論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	A French liaison learning system using TTS for Japanese students
著者(和文)	古井 貞熙
Authors(English)	Josafa de Jesus Aguiar Pontes, Sadaoki Furui
出典(和文)	日本音響学会2011年春季講演論文集, , No. 2-4-1, pp. 1579-1582
Citation(English)	, , No. 2-4-1, pp. 1579-1582
発行日 / Pub. date	2011, 3

A French liaison learning system using TTS for Japanese students

○ Josafa de Jesus Aguiar Pontes, Sadaoki Furui (Tokyo Institute of Technology)

1 Introduction

French is a language with major pronunciation variations at word-final consonants (WFCs). The pronunciation of these graphemes may depend on the context of the neighboring words or the morpho-syntactic roles played by them. It may even depend on the intended meaning. Second language learners quite often struggle to learn when to pronounce them.

To alleviate this problem, we propose a pronunciation Computer-Assisted Language Learning (CALL) system using TTS, with focus on liaison in French. Liaison is the pronunciation of a WFC when the following word starts with a vowel, a mute 'h' or some glides [1]. The graphic consonants of the linking words are -c, -d, -g, -n, -p, -r, -s, -t, -x and -z, which can be illustrated via the words/tokens "blanc", "grand", "long", "bon", "trop", "premier", "très", "petit", "deux" and "chantez" in different contexts.

Liaisons are classified into Obligatory (Ob), Forbidden (Fo) and Optional (Op) contexts [2]. Obligatory means that the WFC must be pronounced. Forbidden means that the WFC must remain silent in a context. And Optional means that pronouncing or not the WFC depends on the word bigram, language style, level of education of the speaker, among others [3]. Given the complexity involved in predicting the pronunciation of the Optional liaisons, they are left out of scope of this investigation. We target only Obligatory and Forbidden contexts.

In a previous study, we have investigated the prediction of the phonetic realizations of WFCs in contexts [4]. We have developed a method for creating a database for French WFCs, for predicting their pronunciations and for classifying their contexts. Based on this method, we developed a grapheme-to-phoneme (G2P) converter, which performs the prediction of pronunciation of the WFCs more correctly than a commercial system. In terms of liaison prediction, our G2P engine obtained comparable performance to the state-of-the-art systems.

We use this engine to produce phonological information for our CALL system. The audio is produced by inputing our G2P conversions to one of the best-performing commercial TTS synthesizer [5]. We assess Japanese students' learning performance after using our system and we compare it to a traditional learning method on the liaison subject.

2 Our liaison learning system

We start from the point in which we have already built an engine that is able to convert French text into sequences of phonetic symbols, as well as produce classification of WFC contexts (Ob, Fo and Op). To introduce our CALL system, we start with the integration of our G2P engine with the commercial TTS synthesizer. Next, we present a couple of functionalities of our system. We finalize this section by presenting how our system gives lessons about liaisons to students.

2.1 Integration of our G2P engine with the commercial TTS synthesizer

In the case of a pedagogical application, very high quality is required for the synthesized speech. Instead of relying completely in the closed solution, we integrate our G2P engine with the commercial TTS synthesizer. We use the G2P conversions produced by our own engine. This allows us to control the phonetic sequence to be synthesized. In other words, we have flexibility to fix incorrect G2P conversions, and therefore, room for improvements.

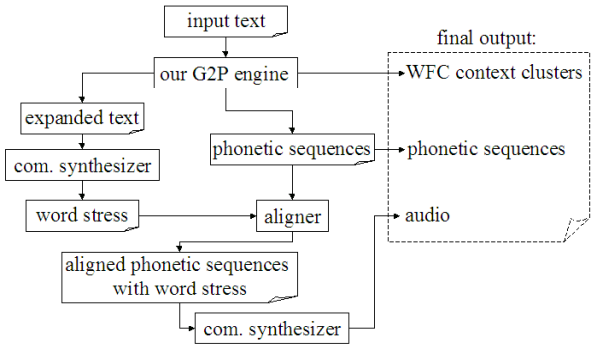

Fig. 1 Communication between our G2P engine and the commercial TTS synthesizer.

In order to produce speech, the commercial TTS needs phonetic sequences accompanied with word stress information. Our G2P engine, however, does not predict the locations of word stress in a sentence. In French, the position of word stress is variable. A word being uttered in isolation has its own stress, but this stress may disappear when the word is uttered in a context. For example, the word "petit" ("little") is pronounced as [p(ə)t`i] in isolation, where the stress is at the vowel [`i]. But in the context "petit enfant" [p(ə)titɑ̃f`ɑ̃] ("little child"), the stress is placed only on the final vowel of the second word [`ɑ̃]. In our case, we use the word stress information generated by the commercial TTS synthesizer instead, at the cost of alignment.

Given an input text, we use our G2P engine to expand it, as shown in Fig. 1. Numbers, abbreviations and compound words are segmented in order to allow the alignment of word stress with our phonetic sequences. Next, we input the expanded text into the commercial TTS synthesizer. This system produces the stress information and the phonetic sequence. The boundaries of each word/token in the phonetic sequence are delimited by a special character. Most of the time, the phonetic word/token boundaries produced by this system correspond to the same segmentation produced by our G2P engine. In this case, the alignment is straightforward. However, in the case of compound

日本人学生のためのTTSを用いたフランス語のリエゾン学習システム
ポンテス・ジョ ザファ・デ・ジェ ズス・アギアル、古井貞熙(東工大)

words, a couple of alignment rules are applied in order to solve the misalignment issues. When the alignment is finally obtained, we input both phonetic sequence and locations of stresses into the commercial TTS synthesizer to produce the final audio.

2.2 Functionalities of our CALL system

Our system allows a user to input French text and have it converted into speech. In addition, it outputs the written phonetic representation of the speech. This allows students to compare the spoken sounds with its corresponding phonetic sequence. That is to say, the visual and audio stimuli work together, allowing the students to identify the sounds and compare them with the similar ones. Another advantage of outputing the phonetic sequences is that, even if the synthesized speech is not perfect, the student can still learn the written pronunciation and realize any mismatch. That is why it is very important to provide correct G2P conversions for the students.

The focus of this investigation, however, is on teaching classification of Obligatory and Forbidden liaison contexts. To do so, our CALL system highlights the WFCs in different colors, according the context classification in which a word falls into. Green color indicates that the pronunciation of a WFC is Obligatory. Red color means that the consonant should not be realized in that context. Orange color is used when the pronunciation is Optional. Our system can classify only simple Optional cases though.

One novelty of our system is that it *displays* to the students which liaison cluster a word/token in context falls into. Depending on the context, the pronunciation may vary. Thus, the student can try to input the same word in different contexts to see how the pronunciation is affected. This makes the study of French pronunciation interactive and dynamic, given the flexibility of our G2P converter to process any French text.

The student can check the pronunciation of a word ending with a final consonant by highlighting the word and clicking on a play button. In this case, the system processes the highlighted word in isolation and plays its audio. In case of a word in an Obligatory liaison context (indicated by a green final consonant), the final consonant of the isolated word is not pronounced, whereas the one in context is. This feature is another novelty of our system, which allows the student to study the pronunciation variability of the WFCs by *listening* to different pronunciations of the same word.

Naturally, the same highlighting mechanism works for a longer sequence of words. The highlighted text is processed independently and the user can listen to only a small chunk of speech of a sentence. This functionality is particularly useful when the student is not familiar with the pronunciation of a certain word. Supposedly, the presence of an unknown word in a sentence prevents the student from learning the pronunciation. Listening to the pronunciation of such a word in a context is not the same as listening to its pronunciation in isolation. Listening to a segment of audio containing a corresponding unknown word is unlikely to help students because of co-articulation effects surrounding this word. However, when this word is synthesized in isolation, the student can listen to its pronunciation without distortion caused by the influence of the phonetic context.

2.3 Lesson about liaison

Our system provides a theoretical explanation about liaisons. It contains the definition of liaison, a list of possible graphemic consonants, its context classifications and respective examples. All examples can be played. The student can read these explanations in Japanese and listen to the audio for each example.

In addition, each category of context unfolds a couple of representative liaison rules. For the Obligatory contexts, 5 main rules are explained. A brief summary of them is shown in Table 1. This lesson also contains 5 main rules for Forbidden contexts (cf. Table 2) and 2 other rules for Optional contexts (cf. Table 3).

Each rule is displayed in the shape of URL links. The naviagation accross the rules allows a student to consult further details about each rule and search for additional sentence examples following the same rule. This functionality is likely to help the students to understand the rules more concretly, and have more chances to practice them. Furthermore, the user can modify such examples as he/she likes and have the modified text processed. By doing so, it becomes easier to understand how different contexts affect the pronunciation of the WFCs.

Table 1 Main Obligatory liaisons rules.

Articles, adjectives, demonstrative adjectives, possessive adjectives or numeral adjectives followed by nouns. Ex. “les _ amis”.
(Subject / direct object / y / en) pronouns followed by verbs. Ex.: “nous _ avons”.
A verb being connected with a vowel initial pronoun through a hyphen. Ex.: “Dit_-il ?”.
Monossilable prepositions following vowel-initial nouns, pronouns, articles. Ex.: “en _ Asie”.
Fixed expressions. Ex.: “pas _ à pas”.

Table 2 Main Forbidden liaisons rules.

Word bigrams belonging to distinct rhythmic groups. Ex. “comment | étudier”.
After the conjunction “et”. Ex.: “et | alors”.
Before words starting with an “aspirated h” or similar behavior. Ex.: “des | haricots”.
Single nouns followed by adjectives. Ex.: “pays | africain”.
Verbs being connected with pronouns in question or imperative sentences. Ex.: “Vont-ils | arriver?”.

Table 3 Rules for Optional contexts.

The words “pas” or “mais” followed by a vowel-initial word, when such a bigram is in the same rhythmic group. Ex. “pas [z] encore ”.
The words “devant”, “après” and “depuis” followed by a vowel-initial word, when this bigram is in the same rhythmic group. Ex. “devant [t] eux”.

The information we present in this lesson is compatible with the explanations found in French grammars for Japanese students. Studying from our system, however, is likely to be more interesting or effective than reading this information from a book. The

student is allowed to *interact dynamically* with our system by inputing his/her own text and by listening to the speech samples. This is not possible with a closed media, i.e. a book or a static E-learning method.

3 Pedagogic evaluation of our CALL system

An E-learning method for Japanese students of French has been proposed by TUFS [6], as part of Global COE Program [7]. Students using this system can study liaison by reading the explanations and by listening to prerecorded pronunciation samples. We call their method as “T system”, while we name ours as “P system”. Both systems have approximately the same Japanese textual explanations about liaisons.

Basically, our evaluation consists of comparing these two systems. T system cannot process a text input by the user, whereas P system is enabled to do so. However, P system may produce about 5% of Obligatory liaison errors, whereas T system is error free, given that the audio has been previously recorded. Which system would the students choose for studying liaison? Which system is more interesting and useful? Which system produces better learning performance? These are some of our evaluation points.

3.1 Evaluation method

We have developed an evaluation system, which controls a couple of tasks assigned to each student. There are nine tasks in total, as follows:

Task 1: A student starts by registering him/herself.

Task 2: He/she answers a *background questionnaire* which inquires about his/her French language learning history. The information we get out of it should be useful only for identifying groups of students having similar characteristics.

Task 3: A *pretest* is presented, which tests the amount of previous knowledge about liaisons

Task 4: Either *P* or *T system* is introduced first to a student. He/she uses one of the liaison learning systems for the study.

Task 5: The student answers a *liaison exercise* with the support of the system which has been introduced first.

Task 6: He/she answers a *posttest*.

Task 7: The *second liaison learning system* is introduced.

Task 8: *Incorrectly answered questions* of the pretest and posttest are displayed. The student has the chance to correct them by consulting any of the learning systems.

Task 9: He/she answers a *subjective questionnaire*, which compares both systems and allows him/her to express his/her impressions.

Our evaluation is divided into two parts: objective and subjective ones. The first part targets objectively finding which system is more efficient in teaching liaison. Tasks 3 to 6 are used the objective part. The second part collects the opinion of the students about both systems. Tasks 7 to 9 address the subjective part.

The *pretest* contains three types of problems. The first one tests the student’s capacity of classifying liaison contexts (Ob, Fo or Op). In the second problem, there are 4 questions. For each one, the student has to choose a phrase out of 4, where liaison is strictly Forbidden. As for the third problem, there is one or more phrases where liaison is not allowed. The student needs to find one or multiple phrase(s), out of five options. In order to avoid random answers, the students are advised to answer "I’m not sure" when the correct answer is unknown. Incorrect answers are “supposedly” penalized with negative scores.

After finalizing the pretest, a half of the students uses P system and the other half uses T system. The students read the Japanese explanations and listen to the French audio samples. Next, they start a training session, where they answer liaison exercises. In the posttest, the types of problems and the distribution of liaison rules are the same as in the pretest and the liaison exercises. Nevertheless, the words of the sentences and their order are not exactly the same.

Table 4 Questions of subjective questionnaire.

No.	Question
1	How intuitive is the navigation of T system and P system?
2	How interesting is T system and P system?
3	How much have you learned from T system and P system?
4	How much can you learn from T system and P system?
5	How much can T system and P system help you to practice French pronunciation?
6	How useful is T system and P system for learning French pronunciation?

Naturally, we can expect incorrectly answered questions from the students. This is an opportunity to help them to learn even more. Our system shows such questions to each student and asks him/her to correct them. The student can consult the system he/she like most. All the questions need to be correctly answered in order to proceed to the next task. This activity gives them the chance to use both systems, which is necessary for answering the subjective questions of Table 4.

4 Evaluation results

Eight post-graduated Japanese students learning French participated of our evaluation. They were separated into two groups, according to the system they have been introduced first. Let us name P group as the students who utilized P system first, and T group the remaining students.

4.1 Objective evaluation

We compare the scores of each group of students in both exams, splitting the results by type of problems.

The first problem has four questions, scoring 0.25 points for each correct answer, or 0, otherwise. Therefore, the sum of scores ranges from 0 to 1. The results are shown in Table 5. The mean for P group is equal to the mean of T group for the pretest. The variance of P group, however, is 5 times larger than T group. Note that the average performance of P group is only 0.44 in the pretest, but after utilizing P system, they doubled their performance (+100%). The relative improvement of T students is +86%, although they start from the same average as their counterpart (0.44). These results are just indicative.

Table 5 Performance of students for problem 1.

Group	Measure	Pretest	Posttest
T	mean	0.44	0.81
	variance	0.02	0.02
P	mean	0.44	0.88
	variance	0.10	0.06

As for the second problem, the scoring criterion is the same as problem 1. Since there is only one possible choice, it is a relatively simple problem. The results are shown in Table 6. T group improved from 0.56 to 0.69, while P group almost doubled the performance again, improving from 0.44 to 0.81. The relative improvement of T group is only +18%, whereas P group is +86%. The performance of students using P system was 4.78 times better than students using T system, although these results are just indicative.

Table 6 Performance of students for problem 2.

Group	Measure	Pretest	Posttest
T	mean	0.56	0.69
	variance	0.06	0.02
P	mean	0.44	0.81
	variance	0.10	0.02

The third problem has 5 sentences, with multiple correct answers. This makes this problem more difficult than the previous one. The scores are different here. If a student's answer is “I am not sure”, 0.0 point is scored for the whole problem. Otherwise, the score is 0.2 points for every match and 0.0 for every non-match.

Table 7 Performance of students for problem 3.

Group	Measure	Pretest	Posttest
T	mean	0.60	0.90
	variance	0.16	0.01
P	mean	0.35	0.90
	variance	0.17	0.01

In average, the performance of P group was worse for this problem during the pretest, as shown in Table 7. Their average score was only 0.35. This seems to confirm the degree of difficulty for this problem. However, T group surprisingly had better performance here. They obtained 0.6 points, although the variance was relatively high and approximately the same as P group. After the learning process, however, both groups coincidentally finalized the posttest with the same mean score and variance, 0.90 and 0.01 respectively. In the case of T group, the relative improvement was only +50%, whereas P group had as much as +157%. These results indicate that the use of P system helped the students to learn 3 times more than those using T system. Again, these results are just indicative.

4.2 Subjective evaluation

Subjective results are obtained from our subjective questionnaire (cf. Table 4). Fig. 2 summarizes the results obtained from the comparison between P and T systems. The horizontal axis indicates the question number in Table 4, while the vertical axis indicates the average score for each group of students. The minimum score is 1 and the maximum is 5 points.

Except for intuitive navigation, the students found P system better or equal to T system in all other evaluated aspects, as can be seen in Fig. 2. The only equally scored aspect of both systems was the amount of learning. Subjectively, they felt both systems had the same learning contribution. Objectively, however, these results are not the same. As we have shown in Section 4.1, the students using P system had better learning performance than those using T system for every problem. It is difficult to subjectively evaluate which system contributed more. Specifically for this measure, we rely on the objective results instead, given that we have data to show where each student has improved.

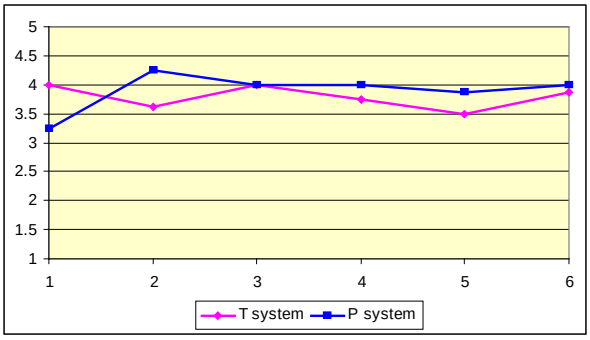

Fig. 2 Comparison between P and T systems in six aspects of our subjective evaluation.

According to Fig. 2, the best score obtained by P system was for the students’ interest. It seems the students appreciated very much the new ideas brought by P system. It seems P system could be more helpful for them to practice and useful for learning French pronunciation than T system. The only drawback pointed by the students was interface/navigation of P system, which needs some improvement in order to become more intuitive.

5 Conclusions and future work

We have implemented a CALL system for teaching liaison in French. We have assessed student’s learning performance after using our system and a traditional method. Our method has been found to be more efficient than the previous one. However, we are not sure whether this tendency is conclusive because we have tested using just a few subjects. We need to test using more students in order to obtain statistically significant results.

6 References

[1] Fouché, P., *Traité de prononciation française*, Paris, Klincksieck, 1959.

[2] Delattre, P., *Principes de phonétique française à l'usage des étudiants anglo-américains*, Middlebury College, 1951.

[3] Pierret, J-M., Phonétique historique du français et notions de phonétique générale, Peeters Louvain-La-Neuve, pp. 98-101, 1994.

[4] Pontes, J. J. A., Furui, S. *Predicting the phonetic realizations of word-final consonants in context – A challenge for French grapheme-to-phoneme converters.* Speech Communication 52, pp 847–862, 2010.

[5] http://www.loquendo.com

[6] http://www.coelang.tufs.ac.jp/modules/fr/pmod2/5-6/1.html

[7] http://cblle.tufs.ac.jp/index.php?id=124