicon, syntax and morphosyntax of the French language. Moreover, this sdg model must be opened to consider different pedagogues' strategies as already mentioned. For example, we must define correctly the sdg to render the difference between errors such as ****troi***, ****quatre plat***, ****tu passe*** where the left final ***s*** can't be justified in the same way:

- Trois (three): this adjective must be always written in that way.
- quatre plats (four dishes): the ending s is the morphogram of the plural of the noun plat,
- tu passes (you pass): here, the ending s is the morphogram of the concord of the verb passer et the second singular person.

This example shows that, at least, three sdg must be defined corresponding to the unpronounced ending ***s***.

To constitute our alphabet, we have studied each of the 172 graphemes of the French language, adding them complementary information (word position and pronunciation) to obtain a first list of sdg. Then we complete it with morphograms and logograms to allow the teachers to adapt the explanations to the learners. We have currently a set of 564 sdg for the French language learning.

4.3. The associated rewriting rules

To detect the learner's eventual errors, the erroneous possible rewritings of each sdg must be identified. The principal problem is to attach to each sdg the rewritings set, those corresponding to different transcription possibilities more or less evident depending on the learning level. The learner's possible behaviors must be predicted to adapt the necessary explanations of the exercise. Reviewing each sdg category can help us to settle some of the problems.

The difficulties concerning the rewriting rules associated to phonogramic sdg are particularly linked to the vowels. The French beginner learner can have some problems with the liaison phenomenon. The liaisons can prevent him from recognizing the words starting with a vowel, he can then write, erroneously, ****un navion*** or ****des zavions*** instead of ***un avion*** (a plane) or ***des avions*** (planes). We have added some sdg to manage this liaison phenomenon as:

Sdg	Correct	Error
z"	Λ	z

Table 1. A liaison rewriting rule.

So, all the difficulties met by learners must be instantiated to achieve a complete modelization of the French language. From a typology of the spelling errors covering all the French language learning we have generated the associated rewriting rules.

For a complete description of the sdg and the associated rewriting rules, please refer to [4].

5. PRESENTATION OF THE DICTATION DIAGNOSIS

After modeling the French language to diagnose exactly the learner's input, the mediated dictation exercise has to present the results to the learner with clues allowing him to better his spelling by acquiring the lexical terms and the grammar rules, for example. Those results, which can be presented on different forms, need to fulfil the pedagogical goals. As shown in figure 3, we model this phase into two stages: the indication and the explanation of the errors. As we want the presentation to meet the pedagogical methods and to be learner's adapted, the respective models appear in this stage.

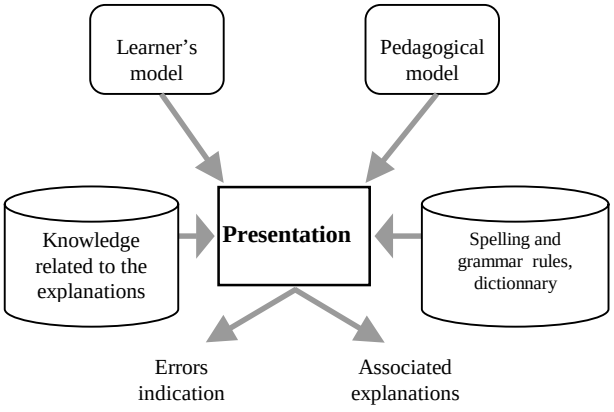

Figure 3: Knowledge related to the diagnosis presentation.

6. DICTOR

On these bases (modelization of sdg and associated explanation rules and communication between knowledge), we've implemented an interactive dictation system, called DICTOR. The dictation is read out to the learner through a speech synthesis and the system corrects the learner's input. The diagnose and the presentation of its results are implemented too and offer the learner explanations on two different levels: the first only shows the erroneous word with a short description of

the kind of mistake (spelling, typographic) and the second, asked by the learner if misunderstanding remains, proposes a more complete explanation on the spelling, lexical or grammatical rules applying to the term in that context.

7. OBSERVATION IN A SCHOOL ENVIRONMENT

The introduction of a system such as DICTOR in a school environment can induce some additional difficulties to the exercise. These can be due to the new situation or to the new tool which is the vocal synthesis system. To evaluate their influence on the exercise, we confronted DICTOR to the end users: the 24 pupils of a 5^{th} level classroom. The observation run during three weeks. Three dictation texts were selected by the teacher to have input duration limited to twenty minutes. We split the classroom into three groups and had each one work into a different situation each week:

- a Sample group, who worked as usually with pen and paper,
- a TELEVOX group who tested a DICTOR system coupled with TELEVOX, a text-to-speech French system based on a TD-PSOLA algorithm and
- an ECHOVOX group who tested a DICTOR system coupled with ECHOVOX, a digital audio system.

To evaluate the perturbation induced by the DICTOR system we had two hypotheses: the spelling mistakes of the DICTOR groups are different from the ones of the Sample group and the DICTOR learners make more mistakes. In table 2, we compare the means relative to the number of words of each dictation (the number of words of each dictation is between parentheses). We note that for two of the three dictations (first and third), the relative means between the DICTOR and the Sample groups are similar but for the second dictation, the relative mean is clearly inferior for the DICTOR group. This can be due to the DICTOR pedagogical module which indicates the spelling mistakes during the learner's typing so he can correct the word. It seems that the contribution of the pedagogical module chosen is non negligible faced to the traditional dictation exercise.

The other point concerns the spelling mistakes made. If we look at them, we notice that the same kind of mistakes appear for the three groups. So DICTOR doesn't seem to disturb the dictation exercise. For a complete description of the spelling mistakes made, refer to [4].

The influence of the kind of vocal synthesis on the dictation can be given by the number of times that the learner listen to each utterance of the dictation That listening-in is linked to the material used (similar for the two DICTOR groups) and to the words composing the utterance itself. If we compare the number of listening-in for a given utterance, it reflects the intelligibility of the speech synthesis.

	Sample		DICTOR	
	Pop	**Mean**	**Pop**	**Mean**
First dictation (11)	7	0,19	16	0,20
Second dictation (19)	4	0,31	11	0,16
Third dictation (41)	9	0,11	10	0,12

Table 2. Relative mean of the spelling mistakes made.

Two of the three dictations show, in table 3, that the use of the text-to-speech synthesis needs more listenings than the recorded voice system but the tendency has been inverse for the third one. The difficulty of the third dictation text has lead the learners to try to comfort themselves by listening.

Listening mean	**1st dictation**	**2nd dictation**	**3rd dictation**
DICTOR/TÉLÉVOX	3,08	3,37	2,71
DICTOR/ECHOVOX	2,34	2,89	3

Table 3. Listening Mean for the three dictations according to the speech synthesis used.

The means for the two synthesis only show a slight difference between them (0,31) when we expected a bigger one due to the mechanic character and the lack of prosody of the synthetic voice of the text-to-speech synthesis.

On the other side, the synthesis perception shows that the learners preferred the digital system in terms of listening comfort (82%) (the text-to-speech ranked 37,5%) even if the dictation task is not affected by this fact.

We asked the children for their opinion on the system in an open questionnaire. The system had a very good acceptation. Sixty-six per cent of them want to use the system at school or at home. Even if the computer attracts them, the learners find that DICTOR is an interesting way to do an exercise that 80% of them don't like.

8. CONCLUSION

We have presented a model of the French language devoted to the dictation exercise. Its pertinence has been put into relevance by the observation made in real conditions in a school environment.

Through this observation of two DICTOR versions, comprising each one a different kind of speech synthesis, we wanted to answer some questions relating to the ergonomic quality of the DICTOR system in terms of usability for the potential end users. The introduction of an interactive system as DICTOR in a school environment arises questions on the adaptation to the users and the users' appropriation of the tool. During the observation, we noted how the learners appropriated DICTOR by using all the possibilities offered.

Another important element of the system is the speech synthesis. We wanted the observation to help us to decide the kind of synthesis to use in such a system. But through the data collected we can only consider that the text-to-speech system and the recorded voice system are equivalent. The use of one or another will only be related to material constraints.

However, we must validate it on a larger period, with a larger population at different language learning levels. The genericity of the model allows us to extend it to other languages and learners: French Second Language for Arabic learners in collaboration with the Institut National d'Informatique d'Algiers (current AUPELF project), English as a Second Language with J. Malet (Sacramento CA).

NOTES

1. "_": word separation character; ant]: the letters ***ant*** end a word; sPL: plural morphogram ***s***; "Λ": the void character.

REFERENCES

[1] Catach N. 1995. *L'orthographe*, Coll. Que sais-je ? 685.

[2] Pécatte J. M. 1992. *Tolérance aux fautes dans les interfaces homme-machine, traitement des chaînes phonétiques, des chaînes orthographiques et des structures syntaxiques*, Ph.D. Toulouse III, January 1992, 25-74.

[3] Pérennou G., Daubèze P. and Lahens F. 1987. Automatic checking and correction of texts: spelling and typing errors; a model: VORTEX, *Technology and Science of Informatics*, vol 6 n°1, 49-67.

[4] Santiago-Oriola C. 1998. *Système vocal interactif pour l'apprentissage des langues - la synthèse de la parole au service de la dictée*, Ph.D., Toulouse III.