



**HAL**  
open science

**Remédiation phonétique et phonologique en FLE par  
une approche multimodale chez les apprenants  
japonophones**  
Sylvain Coulange

► **To cite this version:**

Sylvain Coulange. Remédiation phonétique et phonologique en FLE par une approche multimodale chez les apprenants japonophones . Sciences de l'Homme et Société. 2016. dumas-01387387

**HAL Id: dumas-01387387**

**<https://dumas.ccsd.cnrs.fr/dumas-01387387v1>**

Submitted on 23 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# **Remédiation phonétique et phonologique en FLE par une approche multimodale chez les apprenants japonophones**

**COULANGE  
Sylvain**

Sous la direction de

Monsieur Jean-Marc COLLETTA & Madame Solange ROSSATO

Laboratoire : LIDILEM

UFR de Sciences du Langage  
Département de Français Langue Étrangère

---

Mémoire de master 2 recherche - 30 crédits

Spécialité français langue étrangère

Année universitaire 2015-2016





# **Remédiation phonétique et phonologique en FLE par une approche multimodale chez les apprenants japonophones**

**COULANGE  
Sylvain**

Sous la direction de

Monsieur Jean-Marc COLLETTA & Madame Solange ROSSATO

Laboratoire : LIDILEM

UFR de Sciences du Langage  
Département de Français Langue Étrangère

---

Mémoire de master 2 recherche - 30 crédits

Spécialité français langue étrangère

Année universitaire 2015-2016



## *Remerciements*

*Merci à mon directeur d'avoir bien voulu assurer les quatre séances de formation intensive, et aux membres de mon jury, pour leur aide & leurs conseils. Et bien-sûr, merci à tous les participants à l'expérience : Yuri, Miyū, Sachi, Shibo, Hono, Yūki, Yoshie, Yasuhiro, Sae et Yuki.*

### DÉCLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : ..... COULANGE .....

PRENOM : ..... SYLVAIN .....

DATE : ..... 6 / 07 / 2016 .....

SIGNATURE :



# Sommaire

<b>Introduction</b>	<b>I</b>
<b>Phonologies &amp; apprentissages</b>	<b>5</b>
<b>1 Acquisition de la phonologie en L.E.</b>	<b>7</b>
1.1 Qu'est-ce que « bien parler une langue étrangère »? . . . . .	7
1.2 Mécanismes d'acquisition de la phonologie . . . . .	8
1.3 Mesurer & évaluer la prononciation . . . . .	19
<b>2 Une approche multimodale de la phonologie</b>	<b>27</b>
2.1 La pédagogie du Silent Way . . . . .	28
2.2 Silent Way & phonologie . . . . .	33
2.3 Multimodalité & Silent Way . . . . .	35
<b>3 Le japonais &amp; sa phonologie</b>	<b>39</b>
3.1 Stratification lexicale . . . . .	40
3.2 Le système d'écriture . . . . .	42
3.3 Les voyelles . . . . .	49
3.4 Les consonnes . . . . .	51
3.5 Rythme & prosodie . . . . .	58
3.6 Problématique . . . . .	59
<b>4 Étude de cas</b>	<b>61</b>
4.1 Profil du locuteur . . . . .	62
4.2 Transcriptions des enregistrements . . . . .	62
4.3 Difficultés au niveau segmental . . . . .	64
4.4 Difficultés au niveau syllabique . . . . .	67
4.5 Synthèse de l'analyse . . . . .	68
4.6 Réflexions générales . . . . .	69



<b>Proposition d'une formation Silent Way &amp; observation de l'évolution de la prononciation</b>	<b>71</b>
<b>5 Méthodologie</b>	<b>73</b>
5.1 Les participants . . . . .	74
5.2 Le protocole . . . . .	76
5.3 L'analyse des données . . . . .	81
<b>6 Déroulement de la formation &amp; observations</b>	<b>87</b>
6.1 Découverte de l'outil Kinéphones . . . . .	87
6.2 Pointage des premières difficultés . . . . .	88
6.3 Suite de la présentation du tableau . . . . .	89
6.4 Utilisation d'autres supports en parallèle du tableau . . . . .	91
6.5 Conclusion . . . . .	93
<b>7 Analyse des données</b>	<b>95</b>
7.1 Alignement automatique avec EasyAlign . . . . .	95
7.2 Analyse des productions à T <sub>1</sub> . . . . .	103
7.3 Comparaison des productions à T <sub>1</sub> & T <sub>2</sub> . . . . .	110
7.4 Discussion . . . . .	114
<b>8 Entretiens avec les participants</b>	<b>117</b>
8.1 Apprentissage de la phonologie au Japon . . . . .	118
8.2 Difficultés perçues dans l'apprentissage . . . . .	119
8.3 Retours sur les premiers ressentis devant le tableau . . . . .	120
8.4 Retours sur l'utilisation du tableau . . . . .	121
8.5 Retours sur la formation et l'approche pédagogique . . . . .	122
8.6 Retours sur le groupe . . . . .	123
8.7 Ancrage du tableau dans la mémoire à moyen terme . . . . .	124
8.8 Remarques concernant les pointages de mots . . . . .	128
8.9 Discussion . . . . .	129
<b>Conclusion</b>	<b>131</b>
<b>Bibliographie</b>	<b>133</b>
<b>Table des figures</b>	<b>136</b>
<b>Liste des tableaux</b>	<b>138</b>
<b>Annexes</b>	<b>141</b>

# Introduction

Dans la présentation d'un volume de la *Revue française de linguistique appliquée*, FLAMENT-BOISTRANCOURT (2006) écrit avoir rencontré nombre de jeunes étudiants chinois partis faire leurs études en Europe, titulaires d'un diplôme de niveau avancé de la langue du pays d'accueil et pourtant tout juste compréhensibles à l'oral. Cette situation est loin d'être rare, et elle est même très commune dans beaucoup de pays d'Asie. Si l'on peut recevoir un diplôme en langue tout en gardant de grosses difficultés à l'oral, c'est certainement parce que l'écrit est prédominant dans l'enseignement des langues étrangères. Pourquoi ? Parce qu'il est bien plus simple d'évaluer objectivement un apprenant sur ce qu'il écrit, que sur ce qu'il dit : la parole étant sujette à de grandes variations interlocuteur, et son existence éphémère et volatile n'en facilitant ni le traitement ni le jugement. L'évaluation à la fin de laquelle l'apprenant se voit décerner un diplôme étant en grande partie basée sur l'écrit, il est donc naturel que l'apprentissage lui-même se concentre sur celui-ci ; dans le but d'obtenir ce diplôme, qui permettra à l'apprenant d'aller de l'avant. Car rappelons que les cours de langue en contexte scolaire ou universitaire, ont pour objectif premier non pas d'apprendre une langue étrangère aux élèves, mais de leur faire obtenir les crédits nécessaires à l'obtention du diplôme visé. Dans les écoles de langues ou les cours particuliers, la situation est cependant bien différente.

Face à ce déséquilibre entre écrit et oral en classe, l'enseignement des langues étrangères tente depuis quelques décennies de donner plus d'importance à l'oral. Le développement technologique permet aujourd'hui de donner un nouveau statut à celui-ci, notamment lorsque les apprenants n'ont pas de contact avec un enseignant natif de la langue étudiée. Alors certes il y a un mouvement récent vers l'enseignement-apprentissage de l'oral, mais les progrès sont toutefois modestes et encore bien des choses restent à faire (HIRSCHFELD 2011b ; HIRSCHFELD 2011a ; MO-

RAZ 2011). On assiste d'ailleurs à un paradoxe : plus la didactique des langues se rapproche de l'oral, moins elle s'intéresse à la phonétique et à la phonologie. PISTORIUS et WETZEL-KRANZ (2011, p. 6) qualifient même ces dernières de « "Cendrillon" de l'enseignement et de l'apprentissage des langues » et déplorent le manque de formation des enseignants dans cette discipline.

Pourquoi l'oral est-il si important ? Maîtriser une langue ne se résume évidemment pas au fait de savoir la parler. Mais il est certain que dans le monde dans lequel nous vivons aujourd'hui, une élocution de qualité, voire fluente en langue étrangère est un enjeu essentiel ; que ce soit lors d'un entretien d'embauche où le jugement du recruteur doit se faire rapidement sur ce qu'il voit et entend, ou dans une relation professionnelle quelconque avec une entreprise étrangère, où la connaissance de la langue du partenaire est un atout, facteur de confiance et de réussite.

Se pose alors la question suivante : qu'est-ce que parler couramment une langue ? Sur le plan phonétique, sur quels critères peut-on qualifier une élocution de « bonne » ou « naturelle » alors qu'il existe tant de variation entre les locuteurs d'une même langue ? On pense souvent à l'expression d'un français classique, accompagné d'une grammaire impeccable, sans élision ni oubli de liaison, mais un tel langage est-il naturel ? Ne doit-on pas plutôt considérer comme « naturel » la variété parlée dans notre environnement, que celui-ci soit l'Acadie, le Sénégal ou l'Ardèche ?

MATTER (2006) soutient qu'en langue étrangère, la prononciation authentique (ou comparable à celle d'un natif de la langue concernée) est pratiquement impossible à acquérir pour un adulte, et que ceux qui l'acquièrent ne sont que de rares exceptions. Que la maîtrise de la langue soit des plus avancées, lorsque le non-natif ouvre la bouche, il se trahit, disait GUIORA (1991, cité par MATTER (2006)) en appelant ce phénomène le « phénomène Conrad », du nom d'un écrivain ukrainien qui écrivait un anglais parfait mais qui n'a pu se débarrasser de son accent slave.

Il existe effectivement des contraintes biologiques et psychologiques qui apparaissent avec le temps et qui peuvent ralentir le processus d'apprentissage, mais si l'on est capable d'apprendre à conduire une voiture, d'éduquer ses enfants ou de gérer les tensions avec les individus qui nous entourent, pourquoi ne serions-nous pas capables d'apprendre un code de communication différent et nous habituer à une prononciation différant de notre langue maternelle ? Tout le monde est capable de travailler sa perception des sons, son articulation ou son intonation. Mais il s'agit de prendre le

temps de le faire, et de trouver une méthode efficace et adaptée à son profil. Il n'y a pas, comme on pourrait le croire, de prédispositions héréditaires à pouvoir prononcer une langue sans difficulté ; mais c'est un grand nombre de facteurs qui font de l'apprentissage un processus plus ou moins difficile. L'environnement dans lequel nous vivons et apprenons joue un rôle particulièrement important, tout comme notre comportement vis-à-vis de la langue cible ou l'âge auquel on rentre en contact avec celle-ci. De plus, la ou les langue(s) maternelle(s) et son ou leur système phonétique et phonologique, selon l'éloignement avec la langue cible, va tantôt faciliter l'acquisition ou la rendre plus laborieuse.

La présente étude se donne pour objectif de proposer un enseignement efficace de la phonologie du français langue étrangère, grâce à une approche multimodale de la langue évitant l'influence de l'écriture et la répétition par mimétisme de l'enseignant. Cet enseignement sera proposé à travers une formation intensive, destinée à un groupe d'apprenants japonophones multi-niveaux, et dont la langue maternelle est à l'origine de grandes difficultés phonétiques et phonologiques en français. Un protocole sera mis en place pour permettre une évaluation automatique de la qualité de prononciation des apprenants. Cette évaluation aura lieu avant et après la formation intensive, dans l'objectif de permettre l'observation en détails de l'évolution de la production des participants.

Le premier chapitre présentera les grandes théories de l'acquisition de la phonologie en langue étrangère, l'influence de l'écriture dans l'apprentissage, et un bref état de l'art sur les pratiques d'évaluation de la prononciation en langue étrangère. L'approche choisie pour cette étude sera présentée dans le deuxième chapitre. Après une présentation du système phonologique du japonais en chapitre trois, une étude de cas de la production d'un apprenant japonophone sera exposée dans le quatrième chapitre. Le chapitre cinq sera réservé à la description du protocole mis en place, le chapitre six décrira en détails le déroulement de la formation et la réaction des participants, et le chapitre sept présentera l'analyse des données. Enfin, un dernier chapitre se consacrera à résumer les retours qualitatifs obtenus auprès des participants lors d'entretiens semi-directifs individuels post-formation.



## Première partie

# Phonologies & apprentissages



# Chapitre 1

## Acquisition de la phonologie en langue étrangère

### 1.1 Qu'est-ce que « bien parler une langue étrangère » ?

Les premières études en acquisition d'une langue seconde ont été longtemps centrées sur le rôle que jouait la langue maternelle dans l'apprentissage (LEATHER 1999, MAJOR 2001, cités par TRAN 2011). L'influence phonétique et phonologique de la langue maternelle sur la langue apprise est indiscutable, mais il a fallu du temps pour comprendre qu'il ne s'agit pas d'un simple transfert direct de la phonologie de la langue maternelle vers la langue cible. Cette première est loin d'être la seule source de difficultés. Selon le modèle de l'ontogenèse de MAJOR (1987, cité par TRAN 2011), les difficultés dues à la langue première sont surtout présentes dans les premiers stades de l'apprentissage ; par la suite, au fur et à mesure que l'apprenant progresse, il sera confronté à d'autres facteurs plus importants comme les préférences typologiques universelles, le marquage des phonèmes <sup>1</sup> ou des facteurs sociolinguistiques et psychologiques. Cependant, le fait qu'il apparaît des obstacles dans l'apprentissage ne rend pas pour autant ce dernier impossible.

Mais est-il nécessaire de parler une langue avec la prononciation d'un natif ? Non,

---

<sup>1</sup>Selon la théorie de la marque de Trubetzkoy, dans les années 30, plus un phonème est marqué, plus il est rare et difficile à réaliser ; une marque peut être par exemple un trait de voisement ou de nasalité. L'ensemble de toutes les marques de chaque phonème constitue la carte d'identité de ce son en psycho-acoustique.



bien sûr. Si le bilinguisme a longtemps été considéré comme un double monolinguisme, dont il est impossible de deviner quelle est la langue première et quelle est la langue seconde<sup>2</sup>, on reconnaît aujourd'hui la compétence partielle, et il est nullement nécessaire de bien prononcer une langue pour pouvoir l'utiliser, tant que l'on reste intelligible. Puisque finalement c'est cela qui importe : *la communication*. L'accent est une partie de l'identité et n'a pas à être rejeté. Dans un monde où les cultures se mélangent, ce n'est pas seulement aux locuteurs de faire des efforts pour être compris, mais aussi aux auditeurs pour accepter la différence dans la langue, et faire l'effort de comprendre et de laisser parler. Cependant, le phénomène de *glottophobie* (discrimination par l'accent, BLANCHET 2016) existe bel et bien, et un accent différent de celui de l'interlocuteur est souvent à l'origine d'une attribution causale, d'une catégorisation à un certain groupe accompagné de tous les préjugés que l'on peut lui attribuer. Il est aussi souvent la cause d'un jugement négatif sur les compétences linguistiques globales du locuteur, voire même de ses compétences intellectuelles générales.

En tant qu'enseignant de langue, il est donc nécessaire d'être en mesure de proposer une formation adéquate permettant un travail sur la qualité de la prononciation, sans pour autant l'imposer aux apprenants.

## 1.2 Mécanismes d'acquisition de la phonologie

Pour apprendre à prononcer, il faut des années et des années. Grâce à la science, nous pouvons y parvenir en quelques minutes.

---

La leçon, Eugène Ionesco

---

<sup>2</sup>Prenons comme référence la définition de BLOOMFIELD (1935) pour qui le bilinguisme est un cas extrême où les deux langues sont maîtrisées parfaitement, à tel point qu'il n'est plus possible de distinguer le locuteur s'exprimant en langue étrangère des natifs de cette langue. Remarquons que cette définition est encore bien ancrée aujourd'hui ; l'édition 2002 du Petit Robert définit toujours le terme "bilingue" comme « Qui parle, possède parfaitement deux langues ».

### 1.2.1 La notion de crible phonologique

Au début des années 70, les travaux d'Eimas, Siqueland, Jusczyk et Vigorito sont les premiers à proposer un protocole rigoureux pour observer les compétences linguistiques du nourrisson. EIMAS et al. (1971) prouvent que l'enfant âgé de 1 à 3 mois est déjà capable de discriminer deux syllabes proches du type /ta/-/pa/, et de nombreuses répétitions de l'expérience confirmeront par la suite que l'enfant ne naît pas « *tabula rasa* » attendant l'empreinte d'un environnement qui le doterait progressivement de capacités de plus en plus complexes » mais que des prémices du langage ou du calcul sont déjà présents bien plus tôt que ce à quoi on s'attendait jadis (DEHAENE-LAMBERTZ et VAN OOIJEN 2002, p. 393). Il sera démontré que l'apprentissage de la langue débute bien avant la production des premiers mots, et que l'enfant s'habitue à la prosodie de la langue parlée par sa mère, *in utero*, dès le dernier trimestre de la grossesse, lui permettant dès la naissance de discriminer des phrases anglaises et japonaises (NAZZI, BERTONCINI et MEHLER 1998). Il ne s'agit là que d'une distinction imprécise, et d'autres paramètres interviendront par la suite, permettant de distinguer des langues plus proches. Dès les premiers mois, le nourrisson distingue quasiment tous les contrastes consonantiques des langues humaines, puis au contact de sa langue maternelle, son système auditif va subir peu à peu une réorganisation psychoacoustique de l'espace phonétique, réduisant sa capacité à discriminer certains contrastes étrangers inutiles dans sa langue maternelle (KUHL, WILLIAMS, LACERDA, STEVEN et LINDBLOM 1992, cités par (DEHAENE-LAMBERTZ et VAN OOIJEN 2002)). À partir d'un an environ, l'enfant japonais ne fait par exemple plus la distinction entre /ra/ et /la/ (inutile en japonais), alors qu'il la faisait encore quelques mois avant. On constate donc que le corps, et plus précisément le système auditif du jeune enfant, va rapidement s'habituer aux sons de la langue maternelle pour créer un schéma type permettant de distinguer plus facilement les phonèmes discriminants dans la langue de l'environnement, et ignorant peu à peu les contrastes non-pertinents dans cette même langue.

Dans la figure 1.1, KUHL (2004) présente un grand nombre de réalisations de voyelles de l'anglais, variant selon plusieurs paramètres comme le sexe ou l'âge du locuteur. Dans ce nuage de sons tous phonétiquement différents, vont peu à peu se distinguer des catégories pertinentes pour le traitement de la langue anglaise, chaque

catégorie représentant un phonème vocalique de l'anglais. Et on constate que pour plusieurs centaines de réalisations différentes, on a seulement 10 phonèmes. Cette catégorisation des voyelles spécifique à la langue maternelle constitue un véritable filtre nous permettant de traiter l'information orale en allégeant la charge cognitive nécessaire à la distinction des sons. Cependant par la suite, lorsque l'enfant entendra une langue étrangère, il aura tendance à traiter les voyelles qu'il perçoit comme des variantes appartenant aux catégories qu'il a définies pour sa langue maternelle.

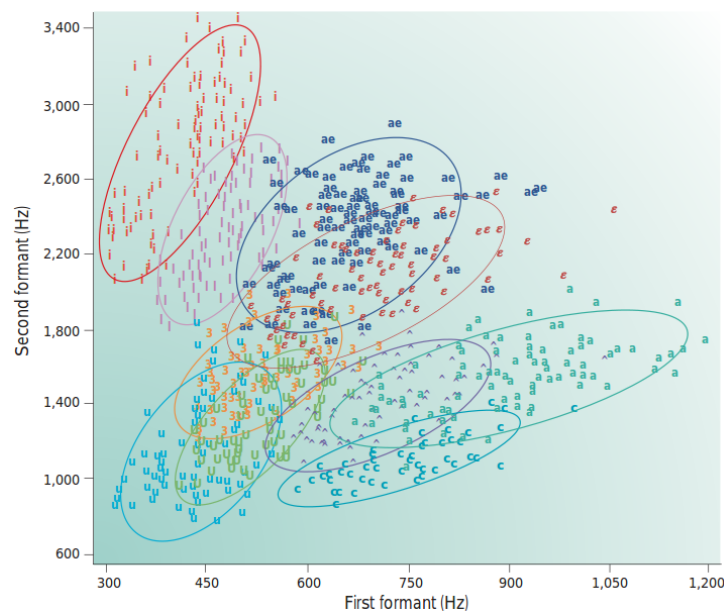


FIG. 1.1 : Catégorisation des voyelles d'après KUHL (2004, p. 834)

Cette surdit  aux phon mes qui ne sont pas pr sents dans le syst me phonologique de la langue maternelle (ou des langues maternelles) est la cause principale de la difficult    percevoir ou prononcer les mots d'une langue  trang re. Le concept de surdit  phonologique appar t d s les ann es 1930, dans les travaux de Polivanov et Trubetzkoy. En 1931, Polivanov la d crit aussi comme la capacit  de repr senter les sons de n'importe quelle langue avec notre propre syst me phon tico-phonologique. Ce qui veut dire que nous percevons les phon mes d'une langue donn e comme une variante des phon mes pr sents dans le syst me phonologique de la ou des langues que nous connaissons et auxquelles nous sommes habitu s ; et les sons qui ne sont pas phon miques dans notre/nos langue(s) ne sont pas, ou tr s difficilement, per ues

(MATTER 2006).

TRUBETZKOY (1939) décrit le système phonologique d'une langue comme une succession de filtres à travers lesquels passent les sons que l'on entend. Il y en aurait trois selon lui : un premier ayant pour rôle de distinguer les phonèmes, et deux autres pour retenir les éléments suprasegmentaux pertinents pour caractériser l'expression du locuteur. Mais chaque langue requiert ses propres filtres, et on a tendance à utiliser inconsciemment les filtres de notre langue maternelle lorsqu'on entend parler une langue étrangère. Il faut donc se construire de nouveaux filtres pour chaque nouvelle langue apprise.

ARCHIBALD (1998, p. 3 cité par TRAN 2011, pp. 8-9), donne un exemple dans lequel le verbe « have » [hæv] en anglais est filtré/interprété différemment par les francophones [æv] et par les germanophones [hæf], chacun entendant le mot selon le répertoire phonémique et les règles phonologiques qu'il possède : absence de la glottale /h/ en français ; assourdissement de la consonne sonore finale en allemand. GERLACH (2004, cité par TRAN 2011, p. 10) donne aussi l'exemple bien connu de « school » /skul/ qui est souvent prononcé [eskul] par les hispanophones, cela étant dû au fait qu'il n'y a pas de combinaisons polyconsonantiques en début de syllabe à part avec une liquide.

Ce filtre n'agit pas seulement sur l'identification du phonème. OTAKE, HATANO, CUTLER et MEHLER (1993, cités par TRAN 2011) démontrent que les auditeurs français, lorsqu'ils sont exposés à la langue japonaise, utilisent une stratégie syllabique pour décomposer le flux de parole qu'ils entendent ; tandis que les auditeurs japonais utilisent les frontières de *mores*<sup>3</sup> comme indice de frontière des mots. La segmentation lexicale est donc étroitement liée à la structure rythmique de la langue source. L'influence de la langue maternelle ne se limite donc pas à la dimension segmentale de la langue. Les Français ont des difficultés à percevoir la différence entre les mots « *tópo* » (taupe) et « *topó* » (avoir trouvé) en espagnol. En effet l'accent tonique ne constitue pas un paramètre pertinent dans le système phonologique du français : les mots isolés portent tous l'accent tonique sur la dernière syllabe, l'accent n'est jamais utilisé pour distinguer deux lexèmes comme *tópo* et *topó* en espagnol. Par conséquent, il est particulièrement difficile pour les apprenants francophones de prendre en compte ce nouveau paramètre qu'est l'intonation lorsqu'ils apprennent une langue

---

<sup>3</sup>Unités phonologiques minimales en japonais (*cf.* Chapitre III).

où l'accent est phonologiquement pertinent. Les hispanophones quant à eux y sont habitués, et ont bien moins de difficultés à percevoir et produire correctement les accents toniques en langue étrangère.

Ainsi, tous les éléments semblables entre la ou les langue(s) maternelle(s) (langue(s) source(s)) et la langue étrangère apprise (langue cible) rendent l'apprentissage plus simple, qu'il s'agisse de phonème, de structure syllabique, de paramètres supra-segmentaux etc. On parle alors de transfert positif entre la langue source et la langue cible (ELLIS 1985, cité par TRAN 2011). À l'inverse, tout ce qui diffère d'une langue à l'autre rend l'apprentissage difficile, et c'est une source d'erreur pour l'apprenant ; c'est le transfert négatif ou l'interférence. Mais cette vision contrastive entre langue source et langue cible dans l'apprentissage a ses limites et est assez critiquée depuis les années 70, car elle suppose que toutes les différences sont à l'origine de difficultés de même niveau. Pourtant certaines différences peuvent simplifier l'apprentissage, et surtout toutes les difficultés ne sont pas de même degré. Une étude de BRIÈRE (1966, cité par TRAN 2011) montre que les anglophones ont moins de difficultés à prononcer une fricative uvulaire sourde [χ] du français ou de l'espagnol que son équivalent sonore [x] du français ou que la pharyngale [ħ] de l'arabe par exemple, alors que ces trois réalisations n'existent pas en anglais. On ne peut donc se contenter de dire que les différences interlangues sont à l'origine de difficultés similaires. D'autres paramètres rentrent en compte comme le marquage du phonème, son lieu d'articulation, ou encore des difficultés sociolinguistiques ou psychologiques. Beaucoup d'études des années 80 montrent qu'une large proportion d'erreurs d'apprenants ne peuvent pas être expliquées par l'interférence de la langue source (TRAN 2011).

### 1.2.2 Des processus d'apprentissage complexes

BLANC (2011) propose de diviser les difficultés des apprenants dans l'acquisition de la phonologie d'une langue étrangère en 6 catégories : les difficultés de niveaux 1-perceptif, 2-productif, 3-cognitif, 4-identitaire, 5-psychologique et 6-pragmatique. Je propose de les résumer dans le tableau de la figure 1.1.

L'acquisition d'un nouveau système phonologique est donc une tâche considérable qui implique l'apprenant dans son entier (BLANC 2011). Dans leur dictionnaire de didactique des langues, COSTE et GALISSON (1976) expliquent qu'« apprendre une

PERCEPTION	percevoir la différence entre plusieurs phonèmes qui ne sont pas distingués en langue source ; observer les éléments suprasegmentaux (intonation, rythmique), par exemple il est important de se concentrer sur la finale du mot en français car c'est lui qui porte la plupart des flexions grammaticales : genre, nombre, personne, temps, mode etc. On a aussi tendance à répéter en fin d'énoncé la partie importante pour s'assurer d'une bonne compréhension (BLANC (2011) rappelle le garçon parisien : « deux cafés, deux ! »)
PRODUCTION	nouveaux mouvements musculaires. Nouvelles positions des organes articulatoires. En français on connaît les difficultés à prononcer les voyelles antérieures arrondies [y] [ø] et [œ] ; la difficulté des sinophones à acquérir l'opposition des consonnes sourdes/sonores (d-t, p-b, f-v etc.) ; la chuintante sonore [ʒ] est aussi difficile pour beaucoup d'apprenants ; les différences de tension articulatoire qui posent notamment problème chez les japonophones qui possèdent un système plutôt relâché ; les groupes consonantiques en initiale comme /sp/ prononcé souvent /ʃp/ par les germanophones ou /esp/ par les hispanophones ; les groupes consonantiques en général, qui posent particulièrement problème pour les japonophones et vietnamophones...
COGNITION	l'apprenant doit créer pour chaque son une représentation globale, ses représentations acoustiques selon le ou les phonèmes avec lesquels il est associé, ses représentations articulatoires, combinatoires, fonctionnelles, et graphique(s).
IDENTITÉ	apprendre un nouveau système phonétique peut rentrer en conflit avec la représentation identitaire de l'apprenant. « parler le FLE de manière absolument identique à un francophone signifie perdre son identité d'allophone. » (BLANC 2011, p. 34) ; de plus, on sait que les apprenants ont de nombreuses représentations sur les langues.
PSYCHOLOGIE	l'apprenant doit surmonter sa frustration, son échec ou sa peur du ridicule ; il doit être ouvert et accepter qu'il peut entendre autre chose que ce qu'il entend dans l'immédiat lorsqu'il ne parvient pas à distinguer deux phonèmes.
PRAGMATIQUE	l'apprenant doit associer une suite de sons à un sens et décider si cette association convient à une situation de communication donnée.

TAB. 1.1 : Les 6 niveaux de difficulté dans l'acquisition de la phonologie en langue étrangère, d'après BLANC (2011)

langue étrangère apparaît à plus d'un titre comme une remise en cause et une réorganisation partielle des conditionnements et des habitudes de perception propres à la langue et à la culture d'origine ». Passer d'une langue à l'autre, ce n'est pas seulement changer de code linguistique, c'est aussi passer d'une identité culturelle à une autre. Pour prononcer une langue étrangère comme un natif de cette langue, il faut revêtir une personnalité propre à celle-ci. MORAZ (2011) affirme qu'il vaut mieux travailler l'acteur qui est en nous, en se concentrant sur le corps dans son entier et ses émotions, et que l'immersion ne suffit pas, il faut s'entraîner, être encadré et dirigé. Moraz ajoute que « la voix elle-même démarre aux origines du corps et de l'être qui diffèrent pour chaque idiome. Par conséquent, il semble judicieux de trouver de nouvelles approches qui permettent aux apprenants de rattacher tel ou tel son à une position corporelle, à un geste, ainsi qu'à une attitude affective. » (2011, p. 38).

Quelque soit son degré de motivation ou d'amour pour la langue, un apprenant ne peut pas maîtriser une langue étrangère d'emblée ; le processus d'apprentissage est long et fastidieux. Ses compétences linguistiques dans celles-ci sont donc en perpétuelle évolution. Tout au long de son apprentissage, l'individu va parler une langue qui n'est ni sa langue maternelle, ni tout à fait la langue cible, bien que sa production s'en rapprochera de plus en plus au fur et à mesure de sa formation. On appelle cet ensemble de compétences dynamiques en langue cible *l'interlangue* (SELINKER 1972, entre autres, cité par FLEGE 1980). Cette interlangue est finalement la version personnalisée de la langue cible de l'apprenant, un continuum entre la compétence nulle (si elle existe) et ce qui pourrait être comparé à la compétence d'un adulte natif de cette langue cible. On peut alors se demander si l'acquisition de la phonologie de la langue cible représente aussi un continuum comparable à celui de la compétence linguistique générale. On pense souvent que le début de l'apprentissage d'une langue est consacré principalement à l'apprentissage de sa phonologie, et une fois qu'elle est plus ou moins acquise, c'est au tour des structures de la langue, de sa grammaire etc. Même si on sent toujours un accent dû à la langue source, on ne revient pas nécessairement sur la phonologie dans la suite de l'apprentissage. Or, celle-ci ne s'acquière pas d'emblée non plus. On doit déjà distinguer les compétences de perception et de production, qui ne sont pas simultanées ; celles-ci sont aussi constituées d'un long processus d'acquisition pendant lequel la qualité de perception ou de production évolue, et tend vers celle que pourrait avoir un natif de la langue étudiée. La

prononciation de tel ou tel phonème ou combinaison phonémique ne peut pas être réduite à la dichotomie correcte/incorrecte, elle évolue d'une forme phonétiquement éloignée vers une réalisation courante chez la plupart des locuteurs natifs de la langue en question (FLEGE 1980).

### 1.2.3 L'influence de l'écriture

Qui travaille sur l'écrit ne fait plus  
fonctionner son ouïe.

---

Renard, 1983, p.44, cité par Moraz,  
2011, p.41

Il existe un autre obstacle au processus d'acquisition de la phonologie d'une langue étrangère : le code écrit. Pourquoi l'enseignement-apprentissage de la prononciation s'appuie-t-il si souvent sur l'écrit ? La conséquence est que l'apprenant se fait une représentation mentale graphique pour arriver à prononcer tel ou tel énoncé, et cela peut l'induire en erreur, car la graphie n'est pas forcément logique et les graphies des langues qu'il connaît déjà peuvent entrer en conflit avec la graphie de la langue nouvellement apprise (RACINE, DETEY et KAWAGUCHI 2012). MORAZ (2011) explique que l'écrit est devenu si important dans l'éducation, qu'il est difficile de concevoir un mode d'enseignement oral ; et on peut voir que c'est effectivement le cas par exemple avec les difficultés que rencontre l'enseignement du mandarin par le tout à l'oral. Celui-ci reste inconcevable pour bien des institutions, alors qu'il permettrait à beaucoup d'apprenants de parvenir à un niveau avancé, sans être ralenti voire dégoûté de la langue par l'apprentissage des sinogrammes. En se précipitant sur l'écrit dès le début de l'apprentissage d'une langue étrangère, il y a de forte chance pour que, au bout du temps de formation, ni l'oral ni l'écrit ne soit correctement maîtrisé. Alors quitte à en arriver là, autant mettre l'un de côté pour réussir l'autre. L'apprentissage de l'écrit sera toujours plus efficace avec de solides connaissances à l'oral, et c'est de plus le processus naturel d'acquisition de la langue maternelle.

Toutefois, passer par le code écrit peut nous permettre de mieux mémoriser une information ; et le seul message oral est trop éphémère et fragile pour suffire à l'apprentissage. Lorsqu'un nouveau lexème est abordé en classe, par exemple, les apprenants demandent souvent à l'enseignant de l'écrire au tableau, afin de mieux le



comprendre et le retenir. Comme si voir les caractères transcrivant le dit mot était nécessaire pour se l'approprier. C'est certainement la situation à laquelle nous sommes arrivés en privilégiant à ce point l'écrit par rapport à l'oral dans l'enseignement. Nous ne faisons plus confiance qu'à ce que nous voyons. Il est vrai que la mémoire visuelle est souvent plus effective que la mémoire auditive, mais n'est-ce pas le résultat d'une culture éducative basée sur l'écrit ? Le résultat de cette mémorisation de la langue par l'écrit est que nous ne retenons que la graphie du lexème en question, et que nous reconstituons sa prononciation à partir de ces graphèmes. C'est un peu comme si nous ne pouvions retenir une mélodie qu'en apprenant par cœur la partition qui en a été écrite, et que lorsque vient le moment de la partager, nous devons d'abord nous remémorer la partition pour la lire et enfin la chanter. N'est-ce pas étrange de procéder ainsi ? L'art de la musique n'est-il pas justement dans le fait qu'elle sort du cœur et non de la tête ? Je pense qu'il en est de même pour la langue. Lorsqu'un natif parle dans sa langue, pense-t-il à toutes les notes indépendantes qui la composent ? Doit-il se remémorer le texte écrit de ce qu'il veut dire pour pouvoir parler ? La langue semble plutôt résonner comme une mélodie inconsciente à laquelle il est habitué, et dont il ne se sert que pour véhiculer le message qu'il veut transmettre. Apprendre une langue par sa représentation graphique uniquement ne permet pas, d'une manière générale, de retenir la plupart des informations suprasegmentales comme le rythme, l'intensité, l'intonation ou le ton. Peut-être que nous nous concentrons trop sur l'outil, et pas assez sur ses fonctions, et c'est peut-être cette approche dénaturisée et décontextualisée de la langue qui fait que tant d'apprenants décrochent au cours de leur formation.

Alors si nous sommes tant attachés au code écrit, pourquoi ne pas en utiliser un plus précis pour noter la prononciation ? La transcription phonétique n'est généralement pas d'une grande aide car la plupart des étudiants ne la maîtrisent pas (ni même les enseignants bien souvent), et sa lecture cursive n'est pas évidente. De plus les caractères de l'alphabet phonétique effraient parfois les apprenants par leur forme inhabituelle, et ceux qui sont directement issus de l'alphabet latin peuvent entrer en conflit avec la prononciation des mêmes caractères dans la langue source ou cible, si celle-ci utilise aussi ces caractères.

Pourquoi alors ne pas utiliser un code écrit semblable à l'alphabet phonétique mais aux caractères radicalement différents de ceux utilisés en langues source(s) et cible

pour éviter toutes confusions et interférences ? Cela pourrait être une solution. C'est d'ailleurs le cas du *bopomofo*, un système de transcription phonétique du mandarin qui a finalement été remplacé par le *pinyin* latin actuel (excepté à Taïwan où il est toujours utilisé). Le *bopomofo* est un alphabet avec un caractère unique pour toutes les attaques syllabiques, les médianes et pour les fins (noyau + coda). Les caractères *bopomofo* sont issus de certains sinogrammes, et ressemblent un peu aux caractères katakana utilisés au Japon (cf. Chapitre III). Il présente l'avantage de pouvoir transcrire l'ensemble des mots du mandarin (et donc des milliers de sinogrammes qui constituent son écriture) avec seulement 37 caractères, sans induire en erreur l'apprenant par sa similarité au script latin par exemple, comme c'est le cas pour le *pinyin*. C'est le système utilisé dans la plupart des dictionnaires d'avant-guerre (avant le *pinyin*) et les dictionnaires taïwanais.

<i>sinogramme</i>	彩虹	<i>détail bopomofo-phonétique</i>	
<i>bopomofo</i>	ㄘ ㄞˇ ㄉ ㄨ ㄥ	ㄘ ㄞˇ	ㄉ ㄨ ㄥ
<i>pin'yin</i>	cǎi hóng	ts <sup>h</sup> aj	x ǝŋ ʌ
<i>phonétique</i>	[ts <sup>h</sup> aj ʌ xǝŋ ʌ]		
<i>signification</i>	arc-en-ciel		

FIG. 1.2 : Les transcriptions *bopomofo* et *pinyin*<sup>4</sup>

Toujours est-il que nous n'avons pas, en français langue étrangère, de caractères phonétiques non-latins évitant la confusion avec l'orthographe, et assez simples pour être facilement acquis et utilisés par les apprenants et les enseignants. Cependant bien des enseignants savent que ce n'est pas grâce à l'écrit que l'on pourra efficacement enseigner la phonologie, et beaucoup d'entre eux se sont penchés sur la place du geste et le rôle du corps dans l'apprentissage de la langue. Dans les Cours de Vacances de l'université de Lausanne par exemple, on propose aux apprenants d'impliquer leur corps tout entier dans l'apprentissage du français. D'après MORAZ (2011, p. 41), « sans mise en mouvement, sans une implication corps et être, le locuteur demeure généralement extérieur à la production », c'est pourquoi cette université propose quantité de modèles, de situations types d'utilisation de la langue à voir, mettre en scène et jouer, et permettant aux apprenants de retenir le schéma prosodique de l'énoncé et

<sup>4</sup>D'après [https://en.wikipedia.org/wiki/Standard\\_Chinese\\_phonology](https://en.wikipedia.org/wiki/Standard_Chinese_phonology)

l'énergie expressive qu'il peut requérir. L'auteure précise cependant que ce type de mise en situation n'est possible qu'à partir d'un certain niveau d'apprentissage, et qu'il semble moins évident de faire ce type d'activités avec un public totalement débutant.

De la même manière, la méthode structuro-globale audio-visuelle (SGAV) propose un apprentissage de la prononciation progressif et intégré à des situations de communication, et non de façon intellectuelle (mémorisation de règles et travail hors contexte). Son approche de la correction phonétique – la méthode « verbo-tonale » – donne un statut particulièrement important à la courbe intonative, la coarticulation et la tension mise en jeu pendant l'élocution (RENARD 1971, cité par BLANC 2011). En jouant sur l'intonation, certains sons peuvent être perçus et prononcés plus facilement : ainsi, d'après l'auteure, le /y/ de « il l'a vu ? » est plus facile à produire que celui de « il l'a vu. », car une montée dans les aigus, et donc de l'intonation facilite la prononciation des voyelles aiguës. La méthode SGAV insiste entre autre sur le fait de travailler avec des dialogues et avec peu ou sans mots isolés sortis de tout contexte langagier. Et encore moins avec des sons isolés comme c'est souvent le cas en classe de langue encore aujourd'hui.

Dans les années 70, Caleb Gattegno propose une approche de la langue radicalement nouvelle appelée l'*approche silencieuse* (ou *Silent Way*), dans laquelle les rôles de l'enseignant et des apprenants sont redéfinis. L'approche de la phonétique et de la phonologie qu'elle propose est unique, elle permet de passer par un support visuel sans passer par l'écrit – les phonèmes sont matérialisés par des couleurs – et c'est cette méthode qui va nous intéresser dans ce mémoire. Le chapitre suivant sera consacré à sa présentation.

Les capacités de notre cerveau sont folles, nous créons des filtres phonologiques permettant d'identifier sans effort les phonèmes pertinents dans la langue que nous parlons, et de caractériser le message sonore que nous percevons en analysant les différents paramètres suprasegmentaux de celui-ci. Nous pouvons créer de nouveaux filtres pour chaque langue apprise, en apprenant à percevoir d'autres phonèmes subtilement différents ou très éloignés de ceux que nous avons l'habitude d'utiliser pour communiquer. Nous sommes théoriquement capables de prononcer n'importe quel phonème de l'ensemble des langues humaines, et nous pouvons imiter n'importe quel accent. Ajoutons que l'identification d'un phonème ne se base pas uniquement sur la perception du son produit, mais aussi sur le contexte dans lequel il apparaît (PETERSON

et BARNEY, 1952, cités par BLANC, 2011). En fonction de ce contexte, nous sommes capables de remettre dans l'ordre des phonèmes, comme des lettres à l'écrit (*cf.* l'article de KUFFER (2006) : « Le creaveu hmauin lit dnas le dsérorde »). Les sons d'une langue ne sont pas simplement le résultat d'une articulation donnée (seule utilisation de l'appareil bucco-phonatoire), mais la coordination de l'articulation, de la respiration, et ils sont intimement liés aux caractéristiques physiques et affectives, c'est pourquoi prononcer une langue à la manière d'un natif ne se limite pas à percevoir et articuler un son, mais il s'agit véritablement de se construire une identité adaptée à cette langue.

En général l'apprenant ne sait pas comment fonctionne son propre système phonologique et comment il prononce les phonèmes de sa ou ses langues sources. C'est pourquoi le rôle de l'enseignant est primordial : il doit faire prendre conscience à l'apprenant des mécanismes qu'il met en place pour prononcer les sons de sa langue, pour ensuite lui faire comprendre les caractéristiques des sons de la langue cibles. Cependant on constate que dans la plupart des classes de langue, l'enseignant ne sait pas lui-même comment il prononce les phonèmes de la langue qu'il enseigne, et il se cantonne à faire répéter ses apprenants jusqu'à ce qu'ils prononcent à peu près comme lui. C'est une méthode qui n'est pas forcément mauvaise, mais ce n'est certainement pas la plus efficace à long terme. De plus on constate que beaucoup d'étudiants abandonnent rapidement, disant souvent ne jamais avoir compris et étant persuadés de leur incapacité (physique ou mentale) à distinguer et prononcer certains phonèmes.

Que pourrions-nous observer dans une classe d'apprenants confrontés à une approche de la phonologie radicalement différente de celle qu'ils ont pu avoir jusqu'à maintenant ? L'accepteraient-ils ? Pourrions-nous observer des évolutions significatives dans leur production ?

### **1.3 Comment mesurer et évaluer la prononciation d'un locuteur ?**

Il y a deux grandes problématiques qui se présentent lorsque l'on souhaite évaluer la prononciation d'un apprenant. Tout d'abord quelle production évaluer ? Le protocole peut grandement influencer la prononciation du locuteur, selon la situation dans

laquelle il se trouve, s'il doit lire, répéter quelque chose ou parler spontanément, s'il connaît ou non le contenu de ce qu'il prononce etc. Et puis se pose la question de quoi évaluer dans sa production. En fonction des paramètres pris en compte dans l'évaluation, les résultats peuvent varier d'une expérience à l'autre. Tout le monde n'évalue pas la prononciation d'un individu de la même manière, car tout le monde n'a pas la même définition de ce qu'est véritablement « une bonne prononciation ». Je propose d'observer les protocoles mis en place dans quelques études plus ou moins récente, dans l'objectif d'avoir une vue d'ensemble de ce qui s'est fait jusqu'à maintenant.

### 1.3.1 Provoquer la parole

La lecture de mots ou de phrases Il semblerait que ce soit la méthode la plus couramment utilisées. Elle permet en effet de s'assurer que tous les locuteurs produisent le même contenu, et d'imposer des formes que l'expérimentateur souhaite observer. On observe souvent l'utilisation de phrases porteuses comme chez TRAN (2011), où chaque mot cible est intégré à une phrase type<sup>5</sup>. Cela permet d'éviter la prononciation isolée du mot, et de prédéterminer l'environnement phonétique direct du mot cible afin de ne pas observer de variantes de réalisation en fonction des mots précédant et suivant. TRAN (2011) propose trois fois le même mot cible dans la même phrase porteuse, à trois temps différents, pour observer si la réalisation varie d'une occurrence à l'autre, ou s'assurer de la stabilité de la prononciation. Cependant, cette méthode présente le biais suivant : le locuteur oralise un texte écrit, et nous savons que la graphie peut influencer la prononciation des mots. Certains chercheurs utilisent notamment la lecture en parallèle d'un autre stimulus, comme la répétition d'un mot prononcé par un natif, pour pouvoir comparer les résultats.

La répétition C'est le cas notamment de RACINE, DETEY et KAWAGUCHI (2012), qui observaient la réalisation des voyelles /y/ et /u/ en français, par des groupes d'apprenants hispanophones et japonophones. Les mots cibles étaient prononcés en répétition puis en lecture, et étaient ensuite présentés à des francophones natifs chargés d'identifier la voyelle prononcée. Les chercheurs constatent alors que les énoncés prononcés en lecture sont globalement moins bien identifiés. FLEGE, MUNRO et MACKAY (1995) ou encore FLEGE, BIRDSONG et al. (2006) mettent en place quant à eux un

---

<sup>5</sup>TRAN (2011) utilise la phrase suivante : « Tu prononces <mot cible> trois fois ».

schémas de type : *Question, Réponse (énoncé cible), Question, Silence (pendant lequel le locuteur doit répéter la réponse précédente)*. Les auteurs disent « espérer que l'insertion d'une question entre la réponse cible et sa répétition empêcherait l'imitation directe par la mémoire sensorielle »<sup>6</sup>. Cette méthode est aussi appelée « *delayed repetition technique* » (*ibid.*).

L'image Il est aussi possible d'utiliser l'image comme support à la production. THOMAS (2015) propose à ses locuteurs de décrire une image de rue bondée avec des anomalies, qui incitent les participants à se focaliser sur des détails communs. Dans cette étude, l'auteur se concentre sur la réalisation des liaisons en français par des étudiants anglophones. AOYAMA et al. (2003) quant à eux présentent à leurs locuteurs japonais vivant aux États-Unis, des images représentant chaque mot cible, avec une traduction du mot en japonais pour s'assurer que le locuteur comprend le mot demandé, ainsi qu'un enregistrement audio de celui-ci dans la langue cible. Chaque image apparaît trois fois de manière aléatoire, mais l'audio n'est diffusé qu'à la première apparition (les auteurs précisent que si les participants ne connaissent pas le mot cible, il peuvent écouter à nouveau le mot cible). À la troisième apparition, la prononciation du locuteur est considérée comme « la plus proche de celle du discours spontané »<sup>7</sup>.

Afin de rassembler plus de productions, THOMAS (2015) demande aussi à chaque locuteur de faire une description favorable de son université à des fins publicitaire en milieu francophone, ainsi qu'une autobiographie de lui-même exprimée au passé.

On peut imaginer un genre de continuum se dessinant entre des méthodes plus contraignantes pour le locuteur, l'obligeant à prononcer certains mots ou certaines phrases dans un cadre bien défini, mais permettant une comparaison des données simplifiées car toutes les productions sont calibrées de la même manière ; et des méthodes guidant beaucoup moins le locuteur, lui permettant de s'exprimer spontanément et librement, mais pour lesquelles le traitement des données est bien plus difficile car les productions sont très différentes d'un locuteur à l'autre. Plus la production du locuteur est libre, plus l'analyse des données semble qualitative.

<sup>6</sup>Traduction personnelle depuis « *It was hoped that interpolating an auditory presented 'question' between the 'answer' (target sentence) and its repetition would prevent direct imitation from sensory memory* » (FLEGE, BIRDSONG et al. 2006, p. 163).

<sup>7</sup>Traduction personnelle depuis : « *The third token of each word was examined because these tokens were considered to be closest to spontaneous speech* » (AOYAMA et al. 2003, p. 3202).

### 1.3.2 Évaluer la prononciation

Une fois les productions enregistrées, il faut déterminer quels paramètres seront pris en compte dans l'évaluation de la qualité de prononciation. Qu'est-ce qui définit « l'accent étranger » dans la production d'un locuteur ? S'agit-il de l'intonation, de la durée de certaines voyelles, de la réalisation de certaines consonnes, du rythme, de l'intelligibilité ? Passé l'évaluation traditionnelle de l'enseignant qui donne un feedback direct sur la production de l'apprenant, deux grandes méthodes d'évaluation se distinguent. Soit il s'agit d'une évaluation basée sur le calcul statistique automatique d'un certain nombre de paramètres prédéterminés (la machine est juge) ; soit les productions sont présentées à un groupe de locuteurs natifs de la langue cible, chargé d'évaluer subjectivement les réalisations. Les deux méthodes se valent : l'une se base sur l'objectivité d'une mesure de la réalité physique du signal sonore, et sur un nombre potentiellement important de réalisations ; l'autre se fonde sur la perception et le ressenti que peuvent avoir des interlocuteurs natifs exposés à ces réalisations.

L'évaluation perceptive peut être réalisée de différente manière. Les natifs, à qui sont soumis les enregistrements, peuvent avoir à noter sur une échelle le degré « d'accent étranger » perçu pour chaque énoncé entendu. AOYAMA et al. (2003) proposent à 16 juges natifs monolingues d'évaluer 384 séries de 4 ou 5 mots cibles sur une échelle de 9 allant de « l'accent étranger le plus fort » à « l'accent étranger le moins fort »<sup>8</sup>. Ils étaient exposés aux enregistrements de 13 mots par 4 groupes de 16 personnes (japonophones natifs et anglophones natifs, adultes et enfants) sur deux périodes d'enregistrements d'une année de décalage. L'objectif était d'évaluer la progression de la qualité de prononciation en anglais, de Japonais résidents aux États-Unis, sur une année et en séparant les enfants (moyenne d'âge=10ans) des adultes (moyenne d'âge=40ans). Dans une autre étude, RACINE, DETEY et KAWAGUCHI (2012) demandent à 28 étudiants francophones natifs d'évaluer 120 mots cibles produits par des japonophones, en commençant par choisir la voyelle qu'ils entendent (/y/ ou /u/) puis en évaluant la qualité de sa réalisation sur un échelle de 5 allant de « très bon représentant de la voyelle » à « autre voyelle » (p. 387).

Les recherches en évaluation automatique de la prononciation se sont développées

---

<sup>8</sup>Traduction personnelle depuis : « strongest foreign accent » et « least foreign accent » (AOYAMA et al. 2003, p. 3202).

dans les années 90, avec l'essor des technologies de reconnaissance vocale. D'après DIGALAKIS et MOUSTROUFAS (2007), les meilleurs algorithmes pour l'évaluation de la prononciation sont basés sur ces technologies. Les logiciels développés sont pour la plupart constitués d'une base de données constituées de textes enregistrés par des locuteurs non-natifs, de différents niveaux, segmentés et annotés par des correcteurs humains. L'apprenant  $\lambda$  susceptible d'utiliser le programme par la suite devra alors lire un des textes de la base de données en s'enregistrant ; le signal alors produit est ensuite comparé aux signaux annotés de la base de données. Les paramètres de durées des phonèmes et des syllabes, la forme du spectrogramme ou la prosodie sont au centre des analyses. La production de l'apprenant est donc évaluée par le biais des dizaines d'autres productions du même texte, annotées en amont par des humains. Le point fort de ce procédé est qu'il peut aisément apporter des informations de remédiations, ou des conseils pour mieux prononcer tel ou tel segment, car tout a été pré-analysé. L'inconvénient majeur cependant, est que l'apprenant ne peut lire qu'un texte contenu dans la base de données, il est donc limité au niveau de sa production. DIGALAKIS et MOUSTROUFAS (2007) tentent d'évaluer la prononciation sur un texte inconnu de la machine. Pour cela, ils décident d'intégrer à la base de données un corpus en langue source (le grec), en plus du corpus natif (anglophones natifs parlant anglais) et non natifs (anglophones non natifs parlant anglais) enregistrés sur les mêmes textes pour calibrer le système de reconnaissance vocale. Le corpus non natif est constitué de la production de 20 locuteurs, et il est évalué par 3 correcteurs (humains) ainsi que par le logiciel. Les paramètres pris en comptes dans l'évaluation sont l'intonation, la prononciation des phonèmes et l'intelligibilité de la phrase, et chaque énoncé se voit attribué une note sur 5 d'après ces paramètres. Comme il est impossible de se baser sur les durées de syllabes ou de phonèmes, puisque les textes utilisés par la suite seront inconnus, les auteurs tentent d'utiliser la reconnaissance vocale pour simuler l'évaluation d'un correcteur humain. En rentrant dans la base de données un corpus dans la langue source, le programme peut dégager des caractéristiques propres aux deux langues (source et cible), et se baser sur celles-ci pour évaluer la production d'un locuteur  $\lambda$ , comme le ferait selon eux le correcteur humain.

KAWAI et HIROSE (1997) ont pour leur part créé un système permettant aux apprenants de japonais langue étrangère, de travailler la production des voyelles longues. Une base de données est d'abord constituée avec des enregistrements non natifs déjà



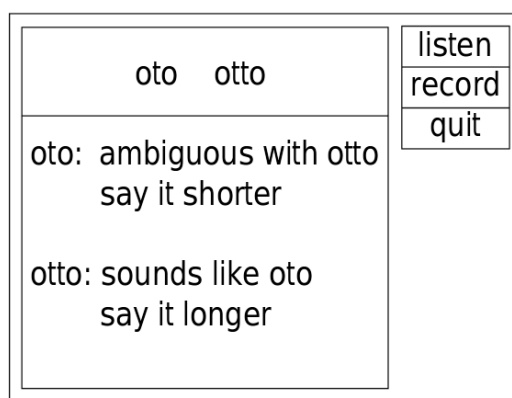


FIG. 1.3 : Interface de l'utilisateur traduite du programme de Kawai et Hirose

alignés et annotés ; puis les textes de la base de données sont présentés à l'apprenant, après une lecture enregistrée, la transcription est alignée automatiquement sur le signal, la durée des phonèmes est calculée puis comparée à celle des alignements de la base de données. En fonction de la différence de durée, un feedback approprié est envoyé à l'apprenant (*cf.* figure 1.3).

### 1.3.3 Discussion

La prononciation peut être évaluée d'un grand nombre de manières différentes, mais chacune d'entre elles possède ses contraintes et ses avantages. On constate qu'une grande partie des études observent la réalisation de mots isolés, ou en contexte fixe et prédéterminé. Parfois les mots produits sont inconnus du locuteurs (*cf.* TRAN (2011)), parfois ils sont traduits dans la langue source. Les stimuli sont majoritairement lus par l'apprenant, ou répétés ; beaucoup d'auteurs sont conscients de l'influence que peut avoir la graphie des mots sur la prononciation, où le mimétisme possible lors de la répétition, mais peu d'études semblent proposer une évaluation de la prononciation d'un énoncé réellement spontané. De nombreux logiciels d'évaluation et/ou correction de prononciation ont vu le jour, mais la majorité d'entre eux imposent un texte prédéterminé à l'apprenant, qui doit alors en lire le contenu pour être correctement évalué. La comparaison de la production d'un apprenant avec les signaux pré-analysés de la base de données, contraint l'utilisateur à ne produire que ce qui a déjà été entré dans la base pour être correctement évalué.

Étant donné l'influence que peut avoir l'écriture sur la prononciation et son apprentissage, il semble capital de pouvoir enseigner et évaluer sans y faire appel. L'approche pédagogique Silent Way offre une opportunité de travailler la phonologie d'une langue sans être influencé par l'écriture, et permet à l'apprenant d'aborder l'objet d'étude dans une attitude rendant l'apprentissage particulièrement efficace, même dans le cas d'une langue phonologiquement distante de sa langue maternelle.



## Chapitre 2

# Une approche multimodale de la phonologie

J'enseigne aux personnes et elles apprennent la langue.

---

Caleb Gattegno

Comment remédier à ces difficultés d'acquisition de la phonologie française par les apprenants de langue maternelle japonaise ? Il faudrait une approche pédagogique qui donne plus d'importance à l'oral et à l'apprentissage de la phonologie. Une approche simple, proposant des outils peu complexes à prendre en main, éviter le métalangage ou le jargon anatomique qui constitue lui-même un apprentissage préliminaire nécessaire et peut freiner l'apprentissage des phonèmes. Il faut une approche qui se concentre sur la substance orale et non sur sa transcription graphique, du moins dans un premier temps, pour ne pas interférer avec l'acquisition des phonèmes. Enfin, il serait bon d'éviter un maximum la répétition enseignant-apprenant(s), qui permet difficilement aux apprenants de mémoriser les phonèmes à long terme.

Nous avons vu que les principales difficultés des apprenants japonophones étaient la distinction de certains phonèmes comme /r/ et /l/ ; /z/, /j/, /dj/ et /dz/ ; /v/ et /b/ ou encore distinguer la différence d'aperture entre /e-ɛ/ et /o-ɔ/. Au début de leur apprentissage, il est aussi important de leur faire prendre conscience de l'indépendance des phonèmes, qui n'existe pas en japonais, où le découpage phonologique s'arrête à la more. Cette prise de conscience simplifiera ensuite la perception et la production des

clusters, ou des groupes consonantiques du français, en évitant l'utilisation de voyelle d'appui entre les consonnes.

L'approche Silent Way, d'abord développée par Caleb Gattegno dans les années 70, propose d'utiliser un tableau dans lequel tous les phonèmes d'une langue donnée sont représentés par des rectangles de couleurs différentes. Au moyen d'un pointeur, l'enseignant peut pointer un son pour en demander la prononciation, cela lui évitant de le prononcer lui-même, et d'influencer l'apprenant à reproduire le son par mimétisme sans réellement se questionner sur ses caractéristiques phonologiques<sup>1</sup>. S'il prend la peine de se poser les bonnes questions, la mémorisation sera certainement meilleure, car l'apprenant retiendra des paramètres fondamentaux et non une forme générale et approximative du phonème.

Quelles sont les grandes lignes de cette approche Silent Way? Quels sont les caractéristiques des outils qu'il propose, et ceux-ci permettraient-ils de remédier aux difficultés rencontrées par nos apprenants?

## 2.1 La pédagogie du Silent Way

Le Silent Way, ou la Méthode Silencieuse, est l'œuvre de toute la vie du mathématicien et philosophe Caleb Gattegno. Ce dernier est surtout connu pour son approche innovante de l'apprentissage des mathématiques, des langues étrangères et de la lecture, mais sa pédagogie s'inscrit dans une réelle philosophie et conception du monde qui constitue la base et la charpente de son travail.

Gattegno considère que seule la conscience est éduicable. Parmi les nombreux sens qu'il donne à cette conscience, on peut retenir que c'est l'état d'être qui conduit à la réflexion dans l'action (BENSTEIN 1996), c'est à dire être présent à nous-même et à l'action que nous entreprenons pour mieux la comprendre. Ainsi l'apprentissage peut se résumer en une suite de prises de conscience qui, une fois automatisées, deviennent connaissances. Le rôle de l'enseignant n'est donc pas de transmettre le savoir à ses apprenants, mais au contraire les guider et les aider dans leur propre apprentissage.

Autrement dit, l'enseignant se concentre plus sur l'apprenant que sur la langue ; et c'est la tâche de l'apprenant que de se concentrer sur elle. L'apprenant ne doit pas

---

<sup>1</sup>C'est-à-dire quels sont les traits qui l'opposent à un autre phonème.

être un spectateur qui regarde l'enseignant mettre les mains dans la langue pour la montrer à la classe ; s'il veut efficacement apprendre, c'est à lui de mettre les mains dedans pour la tourner dans tous les sens. C'est de cette façon qu'il va faire naître des questions et des hypothèses qui vont le faire progresser dans son apprentissage. L'enseignant doit donc observer ses apprenants pour mieux leur présenter la langue et faciliter, voire provoquer des prises de conscience. Une solution à un problème ne représente rien sans le problème qui l'a fait naître. Il en est de même en classe. Si l'enseignant donne une solution (ou une règle) avant même que l'apprenant se soit posé une question, alors cette information risque d'être difficilement retenue et bien maladroitement réutilisée dans un autre contexte.

Gattegno décompose le processus d'apprentissage en quatre étapes. En premier lieu la découverte, une prise de conscience qu'il y a quelque chose d'inconnu à explorer, quelque chose à apprendre. Après quoi a lieu l'exploration elle-même, prendre conscience de ce qui existe pour planifier des tentatives de résolution du problème, procéder à ces tentatives en observant leur résultat, utiliser ses connaissances pour les interpréter, et réitérer l'opération jusqu'à trouver une solution satisfaisante. Une fois la solution trouvée vient l'étape de pratique, qui consiste à répéter de différentes manières la formulation du problème initial et le résoudre afin d'automatiser le processus et le rendre cognitivement plus léger (moins d'attention nécessaire pour le résoudre). À ce stade, l'apprentissage peut être considéré comme terminé, mais une dernière étape consiste à transférer la connaissance acquise dans un autre contexte, ou l'utiliser comme ressource à la résolution d'un prochain problème (ROSLYN et MESSUM 2011).

TAB. 2.1 : Les étapes d'apprentissage selon Gattegno

Étape 1	Découverte du nouveau problème
Étape 2	Exploration & tentatives de résolution
Étape 3	Pratique et automatisation
Étape 4	Transfert de la nouvelle connaissance

On peut retrouver chacune de ces étapes dans toutes les situations d'apprentissage imaginables : une rencontre avec quelqu'un dans la rue, apprendre à conduire une voiture ou encore taper un texte à l'ordinateur. Dans une tâche aussi complexe qu'apprendre une langue étrangère, on retrouve aussi ces étapes, mais échelonnées sur

une longue période et dans différents aspects de la langue. En effet on peut être en phase de découverte d'une forme grammaticale inconnue – étape 1 – tout en sachant pertinemment prononcer les phonèmes qui la composent – étape 4. Tous les éléments produits par un apprenant peuvent donc être répartis sur une échelle d'apprentissage, suivant le point de vue de l'observation (syntaxe, intonation, prononciation etc.). La figure 2.1, inspirée de ROSLYN et MESSUM (2011), illustre cette pluralité de l'apprentissage chez un individu fictif apprenant l'anglais et formulant l'énoncé suivant « I have been living in Nice for twenty-three years ».

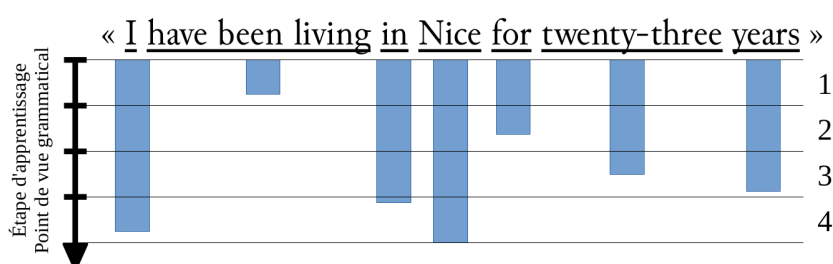


FIG. 2.1 : Illustration du processus d'apprentissage chez un apprenant d'anglais

L'important est de comprendre que l'apprentissage d'une langue étrangère est un processus complexe et dynamique, il n'y a pas « un » niveau d'apprentissage, mais autant que d'éléments produits par l'apprenant et de points de vue d'observation de la langue. Dans la figure 2.1, on peut imaginer que sur le plan grammatical, l'apprenant est déjà très familier avec le pronom personnel « I », il l'utilise spontanément et naturellement dans ce nouvel énoncé : étape 4 d'apprentissage. Le nom propre « Nice » ne pose lui non-plus aucun problème grâce à la transparence avec le français par exemple. En revanche, l'utilisation de « for » n'est pas encore acquise : l'apprenant est toujours en train d'explorer son fonctionnement et faire des hypothèses sur son utilisation – étape 2. La forme « have been -ing » est complètement nouvelle et l'apprenant est en train de se rendre compte qu'il y a un élément inconnu à explorer – étape 1. D'un point de vue phonologique, et pour le même individu au même instant, ce graphique pourrait être tout autre : les phonèmes contenus dans « have been living » sont peut-être déjà bien familiers pour l'apprenant, il a déjà fait un certain nombre d'hypothèses sur la prononciation de ces suites phonémiques, et il est en train de les automatiser – étape 3.

L'apprentissage de tous ces éléments est réparti dans le temps, et il évolue en

permanence. Chaque apprenant a un profil différent et n'acquiert pas les mêmes éléments à la même vitesse. L'enseignant doit observer attentivement la production de chacun de ses apprenants pour évaluer leur apprentissage. Ce faisant, il peut alors pointer les difficultés et guider l'apprenant en fonction de sa situation.

Nous avons vu que la dernière étape d'apprentissage consistait en le transfert de la connaissance acquise dans un autre contexte ou un prochain apprentissage. Ce principe de réutilisation des connaissances pour en acquérir de nouvelles est une des notions clés de la philosophie de Gattegno, selon laquelle tout procès fait partie d'une chronologie bien définie, une hiérarchie temporelle qui régit le monde. C'est la recherche perpétuelle d'un fonctionnement plus économe en temps comme en énergie qui est à l'origine de cet ordre chrono-logique et qui permet l'évolution (NINOMIYA 2015).

L'enseignant peut donc aider l'apprenant dans son apprentissage en l'incitant à utiliser tantôt son bagage d'acquis, tantôt son intuition et son raisonnement pour résoudre le problème, suivant la situation à laquelle il fait face.

Toutes les prises de conscience qui constituent l'apprentissage ne sont pas de même importance. Selon la tension produite par le problème, et le temps et l'énergie mobilisés pour sa résolution, la découverte de la solution (et donc la prise de conscience de la solution) peut être assez intense pour entrer dans les livres d'histoire (comme le fameux *Eurêka* d'Archimède) ou au contraire peut être insignifiante au point que l'on ne s'en aperçoit même pas. ROSLYN et MESSUM (2011) parlent d'un continuum d'importance des prises de conscience selon l'énergie générée par la dis-solution de la tension créée par le problème. BENSTEIN (1996) décrit la prise de conscience comme une « gélification de la conscience », une solidification, à un moment donné, de la conscience, la rendant tangible et plus facilement observable.

Mais pour permettre ces prises de conscience, l'apprenant doit être mentalement présent. NINOMIYA (2015, p. 15) nous donne une définition philosophique de la présence, tirée du dictionnaire de l'Académie française 9<sup>ème</sup> édition, expliquant qu'être présent au monde, c'est « appréhender celui-ci comme un objet de perception, et de se considérer soi-même comme un élément constitutif de ce monde ». L'apprenant doit être totalement ouvert à ce qu'il perçoit d'inconnu, et ne doit pas se focaliser sur ses expériences et connaissances passées pour contourner le problème. Ce n'est pas la connaissance elle-même qui est importante, mais le processus que l'on met en



place pour l'obtenir. Ainsi un transfert direct du savoir de professeur à élève n'aurait que peu de valeur ; mais c'est l'éducation de la conscience – et la manière de prendre conscience – qui est réellement bénéfique pour l'apprenant. C'est donc le devoir de l'enseignant d'éduquer cette conscience. BENSTEIN (1996) décrit le rôle de l'enseignant comme transmetteur de savoir-faire (et non de connaissances), favorisant la tension chez l'apprenant (la naissance de problèmes engendrant la recherche de leur solution), et qui utilise le silence pour fournir l'espace dans lequel l'apprenant peut parvenir à une prise de conscience.

« Être présent » (ou *"being present"*) à une situation peut être opposé à « faire attention » (ou *"paying attention"*) à une situation si l'on considère qu'être présent ne coûte aucune énergie – on est captivé sans forcément y faire attention – et où au contraire « faire attention » nécessite une énergie pour se concentrer sur la situation en question. ROSLYN et MESSUM (2011) proposent comme exemple la lecture d'un bon roman, pour laquelle aucun effort de concentration ne semble nécessaire car l'intrigue de l'histoire nous maintient sous tension. Cependant lorsqu'un évènement extérieur survient, il faut fournir une certaine énergie pour sortir du roman et se préoccuper de l'évènement en question. Il faut alors qu'une nouvelle tension se crée, et quand celle-ci sera supérieure à celle de la précédente activité alors on devient présent au nouvel évènement, et aucune énergie de concentration ne semble plus nécessaire. Constatez que nous sommes généralement bien plus rapidement présents au livreur de pizza qui sonne à la porte, qu'à notre chat qui miaule pour avoir à manger. Tout est question de force de tension créée par le nouvel évènement.

Le Silent Way est donc une approche qui se centre sur les besoins de l'apprenant et sur son rythme d'apprentissage naturel, l'enseignant est là pour le former à des savoir-faire lui permettant d'apprendre la langue par lui-même. Selon cette approche, il est plus pertinent de pointer les erreurs pour guider l'apprenant plutôt que de corriger. L'erreur et la correction sont des phases essentielles de l'apprentissage, et les enlever à l'apprenant rend l'enseignement moins efficace. L'apprenant doit être acteur de son apprentissage, il doit faire des tentatives, des erreurs puis se corriger pour mieux comprendre le fonctionnement de la langue. C'est en cela que cette approche est dite « silencieuse », l'enseignant n'est pas muet, mais les connaissances sur la langue sont directement formulées par les apprenants, et l'enseignant se limite à les guider, à leur donner un constant *feedback* sur leur production, et c'est lui qui met en contact la

classe avec tel ou tel aspect de la langue, dans une chronologie qu'il définit.

## 2.2 Silent Way & phonologie

ROSLYN et MESSUM (2011, p. 88) décrivent le processus d'enseignement / apprentissage d'un nouveau phonème avec l'approche silencieuse. Pour initier l'apprentissage, il faut d'abord que l'apprenant prenne conscience de l'existence d'un nouveau son à apprendre. Cela peut paraître évident, mais sans cette prise de conscience, il ne peut y avoir apprentissage. Vient ensuite la deuxième étape avec l'exploration du son nouveau. L'apprenant va tenter de le prononcer, et en percevant sa production ainsi que le *feedback* de l'enseignant, il va pouvoir tenter d'ajuster l'articulation de ses organes phonatoires jusqu'à ce que le retour soit positif. S'il est à l'écoute de ces trois *feedbacks* (son ouïe, son corps et l'enseignant), l'apprenant peut ajuster sa production pour se rapprocher du son cible. Ce faisant, il va développer des critères musculaires et auditifs pour caractériser le son en question, et ce sont ces critères qui vont former la nouvelle connaissance qu'il pourra alors répéter dans différents contextes pour l'automatiser. Plus il pratique, plus il deviendra à l'aise avec ce son et moins il aura besoin de penser aux différents critères phonologiques qui le caractérisent.

L'enseignant n'est donc plus la référence sacrée de la prononciation des phonèmes de la langue étudiée. Tout locuteur ne prononce pas exactement comme son voisin, et aucune langue n'a de standard scientifique (seulement un standard politique). Si l'enseignant devient le standard de la langue étudiée, alors les apprenants auront probablement plus de mal à comprendre un autre locuteur ou leur prochain enseignant. Et qu'en est-il pour les enseignants non-natifs de la langue cible ? Dans le cas du Japon, cela amène bien souvent à une prononciation « japonaise » devenant la référence, car c'est celle de la plupart des enseignants et donc des apprenants.

Alors comment peut faire l'enseignant pour mettre en contact ses apprenants avec la phonologie de la langue étudiée, sans qu'il ne prononce mot ? Gattegno a mis au point différents outils, dont l'un d'eux sera au centre de ce mémoire. Il s'agit d'un tableau dans lequel tous les phonèmes de la langue étudiée sont représentés par des couleurs. Il existe (ou il pourrait exister) un tableau différent pour chaque langue, voire chaque variété de langue<sup>2</sup>. La disposition des sons et leur couleur peuvent mettre

---

<sup>2</sup>Gattegno a créé entre autres un tableau pour l'anglais britannique et un pour l'anglais des États-

en évidence des relations phonologiques, car ils sont regroupés par point d'articulation et par traits phonologiques. Les affriquées et les diphtongues possèdent deux couleurs. Ce tableau phonologique peut aussi bien être utilisé pour faire découvrir les phonèmes à l'apprenant, que pour analyser des constructions ou remédier à certaines prononciations tout au long de l'apprentissage de la langue.

i	e	ɛ	a	ɑ	ɔ	
y	ø	œ			o	u
	ə					wa
	ɛ̃	œ̃	ã		õ	wẽ
j						w
ɥ						
	p	t	k	ks	s	f
	b	d	g	gʒ	z	ʒ
ɲ	m	n	ɳ			v
	l	ʁ				dʒ

FIG. 2.2 : Tableau phonologique du français de Gattegno, avec transcription<sup>3</sup>

On peut constater sur ce tableau (figure 2.2) que Gattegno propose de regrouper les voyelles en haut, et les consonnes en bas ; les voyelles sont triées par apertures et position avant-arrière de la langue sur la première ligne, puis arrondissement sur l'axe vertical. Les diphtongues sont sur la droite et les nasales sont en bas, toujours en fonction de l'aperture. Pour les consonnes on retrouve les occlusives /p t k/ regroupées avec leur équivalent voisé et nasal en fonction du point d'articulation, à leur droite on trouve /ks/ et les fricatives /s f/ et leur équivalent voisé. En bas du tableau, on peut voir la liquide et l'uvulaire, les semi-voyelles sont entre les voyelles et les consonnes ; et enfin /ɲ/ et /dʒ/ sont alignés avec les occlusives nasales et les fricatives chuintantes.

Les symboles phonétiques n'apparaissent évidemment pas sur le tableau tel qu'il est utilisé en classe. C'est l'enseignant qui va les faire découvrir à la classe lors des premières séances en donnant des indices sur leurs caractéristiques. Par exemple, il pourra montrer l'ouverture de la bouche, la possibilité de tenir un phonème ou la vibration ou non des cordes vocales. Grâce à la disposition logique des sons, les apprenants sont alors très rapidement en mesure de deviner les différents phonèmes du

Unis (LE PAGE 2015).

<sup>3</sup>Source de l'image : <http://www.noborderslearning.com/>



sont utilisés pour illustrer, compléter ou étayer la communication, et constituent de véritables ressources expressives complémentaires à la parole. Pourquoi alors ne pas les utiliser pour aider ou renforcer l'enseignement/apprentissage et la construction de nouvelles connaissances ? Plusieurs méthodes ont déjà été mises au point au cours du dernier siècle, mais elles sont encore peu connues. On citera bien-sûr la méthode gestuelle de Suzanne Borelle-Maisonny, qui propose une représentation gestuelle emblématique des phonèmes du français dans les années 60 ; ou l'approche haptique de la lecture de Florence Bara, Édouard Gentaz et Pascale Colé qui font utiliser à leurs jeunes apprenants le sens du toucher pour rendre plus efficace leur apprentissage et remédier aux déficiences éventuelles de ces derniers.

Comme l'a démontré Piaget dans les années 50, puis entre autres Goldin-Meadow par la suite, l'enfant utilise le geste comme stratégie d'étayage pour parvenir à des tâches complexes comme la production d'énoncés à deux mots (qui commence pas une séquence mot+geste) ou l'explication d'opérations mathématiques dont le concept n'est pas encore acquis (TELLIER 2010). COLLETTA (2009) rappelle que la plupart des acquisitions cognitives du jeune enfant proviennent de sa capacité à établir des liens intermodaux, comme la coordination entre la vision et la préhension, ou les correspondances vision-audition ou audition et mouvements phonatoires. Apprendre tout en faisant des liens entre les différents modes devrait donc être naturel.

L'utilisation du geste en classe a plusieurs fonctions, il peut servir à informer – renseignement lexical ou grammatical par exemple ; il peut évaluer – encourager, confirmer ou signaler une erreur ; et il a bien-sûr la fonction d'animer la classe et les activités. On appelle « geste pédagogique » tout geste produit par l'enseignant dans le but d'aider l'apprenant. TELLIER (2010) le décrit comme fédérateur d'attention mais aussi contributeur à l'établissement d'un climat affectif et moins formel dans la classe. Ce geste peut traduire de manière non-verbale la parole de l'enseignant ou bien être utilisé indépendamment de celle-ci. Il s'agit d'un code commun à la classe, permettant une meilleure cohésion de groupe et une progression efficace de l'enseignement.

Dans l'approche Silent Way, et notamment avec les différents tableaux conçus par Gattegno, le geste a une fonction déictique importante, il focalise l'attention de l'apprenant sur un élément en particulier ou sur un groupe d'éléments (LE PAGE 2015). Pointer les sons du tableau phonologique revient même à les « écrire », et observer ce pointage revient à les « lire » (COLLETTA 2009, p. 5). Le geste de pointage, qu'il soit

fait par l'enseignant ou par l'apprenant, permet aussi une meilleure mémorisation, car il propose une dimension supplémentaire – le geste – en plus de l'audition et de la vision. LAURENT (2004), qui propose une formidable méthode d'apprentissage de la grammaire inspirée des tableaux de Gattegno, toujours dans une approche Silent Way, utilise les coups du peinteur sur son tableau de grammaire pour ajouter une dimension rythmique, démarquant chaque mot constituant la phrase analysée.

On peut ajouter qu'une autre pratique impliquant le geste est souvent utilisée en Silent Way. Il s'agit d'utiliser ses doigts pour représenter différents phonèmes ou morphèmes, voire différents mots ; et afin de matérialiser la structure de la syllabe, du mot ou de l'énoncé et de pouvoir attirer l'attention sur tel ou tel élément, ou encore facilement représenter l'ajout ou la suppression d'un élément.

L'utilisation du geste se révèle donc être un formidable outil pour guider l'apprenant dans son apprentissage, et en particulier dans une approche Silent Way, où l'enseignant a tout intérêt à éviter de donner trop d'informations orales qui pourraient perturber le processus d'apprentissage naturel. Le geste est un outil indispensable pour l'enseignant, afin de gérer le groupe-classe, focaliser l'attention des apprenants, leur expliquer des phénomènes en les matérialisant avec les mains ou en les pointant sur un tableau. Utilisé par l'apprenant lui-même, le geste aide alors à la mémorisation et à l'explicitation d'un concept complexe en cours d'acquisition. La disposition logique des phonèmes sur un tableau, et leur distinction par des couleurs permet d'éviter l'interférence de l'écrit, pour mieux apprendre la langue (d'abord orale) et enfin pour mieux entrer dans l'écrit dans un second temps. La disposition des sons permet aussi de mettre en évidence des relations phonologiques et articulatoires souvent ignorées dans les approches traditionnelles, simplifiant la compréhension et la mémorisation des caractéristiques de chaque phonème. Car il est avant tout important que l'apprentissage soit vécu par l'apprenant. Si c'est lui-même qui formule ses problèmes et ses hypothèses pour les résoudre, alors les connaissances qui en naissent s'ancrent plus profondément en lui, et sont réutilisables plus naturellement. De plus, c'est un réel savoir-apprendre que développe l'apprenant en étant l'acteur de son apprentissage, et c'est bien cela que nous allons chercher à l'école ou à l'université.

En plus de multiplier les modalités d'apprentissage et de se centrer sur les processus que met en place l'apprenant, le Silent Way présente l'avantage d'entrer dans la langue exclusivement par l'oral, et entrer plus tard dans l'écrit grâce à des compé-

tences orales déjà développées. C'est pourtant bien souvent cet écrit qui prédomine dans l'enseignement des langues et l'enseignement en général. Saussure constatait déjà cette incohérence :

Langue et écriture sont deux systèmes de signes distincts : l'unique raison d'être du second est de représenter le premier ; l'objet linguistique n'est pas défini par la combinaison du mot écrit et du mot parlé ; ce dernier constitue à lui seul son objet. Mais le mot écrit se mêle si intimement au mot parlé dont il est l'image, qu'il finit par usurper son rôle principal ; on en vient à donner autant et plus d'importance à la représentation du signe vocal qu'à ce signe lui-même. C'est comme si l'on croyait que pour connaître quelqu'un, il vaut mieux regarder sa photographie que son visage. (Saussure, 1916/1972, p. 45, cité par DETEY (2007, p. 22))

Quant à l'apprentissage d'une langue étrangère dont le système d'écriture diffère totalement des langues que nous connaissons, il y a un intérêt évident à aborder l'oral indépendamment de l'écrit. Et le tableau phonologique de Gattegno devrait faciliter les prises de conscience des phénomènes phonologiques complètement nouveaux, en représentant par différentes modalités la réalité de la langue étudiée. Le japonais est une langue phonologiquement très éloignée du français, ainsi que nous allons le voir dans le chapitre suivant, et la présente étude se propose d'évaluer les effets d'une telle approche sur l'apprentissage du français par des étudiants japonophones.

## Chapitre 3

# Le japonais & sa phonologie

L'origine du japonais a longtemps été controversée et n'est toujours pas réellement établie. Bien que longtemps considéré comme une langue altaïque, le japonais est aujourd'hui souvent mis à l'écart, comme son voisin coréen, et placé dans la catégorie des langues isolées (NOTSU et LEVET 1997). Si son système d'écriture et une importante partie de son lexique proviennent du chinois, le système phonologique du japonais est quant à lui caractéristique des langues austronésiennes (*ibid.*). Bien que les deux langues soient très différentes sur les plans phonologique et morphologique, le japonais a été profondément influencé par le chinois dès le début du v<sup>e</sup> siècle, avec l'arrivée du Confucianisme et de l'écriture.

Afin de mieux comprendre les difficultés phonologiques des locuteurs japonais, il est essentiel d'étudier leur langue. L'écriture est arrivée tardivement sur l'archipel, mais elle a bouleversé le système phonologique japonais qui a dû s'adapter à une arrivée massive de termes chinois et des sinogrammes. Les contacts de langues sont en effet la cause de l'extrême complexité du système d'écriture japonais, et de l'originalité de son système phonologique actuel. Tout au long de ce chapitre, je ferai référence à différentes couches lexicales de la langue. Les combinaisons phonétiques permises en japonais varient en fonction de l'origine du mot – japonais, chinois ou étrangère – et il sera donc fait une brève description des trois « strates lexicales » du japonais avant de commencer une présentation du système d'écriture. S'en suivra une analyse détaillée du système vocalique puis consonantique de la langue, et enfin seront abordés le rythme et la prosodie japonaise.



### 3.1 Stratification lexicale

Il est nécessaire de prévenir le lecteur que la langue japonaise se voit composée de trois couches lexicales différentes. Ces trois couches sont le fruit de contacts interlangues ayant eu lieu tout au long de l'histoire, et qui ont introduit sur l'archipel des mots de langues phonologiquement éloignées, apportant au japonais de nouvelles combinaisons phonémiques. Ces trois couches sont connues sous les noms de *yamato-gotoba*, *kango* et *gairaigo*.

La première couche lexicale est composée de l'ensemble des termes japonais autochtones ; ils étaient déjà utilisés sur l'archipel avant les premiers échanges linguistiques avec le continent. On l'appelle généralement *yamato-gotoba* やまと言葉<sup>1</sup> ou *wago* 和語. Sur le plan phonologique, LABRUNE (2006) constate que dans le lexique *yamato* encore utilisé aujourd'hui, les consonnes palatalisées sont rares, que /w/ n'apparaît plus que devant /a/, /h/ n'existe pas à l'intervocalique, /p/ n'existe plus (KINDAICHI et MAËS (1978) signalent toutefois quelques occurrences dans les adverbes et les impressifs), les consonnes sourdes situées immédiatement après la nasale-more /N/ sont quasiment inexistantes, et aucune obstruante voisée ni le /r/ n'est géminée.

Avec l'introduction de la religion et de la philosophie du Continent et l'augmentation des échanges entre la Chine et le Japon à partir du IV<sup>e</sup> siècle, une très grande quantité d'emprunts à différentes variétés du chinois sont arrivés sur l'archipel. Le japonais s'est alors « sinisé », et un lexique dit sino-japonais, ou *kango* 漢語 a vu le jour<sup>2</sup>. Parmi ces lexèmes, LABRUNE (2006) dénombre une grande quantité de consonnes palatalisées, de nombreuses nasal-more /N/, beaucoup de géminations et d'allongements, l'absence d'un /p/ non géminé ainsi que l'inexistence d'obstruantes sonores géminées. LABRUNE résume ainsi l'ensemble des structures dérivant du chinois :

<sup>1</sup> litt. « mots yamato », Yamato étant l'ancien nom du Japon (KAMATA et YONEYAMA 2001).

<sup>2</sup> Notons que cette paire de sinogrammes signifie « chinois mandarin » en mandarin (litt. Langue des Han)(KAMATA et YONEYAMA 2001).

$$(C)_{(y)}V \left\{ \begin{array}{c} N \\ R \\ i \\ t/k + u/i \end{array} \right\} \quad (3.1)$$

Une voyelle V obligatoire peut-être précédée et/ou suivie de phonèmes optionnels : une consonne C avec ou sans palatalisation *y* ; et suivie d'une nasale-more /N/, d'un allongement de V (noté R par l'auteure) ou encore une consonne /t/ ou /k/ suivie de /u/ ou /i/ (LABRUNE 2006, p. 31). LABRUNE ne présente que les caractéristiques phonologiques observables dans le japonais standard actuel, mais si l'on se penche sur la variation diachronique de la langue, on constate que les Japonais ont assimilé des combinaisons phonémiques chinoises tout à fait originales mais qui ont fini par disparaître au fil des siècles. Lorsque les textes étaient encore rédigés en chinois, il sembleraient qu'ils aient aussi été prononcés de manière assez similaire au chinois de l'époque (LIU et WASHIO 2010). L'intégration d'un grand nombre de lexèmes chinois dans la langue japonaise du peuple a familiarisé ce dernier avec les formes /kwa gwa kwi gwi kwe gwe/ ou encore avec des finales consonantiques comme -k -t -p -m -n ou -ŋ. Plus aucune de ces formes n'existent aujourd'hui, si ce n'est la finale -n qui a plusieurs réalisations nasales. LIU et WASHIO (2010, p. 170) donnent comme exemple les mots \**kigwai* 奇怪 (étrange) et \**bengwe* 變化 (évolution) qui étaient prononcés de la sorte dans la langue quotidienne à partir de l'ère Heian (794-1185). Ces réalisations perdurent jusqu'au XVIII<sup>e</sup> siècle, puis se confondent avec /ki gi ke ge/, /kwa gwa/ quant à lui ne résiste que jusqu'au XV<sup>e</sup> siècle, puis sont fréquemment confondus avec /ka/ et /ga/, et enfin disparaissent totalement à la fin de l'ère Edo (1603-1867). \**kigwai* et \**bengwe* sont devenus aujourd'hui *kikai* et *henka*.

Les finales consonantiques disparaissent aussi vers le XVIII<sup>e</sup> siècle : -ŋ se confond avec -u, -m et -n se mélangent en une seule finale nasale -N, -k devient -ki ou -ku, -t devient -tu et -p devient -hu puis -u. Des transcriptions de missionnaires chrétiens au XVI<sup>e</sup> siècle prouvent sa présence jusqu'à cette période<sup>3</sup>, puis à partir du XVII<sup>e</sup> on ne trouve plus de trace de coda consonantique (LIU et WASHIO 2010).

<sup>3</sup>On peut y lire notamment les mots \**butmet* 仏滅 (mort de Bouddha) ou \**bat* 罰 (amende), aujourd'hui *butumetu* et *batu* (LIU et WASHIO 2010, p. 171).

À partir du xvi<sup>e</sup> siècle, des missionnaires européens, principalement portugais et espagnols, arrivent au Japon pour prêcher la Parole de leur dieu. C'est le début des relations avec l'Occident. À partir de cette période un nombre croissant d'emprunts aux langues européennes va donner naissance à une troisième couche lexicale appelée *gairaigo* 外来語 (mots venant de l'extérieur). On assiste alors à l'apparition progressive de nouvelles consonnes comme /f/ ou /v/ ainsi que de nombreuses combinaisons phonologiquement impossibles jusqu'alors (KINDAICHI et MAËS 1978).

Ainsi le japonais moderne est un mélange de ces trois lexiques, et si certains concepts comme « la télévision » n'existent que dans une des trois couches (*gairaigo* : *terebi* テレビ) certains possèdent des équivalents dans deux, voire trois couches. Par exemple, « le riz » a pour version autochtone *mesi* 飯 ; pour version sino-japonaise *gohan* ご飯 et pour version occidentale *raisu* ライス, de l'anglais *rice*. Il est intéressant de constater que la version sino-japonaise est plus formelle que les deux autres, et que la version occidentale fait plutôt référence à la culture étrangère et/ou à la modernité, en l'occurrence le riz cuisiné de manière exotique (LABRUNE 2006).

### 3.2 Le système d'écriture

Le japonais était une langue exclusivement orale jusqu'à l'adoption des sinogrammes quelques siècles après l'arrivée des premiers textes philosophiques en provenance du continent. Au contact de ces écrits, les Japonais vont adopter le système sinogrammique, et créerons de nouveaux caractères à partir de celui-ci pour optimiser la transcription de leur langue. Les sinogrammes utilisés au Japon sont appelés *kanji* 漢字, et les nouveaux caractères purement japonais constituent deux syllabaires appelés *hiragana* 平仮名 et *katakana* 片仮名.

Selon LIU et WASHIO (2010), les premiers écrits furent apportés par le missionnaire coréen *Wani* 王仁 en 285 après J.-C., venu enseigner la sagesse du Confucianisme dans l'archipel japonais. À partir de 630, et pendant plus de deux siècles, le Japon va envoyer des émissaires pour se former à la religion bouddhique et à la politique chinoise. Dès cette période, un grand nombre de textes vont arriver sur l'archipel et le contact avec l'écriture deviendra rapidement inévitable.

Seulement, le japonais est une langue agglutinante tandis que le chinois une langue isolante, et qui plus est avec un système phonologique radicalement différent

du japonais. L'assimilation des sinogrammes pour transcrire la langue nippone n'a donc pas été chose facile et cela a abouti à un système particulièrement complexe (LIU et WASHIO 2010). LABRUNE (2006, p. 22) le décrit même d'« unanimement reconnu comme l'un des plus complexes qui soit ». Aussi, c'est le chinois classique qui fut écrit tel quel pendant plusieurs siècles par les érudits japonais, qui s'efforçaient de respecter syntaxe et grammaire traditionnelle.

Le *Koziki* 古事記 est le plus ancien texte écrit par une main japonaise, dont on ait la trace aujourd'hui. Il est écrit en 712 par le noble écrivain-bureaucrate *Ō no Yasumaro* 太安万侶 sous l'ordre de l'Impératrice *Genmei* 元明. Il s'agit d'un recueil de mythes sur les origines du Japon et les Dieux. Il est accompagné huit ans plus tard par le *Nihon-syoki* 日本書紀, plus élaboré et plus détaillé que le *Koziki*. Ces ouvrages ont été d'une aide considérable pour les historiens japonais, puisqu'ils constituent les textes les plus complets concernant les débuts de l'histoire japonaise (même si une grande partie reste une compilation de légendes). Ceux-ci sont entièrement rédigés en sinogrammes et plus ou moins selon la grammaire chinoise de l'époque.

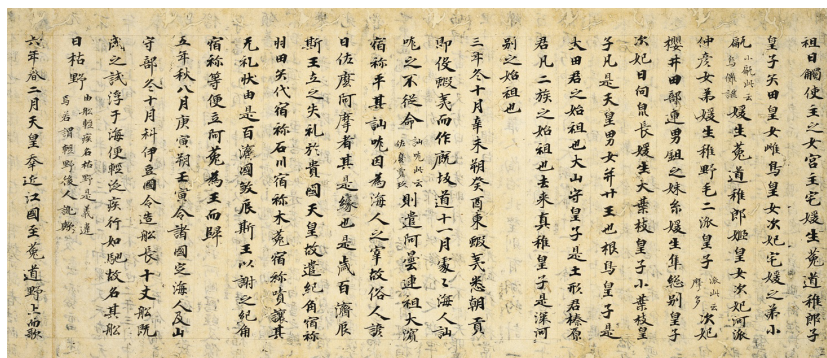


FIG. 3.1 : Extrait du Nihon-syoki (720)

D'après LABRUNE (2006), les emprunts au chinois se seraient faits en trois vagues, donnant naissance à trois lectures différentes, nommées *on'yomi* 音読み ("lecture phonétique") pour de nombreux sinogrammes. La première vague aurait eu lieu pendant l'introduction du bouddhisme entre le v<sup>e</sup> et le vi<sup>e</sup> siècle, donnant naissance à la lecture *go'on* 呉音 ; une deuxième vague entre le vii<sup>e</sup> et le viii<sup>e</sup> siècle, plus importante, apportant la lecture dite *kan'on* 漢音, et enfin une vague d'emprunts plus tardifs donnant la lecture *tō'on* 唐音<sup>4</sup>. Chacune de ses lectures correspond à l'équivalent chinois

<sup>4</sup>LIU et WASHIO (2010) reconnaissent une quatrième catégorie : les prononciations populaires ou

de l'époque, et de la région d'emprunt. Les lectures *go'on* proviendraient du royaume de Wú 吳 (Go en japonais) au sud-est de la Chine, l'un des trois royaumes se battant pour la domination du pays après la chute de la dynastie des Hàn en 220, et jusqu'à la réunification par la dynastie des Jìn en 280 (LIU et WASHIO 2010). Les lectures *kan'on* datent des périodes japonaises Nara (710-784) et Heian (794-1185), et correspondent à la prononciation du chinois standard de la ville de Cháng'ān (actuelle Xī'ān) ancienne capitale des Hàn 漢 (Kan en japonais). Enfin les lectures *tō'on*, se sont introduites au Japon pendant les périodes Kamakura (1185-1336) et Muromachi (1338-1573), avec le mouvement bouddhique du Zen (LIU et WASHIO 2010).

En plus des lectures chinoises, les sinogrammes se verront rapidement attribuer une ou plusieurs lectures purement japonaises, appelées *kun'yomi* 訓読み ("lecture sémantique"). Celles-ci voient le jour pour exprimer en japonais des concepts difficiles et simplifier la compréhension pendant la lecture en passant par le lexique autochtone japonais (LIU et WASHIO 2010).

Le japonais a donc cette particularité de proposer un grand nombre de lectures pour un même sinogramme. Prenons pour exemple le caractère 行 (principalement "ligne" ou "aller"), prononcé en ancien chinois \*hǎŋ, il peut se prononcer en lecture sino-japonaise *gyō* (*go'on*), *kō* (*kan'on*) ou *an* (*tō'on*) et en lecture japonaise *yuki*, *yu* ou *i* selon le contexte sémantique (LABRUNE 2006). Certains sinogrammes n'ont qu'une seule lecture possible en japonais, d'autres sont connus pour en avoir plus de dix<sup>5</sup>.

Face à la complexité d'un tel système d'écriture, certains commencèrent à utiliser les sinogrammes seulement pour leur valeur phonétique afin d'écrire leur langue. Ainsi, un nombre très réduit de caractères suffisait à écrire n'importe quel mot japonais. À chaque syllabe (ou plutôt à chaque more, cf. §2.5) correspondait un ou plusieurs sinogrammes ; par exemple le phonème /a/ pouvait se transcrire avec l'un des sinogrammes suivant 安, 愛, 悪, 阿... sans faire appel au sens de ces caractères – respectivement calme, amour, mal et coin (LIU et WASHIO 2010). Ce besoin d'utiliser les sinogrammes pour transcrire directement la langue japonaise sans passer par le chinois est né de l'intention de transcrire des poèmes. On appelle aujourd'hui ce système *man'yōgana* 万葉仮名, nom dérivé du *Man'yōshū* 万葉集, la première an-

d'usage, ou *kan'yō'on* 慣用音.

<sup>5</sup>Le dictionnaire Kangorin compte par exemple 13 lectures courantes et actuelles pour le sinogramme 生 ("vivre", entre autre) (KAMATA et YONEYAMA 2001).

thologie de waka (poèmes japonais), datant de la fin du VIII<sup>e</sup> siècle, et entièrement rédigée en *man'yōgana*. C'est à cette période (Heian, 794-1185) que le Japon connaît une transformation importante sur les plans social, politique ou encore artistique. Les échanges avec la Chine se font plus rares, et l'aristocratie japonaise se développe. On peut penser que cette volonté de ne plus utiliser la langue chinoise pour écrire le japonais est liée aux tendances générales de l'époque à s'éloigner du continent. Les *man'yōgana* sont une occasion inespérée pour la population d'accéder à l'écriture sans devoir retenir plusieurs milliers de caractères et une syntaxe étrangère.

Ce sont notamment les femmes de la haute bourgeoisie qui vont s'emparer de ce système, en le stylisant et le simplifiant, et de là naîtront les *hiragana*. Il est alors nécessaire de citer *Le Dit du Genji* (ou *Genji monogatari* 源氏物語), roman monumental de la littérature japonaise que l'écrivaine Murasaki Shikibu écrivit au XI<sup>e</sup> siècle. La naissance des *hiragana* est étroitement liée à cette œuvre, quasi-exclusivement rédigée avec ce système (NHK 2015).

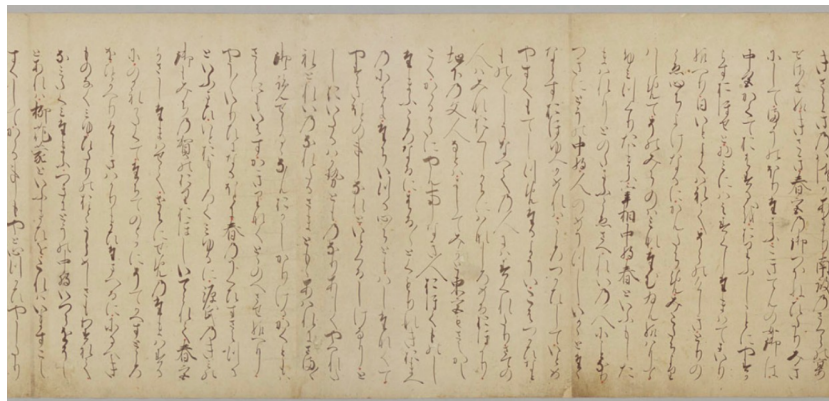


FIG. 3.2 : Extrait du manuscrit du Dit du Genji (XI<sup>e</sup> siècle)<sup>6</sup>

Un autre syllabaire fut créé parallèlement aux *hiragana* : les *katakana*. Il s'agit en réalité du même système d'écriture, mais d'un script légèrement différent créé dans des conditions différentes. Les *katakana* auraient été créés dès le IX<sup>e</sup> siècle par des religieux souhaitant lire des prières en japonais (et non en chinois). Des moines auraient alors utilisé certains *man'yōgana* pour annoter leurs textes. Avec le temps et l'habitude, ces annotations furent abrégées et acquirent leur forme actuelle (LIU et WASHIO 2010). On peut constater que certains *katakana* sont très semblables à leur

<sup>6</sup>Source : Librairie digitale de l'Université Kokugakuin

équivalent *hiragana*, il s'agit dans ces cas-là des mêmes sinogrammes qui en sont à l'origine.

Les *hiragana* et les *katakana* sont deux variantes d'un seul et même système appelé *kana* 仮名. On parle souvent de syllabaire, mais il serait plus rigoureux de parler de « moraire » puisque l'unité phonémique minimale du japonais est la more, et le statut de syllabe est largement remis en question (LABRUNE 2006). Les mores sont donc les plus petites parties d'un mot qu'un Japonais soit capable d'épeler dans sa langue. Par exemple, les épinards sont épelés é-p-i-n-a-r-d-s en français mais leur équivalent japonais *hōrensō* [ho:rēnsō:] est épelé ho-u-re-n-so-u. En *kana*, on va donc écrire un caractère par more : ほ<sub>ho</sub> う<sub>u</sub> れ<sub>re</sub> ん<sub>n</sub> そ<sub>so</sub> う<sub>u</sub> en *hiragana* ou ホ<sub>ho</sub> ウ<sub>u</sub> レ<sub>re</sub> ン<sub>n</sub> ソ<sub>so</sub> ウ<sub>u</sub> en *katakana*. On s'aperçoit alors que s'il s'agissait réellement d'un syllabaire, on l'épellerait hō-ren-sō. Le tableau 3.3 présente l'ensemble des *kana* existants : 48 caractères pour chacun des scripts. Les caractères *we* et *wi* ne sont plus utilisés depuis 1946. Les variantes voisées g- z- d- b- sont dérivées de leur équivalent sourd, auquel on ajoute le diacritique ̣ (*dakuten* 濁点 « signe de voisement » en japonais) et p- est dérivé de h- avec le diacritique ̣ (*handakuten* 半濁点 « signe de semi-voisement »). Ainsi, *ga* s'écrit が<sup>̣</sup> ou ガ, *zu* s'écrit ず<sup>̣</sup> ou ズ, *zi* s'écrit じ ou ジ, *po* s'écrit ぽ<sup>̣</sup> ou ポ<sup>̣</sup>. Le phonème /h/ était à l'origine prononcé [p], d'où le voisement en /b/ avec le *dakuten*, le *handakuten* sera une création plus tardive pour transcrire /p/ (LABRUNE 2006).

La transcription des allongements vocaliques est peut-être la seule chose qui diffère entre les *hiragana* et les *katakana*. Les premiers les transcrivent par une voyelle indépendante à la suite de la more dont la voyelle est allongée (かあ kaa, きい kii, くう kuu, けい kei, こう kou); les *katakana* quant à eux utilisent une barre latérale symbolisant l'allongement et commune pour toutes les voyelles (カー kaa, キー kii, クー kuu, ケー kee ou kei, コー koo ou kou<sup>7</sup>).

Il faut aussi ajouter à ce tableau des *kana*, les formes palatalisées de certaines consonnes. Celles-ci se transcrivent à l'aide du *kana* en -i, suivi d'un *kana* ya, yu ou yo en demi-taille (le caractère est imprimé plus petit). Par exemple *kya* s'écrit きゃ (ki+ya) ou キヤ, *pyo* s'écrit ぴょ (pi+yo) ou ピョ, *tyu* s'écrit ちゅ (ti+yu) ou チュ. Et enfin la gémination d'une consonne s'écrit à l'aide du *kana* tu en demi-taille, suivi de la consonne à géminer : *nippon* s'écrit donc に<sub>ni</sub>つ<sub>tu</sub> ぽ<sub>po</sub> ん<sub>n</sub> ou bien ニ<sub>ni</sub>ツ<sub>tu</sub> ポ<sub>po</sub>

<sup>7</sup>il arrive de trouver aussi la voyelle indépendante à la suite de la more, comme en *hiragana*, lorsqu'il s'agit d'une transcription de sinogrammes.

	hiragana katakana phonétique	hiragana katakana phonétique	hiragana katakana phonétique	hiragana katakana phonétique	hiragana katakana phonétique
	<b>a</b>	<b>i</b>	<b>u</b>	<b>e</b>	<b>o</b>
∅	あ ア [a]	い イ [i]	う ウ [ü]	え エ [e]	お オ [o]
<b>k</b>	か カ [ka]	き キ [ki]	く ク [kü]	け ケ [ke]	こ コ [ko]
<b>s</b>	さ サ [sa]	し シ [ɕi]	す ス [sü]	せ セ [se]	そ ソ [so]
<b>t</b>	た タ [ta]	ち チ [tɕi]	つ ツ [tsü]	て テ [te]	と ト [to]
<b>n</b>	な ナ [na]	に ニ [ni]	ぬ ヌ [nü]	ね ネ [ne]	の ノ [no]
<b>h</b>	は ハ [ha]	ひ ヒ [çi]	ふ フ [fö]	へ ヘ [he]	ほ ホ [ho]
<b>m</b>	ま マ [ma]	み ミ [mi]	む ム [mü]	め メ [me]	も モ [mo]
<b>y</b>	や ヤ [ja]		ゆ ユ [jy]		よ ヨ [yo]
<b>r</b>	ら ラ [ra]	り リ [ra]	る ル [rü]	れ レ [re]	ろ ロ [ro]
<b>w</b>	わ ワ [wa]	ゐ ヰ [i]		ゑ ヱ [e]	を ヲ [o]
<b>N</b>	ん ン [N]				

FIG. 3.3 : Les *kana* japonais et une transcription phonétique possible



ン<sub>n</sub> en *katakana*.

Aujourd'hui ces deux scripts sont utilisés quotidiennement aux côtés des sinogrammes. Les *hiragana* notent les flexions grammaticales (on les appelle alors des *okurigana* 送り仮名), les interjections, ils peuvent aussi noter un morphème lexical que le scripteur ne veut pas ou ne sait pas écrire en sinogrammes. Les *katakana* quant à eux sont utilisés pour les mots d'origine étrangère, les onomatopées, impressifs ou idéophones, les formes dialectales ou argotiques, et ils peuvent avoir une fonction d'emphase. Les jeunes japonais commencent par apprendre les *kana* dès leur première année d'école, puis petit à petit les sinogrammes<sup>8</sup>. Les livres pour enfants sont généralement entièrement rédigés en *hiragana*, puis les sinogrammes sont progressivement introduits, souvent avec leur lecture en *hiragana* suscrits, que l'on appelle alors *furigana* ou *rubi*.

Avec l'arrivée récente de nouvelles combinaisons phonémiques, et de nouveaux phonèmes dans le japonais moderne, des graphies adaptées ont peu à peu vu le jour pour transcrire ces nouveaux sons provenant de l'étranger. Ainsi, on peut par exemple écrire テイ(*te+i* en demi-taille) pour écrire le son [ti]. Ces nouvelles graphies ne sont utilisées qu'avec le syllabaire *katakana* et n'apparaissent que sur les emprunts relativement récents.

On constate donc que de nouvelles combinaisons graphiques sont apparues, mais qu'en est-il sur le plan phonétique ? Bien qu'il soit possible d'écrire et prononcer mon prénom /siruvan/ シルヴァン [ɕirüvãŋ], avec le nouveau graphème /va/, il est plus courant de lire et entendre /siruban/ [ɕirübãŋ] voire [ɕirübãŋ]. Et le fait d'écrire un *va* en *katakana* ne signifie pas qu'il sera prononcé [va]. L'écriture et la réalisation phonétique sont donc deux choses bien distinctes. Mais il semble évident que les nouvelles combinaisons phonémiques ne peuvent se voir attribuer des graphèmes spécifiques qu'après avoir été plus ou moins intégrées à l'oral. C'est donc d'abord par la langue parlée que les nouvelles possibilités arrivent, puis, après un certain temps, se distinguent à l'écrit. Cela n'empêchant pas un grand nombre de personnes, surtout les

<sup>8</sup>Selon le [Ministère de l'Éducation](#), 1006 sinogrammes sont appris durant les 6 années d'école primaire, et 1130 autres pendant les 3 ans de collège. Un total de 2131 kanji constitue aujourd'hui la liste des *zyōyō kanzi* 常用漢字, les « caractères d'usage courant ». Le nombre total de sinogrammes utilisés en japonais n'est pas déterminé, mais les dictionnaires sino-japonais ordinaires en contiennent généralement 6000 (KAMATA et YONEYAMA 2001) c'est aussi le nombre de caractères qu'il est possible de taper à l'ordinateur avec un clavier classique, et les dictionnaires plus exhaustifs, comme le fameux *Daikanwa Ziten* vont jusqu'à 50000 sinogrammes (MOROHASHI 2000).

plus âgées, de ne jamais faire la distinction à l'oral comme à l'écrit. De plus, la période d'emprunt a une grande influence sur la prononciation ou la transcription d'un mot : la forme d'emprunt original est généralement conservée. Par exemple le mot « Tibet » se dit [tɕibetto] et s'écrit avec le *katakana* チ [tɕi] ; pourtant le mot « T-shirt » est couramment prononcé [ti:ɕʰatsɯ] et s'écrit avec la nouvelle combinaison te+i テイ [ti]. La raison à cette différence n'est autre que l'âge de l'emprunt, le premier datant d'une époque où la combinaison /ti/ n'était pas encore répandue. LABRUNE (2006, p. 109) précise que les nouvelles consonnes ou combinaisons phonémiques ne sont quotidiennement utilisées que chez certains locuteurs « dont le profil socio-culturel est plutôt urbain, éduqué et féminin ».

On peut voir qu'il reste impossible de transcrire plusieurs consonnes à la suite, sauf dans le cas du /N/.

### 3.3 Les voyelles

Il y a 5 voyelles en japonais standard actuel : / a i u e o /. Elles sont globalement moins tendues qu'en français, et la labialité est peu marquée. Le /a/ est réalisé comme un [a] central, parfois légèrement postérieur après une vélaire. Le /i/ est un [i] peu tendu, parfois transcrit [ɪ] s'il est considéré comme moins antérieur. Le /u/ est une voyelle postérieure haute légèrement arrondie, beaucoup moins labiale que le /u/ français : on la transcrit [ɯ] ou [ʊ]. Elle est centralisée après /s t z/ et les consonnes palatalisées : [y̥] (ou [y̥]) (LABRUNE 2006). Le /e/ se situe entre [e] et [ɛ] ; /o/ entre [o] et [ɔ].

Les voyelles /i/ et /u/ ont un statut particulier en japonais. Elles modifient sensiblement la réalisation des consonnes /t d h s z/. Elles sont aussi souvent dévoisées ([i̥] [ɯ̥]), voire élidées après les consonnes sourdes. Certains de ces effacements vocaliques sont totalement lexicalisés (comme /gakki/ (semestre) présent tel quel dans les dictionnaires, mais provenant de /gakuki/), d'autres ne sont pas encore lexicalisés (comme /gakusei/, réalisé généralement [gakse:] ou [gakɯse:]) (LABRUNE 2006).

Les groupes consonantiques des emprunts aux langues étrangères sont évités au moyen de trois voyelles d'appui (ou épenthèses) : /u/ (principalement) ; /i/ après [ɕ tɕ] et parfois [k] ; et /o/ après [t d]. Voici quelques exemples :

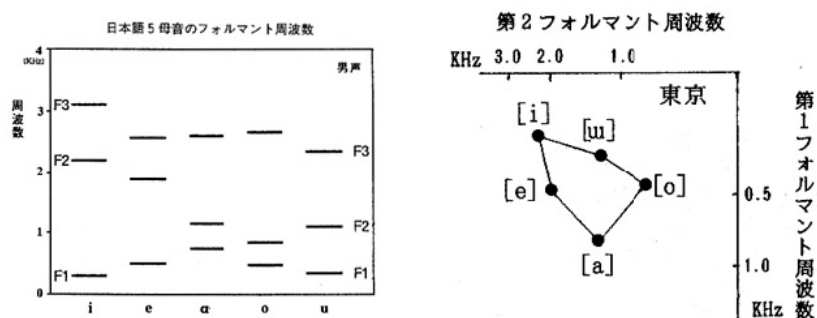


FIG. 3.4 : F0, F1, F2 des voyelles et triangle vocalique du japonais (Sugito, 1995, cité par KAMIYAMA et VAISSIÈRE (2009))

- /furansu/ フランス (France)
- /amusuterudamu/ アムステルダム (Amsterdam)
- /gurunooburu/ グルのーブル (Grenoble)
- /sutoresu/ ストレス (stress)
- /supotto/ スポット (spot)
- /kirishitan/ キリシタン (chrétien, du portugais cristão)

Notons que certaines voyelles épenthétiques sont parfois très réduites et il arrive qu'elles soient même invisibles au niveau phonologique, mais toujours présentes au niveau phonétique (notamment au niveau du déplacement accentuel (LABRUNE 2006)).

Le japonais présente un grand nombre de voyelles longues, surtout dans les mots sino-japonais (*kango*) et les emprunts aux langues occidentales (*gairaigo*). On note très peu de différence de timbre entre une voyelle longue et une voyelle courte, excepté pour le /e/ qui est souvent diphtongué lorsqu'il est allongé : [ej]. Les voyelles longues sont généralement translittérées ā, ī, ū, ē et ō soit respectivement aa, ii, uu, ei, ou. La voyelle ō et bien souvent la voyelle ē ne sont pas diphtonguées bien qu'elles s'écrivent en *hiragana* /ou/ おう et /ei/ えい.

L'existence de diphtongues à proprement parler est controversée. Beaucoup de phonéticiens japonais reconnaissent leur existence mais ne les définissent pas de la

même manière. Quoi qu'il en soit, toutes les combinaisons possibles entre les cinq éléments du système vocalique sont présentes (iu, ie, io, ia, ui, ue, uo, ua, ei, eu, eo, ea, ai, au, ae, ao, oi, ou, oe et oa) (LABRUNE 2006).

### 3.4 Les consonnes

Les nombres de consonnes et de possibilités phonotactiques varient beaucoup en fonction des strates lexicales (mots japonais autochtones, mots sino-japonais ou emprunts aux langues européennes).

L'inventaire des phonèmes consonantiques du japonais est rendu particulièrement complexe par le nombre d'allophones que possèdent la plupart des consonnes. Les réalisations de celles-ci ont en effet tendance à varier selon le contexte vocalique. De plus, tous les auteurs ne prennent pas en compte de la même manière les réalisations phoniques réalisées par les japonophones dans la prononciation des emprunts récents aux langues européennes. Les phonéticiens parlent de système minimal ou maximal, selon qu'ils considèrent ou non les nouvelles réalisations. Par exemple, l'occlusive apico-dentale [t] est communément considérée comme unité phonologique indépendante du japonais /t/, pourtant elle devient affriquée chuintante devant /i/ ([tɕ]) ou sifflante devant /u/ ([tɕʰ]) et il arrive que ces réalisations soient considérées comme autres phonèmes indépendants, plutôt qu'allophones d'un archiphonème /T/ (indépendance de la consonne /tɕ/ chez LABRUNE (2006) qui considère pourtant [tɕ] comme un allophone de /t/; KINDAICHI et MAËS (1978) considèrent les deux comme allophones de /t/; HATTORI (1960) ou BLOCH (1950, cité par KINDAICHI et MAËS (1978)) les considèrent comme trois phonèmes à part entière /t/, /tʃ/ et /tʃʰ/).

Pour établir un inventaire des consonnes japonaises, il est donc nécessaire de poser sa propre définition d'une unité consonantique. Le présent travail n'a pas pour objectif de remettre en question les inventaires précédents, et il sera fait principalement référence à l'inventaire de KINDAICHI et MAËS (1978)<sup>9</sup>.

En plus des 5 voyelles présentées dans la section précédente, KINDAICHI compte 17 phonèmes consonantiques.

L'auteur résume le système consonantique japonais en quatre ordres articula-

<sup>9</sup>Cela en raison de la simplicité et l'économie de leur système phonique, privilégiant les archiphonèmes aux multiples phonèmes indépendants.

	Bilabiales	Apico-dentales	Dorso-alvéolaires	Dorso-vélaire	Liquide r
Sourdes	p	t	s	k	
Sonores	b	d	z	g	Laryngale h
Nasales	m	n		ŋ	

Semi-voyelles	
Bilabiale	Palatale
w	j
Phonèmes-mores	
N	Q

TAB. 3.1 : Tableau des phonèmes consonantiques d'après KINDAICHI et MAËS (1978, p. 11)

toires (bilabial, apico-dental, dorso-alvéolaire et dorso-vélaire) pour produire un son sourd ou voisé, oral ou nasal ; à quoi s'ajoutent la liquide /r/ et la laryngale /h/, les semi-voyelles /w/ et /j/, ainsi que deux phonèmes spéciaux dits phonèmes-mores ou phonèmes moriques. À ce tableau, on ajoutera la nouvelle consonne /v/ et de nouvelles combinaisons phonémiques qui seront présentées en fin de section.

### 3.4.1 Consonnes variant peu (ou pas) selon le contexte vocalique

Les consonnes /p b m n k g r/ ne présentent pas de variation remarquable en fonction de la combinatoire vocalique, elles sont donc réalisées de manières assez similaires quelque soit la voyelle qui suit ou précède.

La consonne /p/ est réalisée en occlusive bilabiale sourde dans tous les contextes. La distribution de ce phonème est intéressante : il n'est présent, dans le lexique *yamato*, que dans certains adverbess et morphèmes expressifs (p. ex. /perapera/ ぺらぺら "couramment"; /poroporo/ ぽろぽろ "bruit d'un liquide s'écoulant en abondance"). Dans les morphèmes sino-japonais, il n'apparaît qu'après un phonème-more (/teNpu/ 添付 "pièce jointe"; /seQpuku/ 切腹 "suicide rituel"<sup>10</sup>). Il n'apparaît sans aucune restriction dans les emprunts aux langues occidentales (/pari/ パリ (Paris); /ekusupōto/ エクスポート (export)).

La consonne /b/ est l'équivalent sonore de /p/, bien que parfois plus proche de la spirante bilabiale [β] à l'intervocalique.

<sup>10</sup>C'est le terme populaire et informel *harakiri* (腹切り) qui est parvenu jusqu'en Occident et qui est aujourd'hui dans nos dictionnaires. Cette version est pourtant très peu usitée sur l'Archipel.

Les consonnes /k/ et /g/ sont très similaires à leur équivalent français : c'est une occlusion créée par la partie centrale de la langue (le dos ou dorsum) contre le palais mou. LABRUNE les considère toutefois légèrement plus vélares qu'en français lorsqu'elles sont suivies de /e a o/. Devant /a/ notamment, en français ou en anglais, celles-ci sonnent presque palatalisées pour une oreille japonaise et cela explique le fait que la plupart des emprunts donnent /kya/ ou /gya/ en japonais là où ils sont originalement prononcés [ka] ou [ga] (p. ex. /kyarameru/ キャラメル (caramel), /kyarakutaa/ キャラクター (character), /gyaru/ ギャル (girl)). Le /g/ est souvent légèrement relâché à l'intervocalique, donnant un [ɣ]. De plus /g/ est souvent prononcée [ŋ] à l'intervocalique ou après une nasale-more par des locuteurs d'un certain âge ou des présentateurs de radio ou de télévision. Mais cette réalisation « a virtuellement disparu du système phonologique des usagers du japonais standard âgés de moins d'une quarantaine d'années » (KINDAICHI et MAËS 1978, p. 24) et c'est [g] ou son allophone [ɣ] qui est réalisé à la place. KINDAICHI considère /ŋ/ comme une consonne indépendante et non comme un allophone de /g/. Il reconnaît que l'utilisation du premier est quasi-complémentaire au second mais plusieurs exceptions amènent l'auteur à en faire un phonème à part entière. Les détails du débat ne seront pas traités ici, mais comme la réalisation [ŋ] est fréquemment remplacée par [g] ou [ɣ] par une grande partie des locuteurs, on peut en conclure qu'il n'y a pas ou plus de réelle opposition phonémique entre ces réalisations. Je la considérerai donc comme un simple allophone de l'archiphonème /g/.

Les nasales /m/ et /n/ sont réalisées comme leurs équivalents français, sans varier remarquablement selon le contexte vocalique. Il en est de même pour /k/, qui avec /p/ est légèrement plus aspiré qu'en français.

La liquide /r/ est réalisée d'un unique battement de la pointe de la langue contre les alvéoles, et correspond donc au [r] de l'alphabet phonétique international. KINDAICHI précise que certains locuteurs réalisent une latérale [l] en toute position, sauf devant /i/. LABRUNE ajoute que ces locuteurs sont souvent de jeunes locutrices, et que cette latérale est aussi courante lorsque /r/ est palatalisé. LABRUNE observe aussi d'autres réalisations pour ce phonème, comme [r] dont le roulement est vulgaire et masculin, les rétroflexes [ɽ] ou [ɽ] devant /u/, [d] à l'initiale chez certains enfants ou encore [ɽ] dans certains dialectes Ryūkyū (au sud du pays, près d'Okinawa). Il est intéressant de constater que très peu de mots japonais autochtones possèdent un /r/

en début de mot, cela est dû au fait qu'il n'existait pas en japonais archaïque. Le /r/ à l'intervocalique dans cette strate lexicale est par contre très fréquent (LABRUNE 2006).

Ces sept consonnes connaissent une variante palatalisée accompagnée d'un glide *j* entre le point d'articulation et la production de la voyelle. KINDAICHI réfléchit alors à la pertinence de considérer ces variantes comme phonèmes indépendants — dans quel cas le glide ferait partie intégrante de la réalisation de la consonne ; ou bien de les considérer comme allophones respectifs de /p b m n k g r/ suivis de la semi-voyelle palatale, qui serait alors à l'origine de cette palatalisation du phonème précédant. Encore une fois, par choix d'économie, il choisit la deuxième option, tout en faisant remarquer que les oppositions "consonne palatalisée" / "non palatalisée" disparaissent devant /e/ — devant laquelle la consonne n'est jamais palatalisée — et /i/ — forçant la palatalisation dans tous les cas de figure.

### 3.4.2 Consonnes variant selon le contexte vocalique

Lorsque les consonnes / s z t d h / sont palatalisées, leur réalisation phonétique est modifiée. Certaines d'entre elles se transforment aussi lorsqu'elles sont suivies de /u/.

/s/ est une fricative alvéolaire sourde devant a, e, o et u, il ne diffère pas du /s/ français. Lorsqu'il est suivi de /i/ ou /y/, il devient alvéo-palatal [ɕ]. Cet allophone est souvent transcrit [ʃ] — c'est notamment la transcription qu'adopte KINDAICHI — mais la langue est un peu plus en avant, et c'est l'avant de son dos (ou pré-dos) qui se rapproche du palais dur, plutôt que la pointe de la langue (ou l'apex). On est donc plus proche du <x> du mandarin que du <ch> français.

/z/ est généralement considéré comme la variante voisée du /s/, bien que de nombreux phonologues japonais, comme HATTORI (1960), la décrivent comme une affriquée sifflante voisée [dʒ]. Le caractère affriqué de la consonne n'est pas évident, KINDAICHI admet qu'il y a une légère occlusion dans certaines réalisations, notamment en attaque, mais qu'elle n'est en aucun cas l'équivalent voisé du [ʒ] dont il sera question plus bas. Comme pour /s/, lorsque /z/ est palatalisé, le lieu d'articulation recule vers le palais et la consonne devient alvéo-palatale [ʒ]. Les phonologues la considèrent souvent comme affriquée, mais là encore, l'occlusion est faible. LABRUNE (2006) la transcrit [dʒ] mais VANCE (2008) la note [jʒ], avec un point d'occlusion plus

en arrière vers le palais que [d] qui se situe au niveau des alvéoles. BLOCH (1950, cité par KINDAICHI et MAËS 1978) pose deux phonèmes indépendants /ʒ/ pour [z] et /j/ pour sa variante affriquée.

Les occlusives /t/ et /d/ sont aussi intéressantes. Leur réalisation devant /a e o/ est apico-alvéolaire, c'est la pointe de la langue qui bloque le flux d'air au niveau des alvéoles puis se relâche, avec ou sans vibration des cordes vocales. Lorsqu'elles sont suivies de /u/, elles deviennent affriquées sifflantes [tʃ] [dʒ], et lorsqu'elles sont palatalisées elles deviennent affriquées chuintantes [tɕ] [dʑ]. KINDAICHI soutient qu'il n'y a aucune opposition entre /du/ et /zu/ ni /di/ et /zi/ en japonais moderne<sup>11</sup>, bien qu'il y ait un *kana* pour chacune de ces mores (respectivement つ, ず, ぢ et じ en *hiragana*). Là encore, VANCE (2008) propose de transcrire phonétiquement le /t/ palatalisé avec une occlusion palatale [tɕ]. BLOCH (1950, cité par KINDAICHI et MAËS 1978) propose deux phonèmes indépendants /č/ (tɕ) et /c/ (ɕ) pour les affriquées. HATTORI (1960) considère aussi l'affriquée sifflante comme indépendante – il la note aussi /c/ – mais pas l'affriquée chuintante.

La laryngale /h/ quant à elle, est réalisée devant /e a o/ en expirant énergiquement de l'air à travers la glotte ouverte, sans vibration des cordes, elle correspond donc à la glottale [h]. Devant /u/, la légère protrusion des lèvres entraîne friction bilabiale [ɸ]. Lorsqu'elle est palatalisée, la friction est palatale [ç]. BLOCH (1950, cité par KINDAICHI et MAËS 1978) considère cette dernière comme phonème à part entière, qu'il note /x/.

Enfin, la semi-voyelle /w/ n'est présente phonologiquement que devant /a/ en japonais moderne. Le *kana* pour /wo/ n'est plus utilisé qu'en particule pour signaler un objet direct, et il est prononcé [o]. Les *kana* pour /wi/ et /we/ ne sont plus utilisés, s'ils sont lus, ils sont prononcés respectivement [i] et [e], il n'y a donc aucune opposition sur le plan phonologique avec /i/ et /e/. /w/ est légèrement moins arrondi que le [w] français de « oiseau ». D'après LABRUNE (2006) il s'agit en fait d'une semi-vocalisation de [u] que l'on pourrait transcrire [u̯].

<sup>11</sup>Contrairement au japonais parlé à l'époque du Shogunat Tokugawa (xviii<sup>e</sup> siècle) où on distinguait /du/ [dʒü] et /zu/ [zʒü] & /di/ [dʒi] et /zi/ [zʒi] (KINDAICHI et MAËS 1978, p. 23).



### 3.4.3 Les phonèmes-mores

Les phonèmes-mores sont des phonèmes pesant à eux seuls une more. Il en existe deux en japonais : /N/ (la nasale-more (LABRUNE 2006) ou le *haneru oto* (KINDAICHI et MAËS 1978)) et /Q/ (la première partie d'une gémiation ou le *tumaru oto* (LABRUNE 2006)). Si l'on demande à un japonophone d'épeler un mot japonais, il comptera donc ces derniers comme deux unités minimales, à l'instar de /ka/ ou /zyo/ par exemple.

La nasale-more, ou ce que les phonologues japonais appellent communément le *haneru oto* 撥ねる音 (litt. « son qui saute ») est une consonne nasale indépendante, dont les réalisations phonétiques sont multiples suivant le contexte phonémique. Elle peut intervenir entre une voyelle et une consonne, entre deux voyelles ou en finale, mais jamais à l'initiale. On fait généralement appel à ce phonème par un /N/, à ne pas confondre avec /n/ qui est toujours suivi d'une voyelle. LABRUNE décrit une réalisation proche de [N] en diction lente ou avant une pause, [m] devant les bilabiales / p b m /, [n] devant les alvéolaires / t d n / et [ŋ] devant les vélares / k g /. Un débat important a lieu quant à sa réalisation devant les autres consonnes / h s z r w y / et à l'intervocalique, certains soutiennent que /N/ est alors une nasalisation de la voyelle précédente (SAITŌ, 1997, 2003, cité par LABRUNE (2006)); d'autres considèrent sa réalisation comme une voyelle haute nasale [ũ] ou [ĩ] (IMADA, 1981; HASHIMOTO, 1950, cités par LABRUNE (2006)). KINDAICHI quant à lui décrit une nasalisation seulement dans les cas de l'intervocalique, devant le /s/ et son allophone [ç] et devant les semi-voyelles :

<i>bilabiale, alvéolaire ou vélaire</i>	<i>voyelle nasale</i>
/seNpai/ [sempaj] « aîné »	/seNsou/ [seũso:] « guerre »
/seNbei/ [sembe:] « gâteau sec »	/daNsei/ [daãse:] « homme »
/deNtou/ [dento:] « tradition »	/deNsya/ [deĩça] « train »
/aNnai/ [annaj] « guider »	/saNeN/ [saẽeŋ] « trois yens »
/aNko/ [aŋko] « pâte de haricots »	/daNu/ [daũu] « pluie de balles »
	/hoNyomu/ [hõjõmu] « lire des livres »

TAB. 3.2 : Quelques réalisations de /N/ à l'intervocalique inspiré de HATTORI (1960)

La consonne /Q/ correspond à la première partie d'une géminée. Les phonologues japonais l'appellent *tumaru oto* 積まる音 (litt. « son compacté, entassé ») car elle est le résultat de la chute d'une voyelle intermédiaire, amenant deux consonnes à se suivre directement. Elle est toujours réalisée par une occlusion ou une constriction glottale [ʔ]. À proprement parler, elle n'est pas la première partie d'une réelle géminée, car en

diction lente, lorsqu'on épelle un mot contenant cette consonne — p. ex. /aQta/ « il y a eu » — on prononcera a-Q-ta, où Q est une resserrement ou une fermeture de la glotte entraînant un silence pendant lequel la langue se place au lieu d'articulation de la consonne suivante ; et on ne prononcera pas at-ta comme dans le cas d'une réelle géminée. D'un point de vue distributionnel, on trouve généralement /Q/ entre une voyelle et les consonnes /p b t k g s/ (1-3), et plus rarement entre /N/ et une consonne (4), voire même en finale (5)<sup>12</sup>

1. /gaQkou/ 学校 [gakko:] « école »
2. /maQtya/ 抹茶 [mattea] « thé vert matcha »
3. /gaQsyōu/ 合唱 [gaççō:] « chœur »
4. /nanQte/ なんつて [nan̄tte] ou [nans̄<sub>ɸ</sub>te] « c'est ce qu'il a dit »
5. /aQ/ あっ[aʔ] « ah ! »

### 3.4.4 La consonne /v/ & les nouvelles combinaisons phonémiques

Les premiers emprunts aux langues occidentales (les premiers *gairaigo*) se sont intégrés à la langue japonaise en s'adaptant presque totalement à son système phonologique. Cependant, après la deuxième guerre mondiale, certains nouveaux emprunts vont être adoptés avec de nouvelles combinaisons, qui vont peu à peu de généraliser, allant jusqu'à s'écrire avec une graphie spécifique nouvelle. Selon LABRUNE (2006) les combinaisons [ti] ou [tu] ont existé jusqu'au XVe siècle, puis on disparaît en japonais moderne. Plus tard, les emprunts aux langues européennes présentant ces combinaisons ont été assimilés respectivement [tɕi] ou [tɕü] et sont encore utilisées sous cette forme aujourd'hui : /tiimu/ [tɕi:mu] (team) ; /tuua/ [tɕüa:] (tour). Pourtant de nombreux emprunts récents arrivent sous leur forme « plus occidentale », certainement par effet de mode et de modernité, ainsi on trouve maintenant les mots /mirukutii/ [mirukuti:] (milk-tea) ; /tiisyatu/ [ti:ç̄ats̄ü] (T-shirt) ; /tuuruuzu/ [tu:ru:z̄ü] (Toulouse). Notons que la réalisation [tu] reste rare, et qu'elle est souvent légèrement affriquée.

Pour LABRUNE (2006), les consonnes /v/ /f/ et /ɸ/ sont devenus, ou sont en train de devenir de nouveaux phonèmes à part entière de la langue japonaise. /f/ et /ɸ/

<sup>12</sup>Mais peut-on réellement l'identifier à /Q/ dans ce cas où aucune consonne ne suit ? Quoiqu'il en soit, d'un point de vue graphique, le même caractère est utilisé : あっ/aQ/.

étaient jusque-là de simples allophones de /h/ et /t/ lorsqu'ils sont suivis de /u/, mais ces dernières années, on entend de plus en plus de réalisations avec les quatre autres voyelles, voire même une palatalisation pour /fyu/ dans les mots /fyuuzyon/ [ɸʲu:ɕō] (fusion) ou /fyuutyaa/ [ɸʲu:ɕja:] (future). La consonne /v/ est aussi devenu un phonème, bien qu'elle soit souvent réalisée en bilabiale fricative voisée voire en occlusive. Il ne semble pas y avoir d'opposition entre /v/ et /b/ dans le japonais actuel, et à l'oral comme à l'écrit la première est souvent remplacée par la deuxième. Les phonologues semblent s'accorder sur le statut de phonème de /v/, ce qui n'est pas encore le cas pour /f/ et /ɸ/.

On peut trouver d'autres combinaisons, plus rares cependant, LABRUNE (2006) répertorie les palatalisations avec /e/ : /ye/ [je], /sye/ [ɕe], /zye/ [ɕe], /tye/ [tɕe]; la réalisation [d] devant /i/ et /u/ : /di/ [di], /du/ [du]; les palatalisations /tyu/ [tɕu] et /dyu/ [dɕu]; et enfin les combinaisons wi, we, wo et kwa, kwi, kwe, kwo ainsi que leur variante voisée gwa, gwi, gwe et gwo. Rappelons qu'il s'agit là du système phonologique « le plus maximal » qui soit, et que la plupart de ses réalisations ne sont pratiquement présentes que dans les noms propres étrangers.

### 3.5 Rythme & prosodie

KINDAICHI et MAËS (1978, p. 9) définissent la more comme « le plus petit segment phonique susceptible, dans des conditions normales d'articulation, d'être prononcé isolément ». C'est donc l'unité prosodique minimale de la langue japonaise. Elle est souvent confondue avec la syllabe occidentale, mais elles n'ont pas exactement les mêmes caractéristiques. Comme nous l'avons vu, un allongement vocalique a le poids d'une more entière, tout comme une gémée, ou la nasale-more /N/. Ce sont les mores et non les syllabes qui sont au nombre de 5-7-5 dans les haïkus (poèmes courts japonais). Il arrive que certains phonologues japonais parlent aussi de syllabe en contexte japonais, pour interpréter le système accentuel de la langue. Celui-ci ne semble pas tout à fait correspondre aux mores, et il semblerait qu'il ait donc une seconde unité prosodique. On parle alors de *double articulation rythmique* (KINDAICHI et MAËS 1978). Notons que LABRUNE (2006) ne reconnaît pas la syllabe lourde (syllabe composée de trois constituants comme CVC ou CVV) mais seulement la syllabe légère (du type CV). D'après elle la syllabe « ne constitue pas une unité pertinente en

phonologie japonaise ».

Selon LABRUNE (2006), les structures moraiques admises en japonais sont les suivantes : V, CV, CyV, N, Q et R, où V est une voyelle, C une consonne, y une palatalisation, N correspond à la nasale-more /N/, Q est une première partie de consonne géminée, et R représente l’allongement vocalique. Les mots *katana*, *sanpo*, *kūki* ou *katto* sont donc tous trimores.

Jetons tout de même un coup d’œil aux structures syllabiques admises en japonais selon KINDAICHI et MAËS (1978). Celles-ci font certainement plus sens pour les Occidentaux que nous sommes pour se donner une idée des structures possibles du japonais (figure 3.3).

V	CV	CyV
VR	CVR	CyVR
VN	CVN	CyVN
VQ	CVQ	CyVQ

TAB. 3.3 : Listes des structures syllabiques admises en japonais d’après KINDAICHI et MAËS (1978, p. 12)

KINDAICHI ajoute plusieurs autres structures plus rares, seulement attestées dans les onomatopées ou les emprunts récents, ou encore dans le cas d’une flexion passée de certains verbes (comme /tooQta/ : « il est passé ») :

VRQ	CVRN	CyVRN
VRN	CVRQ	CyVRQ
VNQ	CVNQ	CyVNQ
VRNQ	CVRNQ	CyVRNQ

TAB. 3.4 : Structures syllabiques rares, d’après KINDAICHI et MAËS (1978, p. 12)

On constate qu’il ne peut donc y avoir de groupe consonantique de type CCV ou VCC, et toutes les syllabes sont ouvertes, à l’exception de N et Q en coda.

### 3.6 Problématique

Même si le japonais possède une structure syllabique assez simpliste, un système vocalique limité à 5 phonèmes et un système consonantique apparemment assez

réduit, on constate que l'étude de sa phonologie n'en est pas simple pour autant. L'histoire du pays a chamboulé plusieurs fois son système, l'écriture a notamment joué un rôle important dans la transformation de la langue, et aujourd'hui peu de phonologues sont en accord dans la définition de la phonologie japonaise. Il a donc été tenté ici de résumer la situation, en optant plutôt pour la description d'un système économique privilégiant les allophonies, comme celui que proposent KINDAICHI et MAËS (1978), tout en gardant un œil sur les travaux d'autres grands phonologues japonais ou étrangers.

Avec un tel système phonologique en langue maternelle, on peut légitimement penser aux difficultés que des apprenants japonophones du français peuvent rencontrer au cours de leur apprentissage. Citons par exemple le système vocalique du français avec ses 16 voyelles (au maximum), ou ses 20 consonnes et semi-consonnes, ou encore les 16 gabarits différents de structures syllabiques dont quelques monstres tels que CVCCCC ou CCCVCC<sup>13</sup>.

Afin de savoir quelles sont les difficultés réelles des apprenants japonophones, une courte étude de cas sera présentée dans le chapitre suivant. Après une analyse détaillée des difficultés des apprenants, nous pourrons réfléchir à une approche plus adaptée de la phonologie, que nous mettrons ensuite en place dans la partie expérimentale de notre étude.

---

<sup>13</sup>Comme « dextre » ou « script ».

## Chapitre 4

# Étude de cas sur la production orale d'un apprenant japonophone

Après avoir survolé les différentes caractéristiques de la phonologie japonaise, on peut émettre un certain nombre d'hypothèses sur les difficultés des japonophones à acquérir le système phonologique français :

- manque de tension dans l'articulation (les voyelles et consonnes sont globalement relâchées en japonais et tendues en français);

- difficultés à enchaîner plusieurs consonnes (aucun cluster consonantique en japonais, sauf avec /N/);

- difficultés à prononcer les combinaisons CV qui n'existent pas en japonais ou qui sont très récentes dans son système, et difficultés à les distinguer de leur équivalent japonais traditionnel (par exemple [si] et [ʃi], [zi] et [ʒi] ou [ʒi], [tu] et [ʈu] etc.);

- difficultés à réaliser les voyelles et consonnes qui ne sont pas présentes dans le système japonais (on peut penser que tous les phonèmes japonais sont légèrement différents de ceux du français, mais certains posent probablement plus de problèmes, comme les voyelles ouvertes  $\epsilon$ ,  $\text{œ}$ , le schwa  $\text{ə}$ , et  $\text{ɔ}$ ; les voyelles fermées  $y$  et  $u$ ; les consonnes  $\text{ɸ}$  ou  $\text{ɸ}$ ,  $\text{ɰ}$ , et la semi-consonne  $\text{ɥ}$ );

Mais comme le suggère BRIÈRE (*cf.* Chapitre 1), les différences interlangues ne sont pas toutes à l'origine de difficultés dans l'apprentissage, et surtout pas de difficultés du même degré. Afin de vérifier ces hypothèses, nous allons donc observer la production réelle d'une apprenante japonophone. Cette étude a pour objectif d'observer

ver les principales difficultés de prononciation d'un individu japonophone. La qualité de la prononciation variant souvent chez les apprenants selon qu'ils lisent ou parlent spontanément, le corpus audio étudié sera composé d'un extrait d'enregistrement de texte lu, et d'un enregistrement de discours spontané. Dans le deuxième cas, l'étudiante ignorait qu'elle était enregistrée, sa prononciation est donc naturelle. Après avoir présenté le profil du locuteur, le corpus utilisé et la transcription intégrale des enregistrements, seront analysées les difficultés au niveau phonétique puis au niveau syllabique en fonction de ce qui est observable dans les enregistrements.

## 4.1 Profil du locuteur

Le locuteur est de sexe féminin, elle a 24 ans au moment de l'étude. Elle est née dans la préfecture de Hyōgo au Japon, mais a vécu la plus grande partie de sa vie dans la ville de Fukuoka. Elle parle un japonais plutôt standard avec sa famille, mais utilise parfois le dialecte local de Kyūsyū avec ses amis. Après trois ans d'études du français au Japon en tant que spécialiste à l'université, elle séjourne en France depuis presque un an, suivant un Master de littérature. Elle a obtenu le diplôme de langue française DELF B2 au Japon avant de venir. Malgré son niveau vraisemblablement élevé de français, elle semble avoir encore beaucoup de difficultés de prononciation. L'étude de son cas paraît donc pertinente pour connaître les difficultés réelles rencontrées par les japonophones dans l'acquisition de la phonologie française.

## 4.2 Transcriptions des enregistrements

### 4.2.1 Texte original de l'enregistrement n°1 (lecture, 1'07")

Le texte utilisé pour l'expérience provient d'une plaque d'information à l'entrée d'un monument historique : « Commune de Gières, fort du Mûrier. Construit entre 1873 et 1878, le fort du Mûrier a été acheté à l'armée par la commune de Gières en 1978. Il a été inscrit en 1993 par le Ministère de la Culture à l'inventaire supplémentaire des monuments historiques. La municipalité de Gières veut en faire un équipement public à caractère culturel pour les habitants de l'agglomération grenobloise. Le fort est actuellement en cours de restauration grâce à des crédits de l'État, de la

région Rhône-Alpes, du département de l'Isère et de la commune de Gières ».

#### 4.2.2 Transcriptions orthographique et phonétique fine de l'enregistrement n°1

1 *Commune de Gières, fort du Mûrier*

[kɔmiyndeɪ.kɔmiyndəʒi<sup>(i)</sup>εχ.pɔχdymy<sub>u</sub>(ʁ/r)je.] ||

2 *Construit entre 1973 et 196...*

[kɔst<sub>o</sub>ɾɥi.ātχəmilnœfsā:swasātχezemilnœfsāsɔw.] ||

3 *1873 et 1878*

[milqisāsɔswasātχezemilqisāsɔswasāt̃di(ʒ/z)qit]

4 *le fort du Mûrier a été acheté à l'armée par la commune de Gières en 1978*

[lɔfɔɾdymyɾje.aeteaf<sub>s</sub>te.alaxmepa:lakɔmyndəʒiε.āmilnœfsāsɔswasāt̃di(ʒ/z)qit]

5 *il a été inscrit en 1993*

[ilaeteēs<sub>u</sub>k<sub>u</sub>ri.omilnœfsā:katrəvēt(r/r)εz]

6 *par le Ministère de la Culture*

[pəβləministεχdəlaky<sub>s</sub>t<sup>ε</sup>γχ]

7 *à l'inventaire supplément supplémentaire des monuments historiques*

[alēβātεχ(β/β)ɣplemāt.(β/β)ɣ.ple.mātεχdē:monymā(ʒj/z)istɔrik] /

8 *la municipalité de Gières veut en faire un équipement publique*

[lamyni(β/β)ipalitedə(ʒj/z)βεχβɔtāfεβœ:ekipəmāpɣblik]

9 *à caractère culturel pour les habitants de l'agglomération l'agglomération...*

[akəβakteχkyltyβεlpy:lezabitādəlag<sub>u</sub>βo:.lag<sub>u</sub>ro.m<sup>w</sup>eras] /

10 *l'agglomération grenobloise*

[lag<sub>u</sub>lɔmεβasjɔgrənobrwa:z]

11 *le fort est actuellement en cours de restauration grâce à des crédits de l'État*

[lɔfɔχ?.e:aktɔɣyεlmā.ākuxdə:(β)εstɔβasjɔgvasadekredidø:leta]

12 *de région Rhône-Alpes, du département de l'Isère et de la commune de Gières*

[dəχεʒjɔχɔnalp.dy:depaχtəmādə:lizeχ.edəlakɔmiyndə:ʒjijεχ]

#### 4.2.3 transcription fine de l'enregistrement n°2 (spontané, 40")

13 *la question chinoise c'était*

[lak<sup>w</sup>εstɔj(β/β)inwa:z<sub>u</sub>sete:]

14 *comment on dit en chinois : « il arrive »*



[kom(̃/ā).̃diʔā̃j̃inwaʔilari:v]

15 *et un étudiant chinois m'a rép/ a répondu heu « nihao »*

[e:õnetɤdjā̃j̃inwamarep.arepõd̃y.h̃.̃ni:ha:w] ||

16 *et il a imité (rire) et et sono et et avant il a il a expliqué que s/*

[e:ila.imate (rire) e:e:sõnõ.e:eavāila.ilaek,splikekø:s:]

17 *si on si il arrive en italien*

[si.̃ʔsi.ilarivʔāitak̃j̃ē]

18 *il ça peut être Robert, un garçon, un mec euh hum*

[il.sā:pøet̃k̃:ʔobe:χ̃.õgax̃s̃.õm̃e:k̃:əm̃]

**Légende :** / : pause ; || : coupure

j̃ : palatalisation

õ : brève

ɥ̃ : prononciation forcée

w : labialisation

ü : centralisation

r : battue alvéolaire

(k̃/r) : distinction difficile

ʃ̃ : dévoisement

ç/z : fricative alvéo-palatale

(k̃) : peu perceptible

ʒ̃ : postériorisation

β : fricative bilabiale voisée

t̃õ : épenthèse

ʧ̃ : antériorisation

ʔ : coup de glotte

t̃ : consonne non relâchée

ā : fermeture

u : voy. fermée post. non-arr.

. : ici utilisé pour marquer une pause entre deux syllabes

### 4.3 Difficultés au niveau segmental

#### 4.3.1 Les consonnes

**Les fricatives uvulaires [k̃] & [χ] et l'approximante latérale alvéolaire [l]** On peut constater que la fricative uvulaire et l'approximante latérale alvéolaire sont source de difficulté pour la locutrice. En coda, l'uvulaire est généralement bien prononcée, plutôt sourde, mais il arrive parfois qu'elle soit inaudible (« par la commune » [pa:lakõ-myn] 1.4 ; « pour les habitants » [py:lezabitā] 1.9). Lorsqu'elle est réalisée en attaque cependant, elle est souvent remplacée par une battue alvéolaire voisée (r), qu'il s'agisse d'une attaque simple (« historique » [istõrik] 1.7, « l'aggloméras... » [lag̃ỹ.ro.m̃<sup>w</sup>eras] 1.9 ou « il arrive » [ilari:v] 1.14) ou d'une attaque complexe (« construit » [k̃õst̃õrqi] 1.2, « quatre-vingt-treize » [katrõṽēt(c/r)ɛz] 1.5, « grenobloise » [gr̃õnõbrwa:z] 1.10). On trouve globalement peu d'uvulaire voisée, il s'agit soit de χ soit de r. On peut penser que ce problème est dû au fait qu'il existe un son relativement proche en japonais : la fricative glottale sourde [h], mais il n'existe pas de variante sonore [h̃] ; et l'uvulaire voisée est plutôt associée à la battue alvéolaire du japonais. D'un point de vue

phonologique, notre étudiante est en cours d'acquisition de l'uvulaire, puisqu'elle le prononce parfois correctement, mais lorsque l'environnement est trop complexe ou qu'elle relâche son attention, la consonne est substituée par la battue alvéolaire ou devient inaudible.

À la ligne 1 du premier enregistrement, l'étudiante prononce un léger /r/ qui ne devrait pas être prononcé, rhotacisant la voyelle finale ([my(ʁ/r)je]). Cela vient certainement du fait que l'étudiante lisait le mot « Mûrier » qu'elle ne connaissait probablement pas, elle en a donc prononcé toutes les lettres.

Enfin, la prononciation du /l/ ne semble pas constituer de difficulté particulière, mais lorsqu'elle se trouve dans une attaque complexe, elle aussi est parfois remplacée par une battue alvéolaire (exemples précédents lignes 9 et 10). On observe aussi à plusieurs reprises que la locutrice prononce clairement une uvulaire au lieu d'une latérale alvéolaire ([lag̃̃ɔ:] l.9 pour « l'agglomération » ou [itaʁjɛ̃] l.17 pour « italien »). S'agit-il d'une difficulté de distinction des deux phonèmes ?

**Le cas des coronales [s]-[ʃ] & [z]-[ʒ]** En ce qui concerne les fricatives alvéolaires, on constate qu'elles sont constamment chuintantes, réalisées en palato-alvéolaires, devant /i/, /y/ et /ɥ/ (« dix-huit » [di(ʒ/z)ɥit] ll.3-4, « supplémentaire » [(ʃ/ç)ɥ.ple.māteχ] ou « monuments historiques » [monjymā(ʒ/z)istɔrik] l.7). Elles sont pourtant correctement réalisées dans les autres environnements (« soixante-treize » [swasātχɛz] l.2, « restauration » [(ʁ)ɛstɔʁasjɔ̃] l.11, « ça peut être » [sɑ:pœtɛʁə:] l.18). On retrouve ici la modification consonantique présente en japonais, avec les mêmes consonnes devant la voyelle /i/. Le phénomène pourrait aussi être une conséquence du dialecte de Kyūsyū, dans lequel les syllabes /sa/ /si/ /su/ /se/ /so/ sont souvent prononcées [ɕa ɕi ɕü ɕɛ ɕɔ] et leur équivalent voisé [ḑa ḑi ḑü ḑɛ ḑɔ] avec un [d] très léger en début de syllabe (au lieu de [sa ɕi sü se sɔ] et [ḑa ḑi ḑü ḑɛ ḑɔ] en japonais standard).

**La fricative labio-dentale [v]** Voici une autre consonne qui semble poser problème à notre étudiante : la fricative labio-dentale [v]. Elle est prononcée proche du [b] (bilabiale et non labio-dentale) mais est plus fricative qu'occlusive (elle semble pouvoir être tenue, il n'y a pas d'occlusion complète des lèvres). Je la transcrirai [β] : [lɛβātɛχ] l.7 pour « l'inventaire » et [βɔtäfɛβ] l.8 pour « vont en faire ». La fricative labio-dentale

n'existe pas en japonais traditionnel, elle a été importée depuis les langues occidentales au cours des dernières décennies, mais n'est pas encore totalement assimilée (cf. chapitre III). L'étudiante japonaise ici enregistrée semble être là encore en cours d'acquisition de la prononciation correcte, elle est à mi-chemin entre le [b] et le [v].

### 4.3.2 Les voyelles

**Centralisation de [u] & [y] vers [ü]** Devant les 16 voyelles du français, le japonophone et ses 5 voyelles peine à distinguer certains sons. Il y en a deux notamment difficiles, le [y] et le [u]. La voyelle la plus tendue du japonais est un [u] centré et peu arrondi, et à l'acuité plus élevée (un peu plus aiguë). À l'université de Kobé au Japon, on le note en général [ü] soit [ʊ]. Ici je garderai la première transcription. Notre étudiante semble bien maîtriser les deux voyelles et ne les inverse pas non plus. Pourtant, on sent légèrement qu'elle peine plus à prononcer le [y], qui est très en avant et la force à labialiser au maximum. Le japonais est une langue aux voyelles très relâchées, il n'y a pas d'important étirement des lèvres ni d'importante labialisation. Notre étudiante va alors souvent avoir recours à la palatalisation pour anticiper la labialisation du [y].

Dès le premier mot de l'extrait du premier enregistrement (« commune » [kɔ̃mjɛ̃] l.1), on assiste à une palatalisation du /m/ qui anticipe la voyelle arrondie qui suit. Le même phénomène a lieu ligne 4 [d̟j] où le /d/ est aussi légèrement palatalisé, ou encore à la ligne 7 dans « supplémentaire » [(j̟/ɕ)j̟.ple.m̟teɕ]. Pourquoi cette palatalisation ? On peut constater que lorsque la palatalisation n'a pas lieu dans des syllabes similaires, l'étudiante semble faire plus d'effort et la voyelle est souvent fortement articulée (« Mûrier » [my̟ɕje] l.4), le résultat est aussi peu naturel car l'attention est concentrée sur la première syllabe, et la deuxième est alors délaissée et prononcée à la japonaise [ɕje] au lieu de [ɸje]. Une seconde hypothèse explique également cette palatalisation devant [y], il s'agit du fait que l'on peut parfois trouver un son proche de [y] en japonais lorsque la consonne précédant /u/ est palatalisée : CyV (LABRUNE 2006). À la syllabe /mu/ ([mu]) correspond la syllabe /myu/ ([mj̟]), [j̟] étant plus proche du [y] français, la solution est donc de prononcer la variante palatalisée CyV pour se rapprocher de la prononciation française de la voyelle suivante.

## 4.4 Difficultés au niveau syllabique

### 4.4.1 Épenthèses et problèmes de structuration syllabique

Comme nous l'avons brièvement vu dans le chapitre précédent, le japonais possède une structure syllabique (ou moraique) relativement peu complexe. On dit qu'il en a hérité des langues austronésiennes (comme les langues philippines) tout comme son système phonologique archaïque. En effet, une grande partie des langues de cette famille possèdent une structure syllabique extrêmement simple, du type CV. Le japonais aussi est exclusivement composé de syllabes CV mis à part le /n/ final qui peut venir fermer une syllabe ou les consonnes géminées.

Comment l'apprenant japonophone peut-il alors prononcer des mots comme « spectre » (CCVCCC) ? Il va souvent se voir intercaler une voyelle brève intermédiaire, appelée épenthèse. La transcription du mot précédent devient alors スペクトル /sʌpɛkʌtɔɾɯ/ ([s<sub>ü</sub>pɛk<sub>ü</sub>tɔɾ<sub>ü</sub>]) et fait partie tel-quel du jargon scientifique japonais depuis plusieurs décennies. Lorsqu'interviennent des structures syllabiques complexes, notre étudiante japonaise, malgré son bon niveau de français, va intercaler inconsciemment une ou plusieurs voyelles d'appui entre les consonnes. Le plus bel exemple est le mot « agglomération » aux lignes 9 et 10 dans « les habitants de l'agglomération grenobloise... l'agglomération grenobloise » [lezabitädəlag<sub>ü</sub>ɔ:ləg<sub>ü</sub>.rɔ.m<sup>w</sup>eras] / [lag<sub>ü</sub>lɔmɛɔsɔjɔgrənobɾwa:z]. Lorsque la structure syllabique est complexe, l'étudiante fait souvent moins attention à la prononciation des phonèmes : « soixante-treize » [swasātɕɛz] mais « construit » [kōst<sub>ɔ</sub>rɕi] à la ligne 2. Elle sait prononcer la combinaison [tɕ] mais l'ajout de la semi-consonne [ɕ] la perturbe et elle reprend le /r/ japonais... Il peut s'agir aussi du fait que la combinaison [tɕɕi] nécessite que l'on passe d'une occlusive alvéolaire sourde (antérieure) à une fricative uvulaire plus ou moins sourde (postérieure) et directement à une approximante labio-palatale voisée suivie d'une voyelle antérieure. De quoi faire un nœud avec sa langue...

On constatera que la voyelle centrale [ü] est souvent utilisée comme voyelle d'appui (« inscrit » [ɛsk<sub>ü</sub>ri] l.5 ; « chinoise c'était » [(̃/ɕ)inwa:z<sub>ü</sub>sete:] l.13), mais on trouve aussi [o] et [ɔ] (l.4, l.16).

#### 4.4.2 Intrusions japonaises & anglaises

**Hésitations japonaises** C'est probablement le cas pour chacun d'entre nous d'hésiter dans sa langue maternelle. Dans les enregistrements étudiés ici, on peut entendre de nombreux [e:], il s'agit simplement de l'équivalent de notre « heu... ». Il ne constitue pas, selon moi, une erreur à corriger, mais il n'est pas impossible de comprendre un « et... », et selon le contexte, il pourrait y avoir un malentendu. Il est donc bon de savoir hésiter dans la langue de son interlocuteur. Dans l'enregistrement du discours spontané, il y a bien plus d'hésitations. On peut trouver une autre intervention d'un mot japonais [sono] l.16 (litt. « cela ») qui est souvent utilisé lorsqu'on réfléchit ou qu'on ne trouve pas ses mots. À part peut-être dans un contexte musical, il semble difficile d'arriver à un malentendu avec ce mot-ci, il est donc encore moins problématique que le précédent.

**Un anglais isolé** À la ligne 13, dans l'enregistrement du discours spontané, on entend le mot « question » prononcé à l'anglaise [k<sup>w</sup>estejɔ̃] avec une fin vocalique plutôt française. Cette prononciation peut avoir plusieurs origines, soit de lecture à la vue de « que » (mais l'erreur serait alors sans doute moins isolée), soit par une confusion avec la prononciation anglaise du même mot (la graphie étant identique). Quoiqu'il en soit, il ne s'agit probablement pas d'une semi-voyelle d'appui comme la palatale [j] puisque la syllabe [kɛ] existe en japonais et ne présente donc aucune difficulté de prononciation pour l'étudiante.

### 4.5 Synthèse de l'analyse

Une analyse minutieuse des enregistrements de cette étudiante japonaise aura donc permis de faire ressortir les différentes difficultés, tantôt au niveau segmental, tantôt au niveau syllabique de sa production orale. Plusieurs réalisations non-naturelles chez un natif restent occasionnelles : parfois l'étudiante prononce correctement, parfois, lorsque l'environnement phonétique ou la structure syllabique sont trop complexes, elle relâche son attention sur certains phonèmes et la prononciation du mot en est affectée. L'acquisition de ces phonèmes est partielle, et leur réalisation dépend du contexte. Ces observations correspondent bien au concept de continuum

dans l'apprentissage de la phonologie (*cf.* FLEGE (1980), Chapitre 1).

Une analyse de la prosodie de l'étudiante a aussi été réalisée, mais celle-ci ne semblait pas présenter de problème particulier (pour le discours spontané), il ne semble donc pas pertinent de travailler cet aspect avec ce type de locuteur.

Cette analyse met en évidence que, mis à part quelques rares phonèmes plus problématiques isolément (comme les voyelles /u/ et /y/ ou la fricative labio-dentale), les principales difficultés ressortent dans les combinaisons de phonèmes. On retiendra certaines associations CV (notamment avec l'uvulaire en attaque et les coronales devant /i/ et /y/) et la plupart des combinaisons pluriconsonantiques (attaques et codas complexes).

## 4.6 Réflexions générales

Nous avons pu voir jusqu'à maintenant que l'apprentissage de la phonologie d'une langue étrangère est une tâche complexe qui peut être influencée par la phonologie de notre langue maternelle et par les systèmes d'écriture de l'une ou l'autre langue. Il a été constaté que l'apprentissage de la phonologie (et l'apprentissage en général) s'appuyait de manière importante sur l'écrit, et que les apprenants se retrouvent souvent à apprendre et parler une langue à travers sa représentation graphique. Le Silent Way est une approche permettant notamment d'appréhender la phonologie sans aucun recours à l'écriture, et présentant les phonèmes de manière à solliciter les prises de conscience chez l'apprenant. Une approche de ce genre pourrait certainement faciliter l'apprentissage du français pour les locuteurs de langue maternelle japonaise. Ces-derniers sont en effet confrontés à une double difficulté : d'une part le système phonologique de leur langue maternelle est très éloigné du français et sensiblement moins complexe sur le plan syllabique, d'autre part leur système d'écriture les habitue à une représentation graphique de la langue totalement différente du système français, tout cela étant à l'origine de grandes difficultés dans l'apprentissage.

L'objectif de cette étude est donc de proposer un cours Silent Way à des apprenants japonophones, et d'en observer les effets sur leur production orale.

Un des paramètres clé du protocole qui a été mis en place pour évaluer leur production, est certainement l'importance donnée à la spontanéité et au naturel de la parole. L'observation de la production des participants s'est faite sur un discours se

voulant le plus spontané possible. Aucune « phrase porteuse » ou énoncés redondants déconnectés de la réalité n'a été utilisé. Aucune lecture de mots ou d'énoncés, ni répétition après écoute n'a eu lieu. Toute la problématique du protocole était de trouver comment *provoquer* la réalisation de certains mots cibles, sans influencer le locuteur, et dans un contexte le plus naturel possible ; tout en répétant l'expérience avec un grand nombre d'individus et permettre une analyse des données à grande échelle. Le protocole doit pouvoir analyser et comparer la production de nombreux locuteurs et ce de manière la plus objective possible.

Si l'analyse automatique permet d'analyser un grand nombre de productions, et d'éviter tout biais de perception subjectif, alors comment rendre automatique l'analyse d'un discours libre, et présentant donc tant de variation interlocuteurs ?

La décision prise pour ce mémoire a été de cibler un certain nombre de mots présentant des groupes consonantiques (ou clusters) prononcés spontanément par tous les locuteurs. Rappelons cependant que c'est bien sur les problèmes qu'il faut se concentrer, et non sur des réalisations isolées de mots cibles. Il est donc essentiel de dégager des tendances à partir de nos observations.

La principale contrainte technique dans cette étude est de vouloir automatiser au maximum l'analyse de la production des locuteurs. Mais y a-t-il réellement un intérêt à l'automatiser ? Une analyse manuelle fine est évidemment plus précise et souvent plus simple à réaliser, mais en plus d'être extrêmement chronophage, elle se limite à l'observation d'un nombre limité d'individus et donc de réalisations. Si l'on veut dégager des tendances générales de prononciation, une analyse automatique sur un large échantillon d'individus est nécessaire. Les résultats sont en effet plus réalistes car ils prennent en compte plus de cas de figure et plus de profils. L'automatisation permet aussi de s'appuyer uniquement sur la réalité physique du son, et écarte toute subjectivité du ou des observateurs. Les technologies actuelles permettent une infinité d'analyses, mais elles sont encore relativement peu exploitées. Plus le nombre d'expériences augmentera et meilleures deviendront les analyses, grâce au partage des connaissances, des outils créés et des protocoles testés.

## Deuxième partie

### Proposition d'une formation Silent Way & observation de l'évolution de la prononciation





# Chapitre 5

## Méthodologie

Il a été proposé dans cette étude un cours intensif de phonologie française à un public multi-niveau d'apprenants japonophones. L'objectif est d'observer les évolutions dans la prononciation des participants au niveau de la structuration syllabique, afin de déterminer les effets d'une telle approche sur la prononciation de ce public. Les capacités de perception ne font pas l'objet de cette présente étude, bien qu'elles soient intimement liées à la prononciation. Il a donc été déterminé un temps d'observation de la production des participants avant ( $T_1$ ), puis après la formation ( $T_2$ ), et l'intégralité de la formation a été filmée et enregistrée pour permettre une étude du comportement des participants et leur prononciation pendant les séances de cours.

En raison des contraintes humaines et temporelles, il a fallu réduire le temps de formation à deux semaines (4 séances), et 7 des 10 participants seulement ont pu assister à l'ensemble des séances. Afin d'obtenir des différences significatives dans la prononciation des participants, il faudra donc réitérer l'expérience sur une plus longue période pour laisser un temps d'automatisation des prises de conscience qui auront été faites pendant les séances de cours. Enfin, un nombre supérieur de participants permettrait d'envisager une généralisation des observations à la plupart des apprenants japonophones du français langue étrangère. Précisons que la méthodologie mise en place pour cette étude se veut être adaptée à une étude de plus grande échelle.

Ce chapitre va donc présenter en détail le protocole mis en place, et laisser l'analyse des données aux chapitres suivants.

## 5.1 Les participants

Un total de 10 locuteurs ont été réunis pour l'expérience. Ils sont tous de nationalité et de langue maternelle uniques japonaises. La tranche d'âge s'étend de 21 à 33 ans, N1, N4, N5, N6 et N8 sont étudiants échange sur l'année à l'Université Grenoble Alpes (UGA), N2 et N7 sont enseignants de japonais dans cette université, N5 et N10 sont étudiants individuels, et N3 est femme au foyer. La parité n'a pas pu être respectée dans cette étude, où seuls deux hommes pour huit femmes ont pu participer. Certains connaissent l'expérimentateur (moi-même) mais personne ne connaît encore l'enseignant en charge de la formation intensive. Tous apprennent le français, mais le nombre d'années d'apprentissage varie entre un an et demi et huit ans. Notons que le nombre d'années d'apprentissage du français ne reflète pas le niveau acquis, cependant il n'a pas été jugé nécessaire d'évaluer les compétences globales des locuteurs, étant donné que seule la prononciation fait l'objet de notre étude. Le nombre d'années, les lieux et les conditions d'apprentissage seront prises en compte dans l'interprétation des résultats. L'illustration 5.1 présente les lieux où les locuteurs ont passé la plus grande partie de leur vie, il ne s'agit pas forcément de leur lieu de naissance<sup>1</sup>. L'illustration 5.2 présente le nombre d'années considérées par le locuteur comme années d'apprentissage du français, quelle que soit l'institution ou le moyen d'apprentissage.

L'expérience s'étant déroulée à Grenoble, les locuteurs ont en commun le fait d'y habiter, et donc de se trouver dans un environnement francophone. Les durées totales de séjour en France vont de 7 mois (pour N1, N4, N6 et N8) à 5 ans et 9 mois (pour N5), avec une moyenne de 1 an et 8 mois. En ce qui concerne les institutions dans lesquelles le français a été appris, on trouve l'université japonaise (N1, N4, N6-10), le lycée japonais pour N4 et N5, l'Alliance Française au Japon pour N3 et des institutions en France comme le Cuef de Grenoble ou l'université Lyon III, pour l'ensemble des dix locuteurs.

Enfin tous les locuteurs ont appris l'anglais, et parfois d'autres langues sur de plus courtes périodes. Tous ne parlent pas la même variété de japonais en famille, N1, N4, N5, N8 et N10 disent utiliser un japonais plutôt standard, mais parfois aux côtés d'un dialecte, les autres ne signalent qu'un dialecte. Comme l'ont remarqué entre autres

---

<sup>1</sup>Source de la carte du Japon : <http://www.statistiques-mondiales.com/>



FIG. 5.1 : Lieux de vie principaux des locuteurs

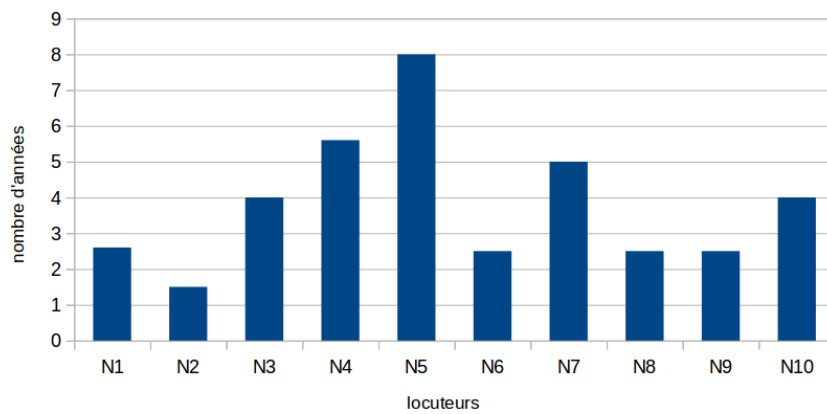


FIG. 5.2 : Nombre d'années d'apprentissage du français par locuteur

(RACINE, DETEY et KAWAGUCHI 2012), il est fort probable que le dialecte maternel influence le locuteur japonophone dans sa production en langue étrangère. Ce paramètre est donc à prendre en compte. Un tableau détaillé des profils par locuteur est disponible en annexe.

## 5.2 Le protocole

Afin d'évaluer les effets d'une approche pédagogique sur la prononciation des locuteurs, le protocole expérimental a été divisé en 3 phases :

- Observation de la prononciation pré-formation en entretien individuel (T<sub>1</sub>)
- Observation du comportement et de la prononciation des participants pendant la formation intensive
- Observation de la prononciation post-formation en entretien individuel (T<sub>2</sub>)

Ajoutons qu'il serait pertinent d'observer à nouveau la prononciation des participants plusieurs mois après la formation (dans une 4<sup>ème</sup> phase d'observation post-post-formation), afin d'évaluer l'influence à long terme de l'approche. Pour des raisons de temps, cette 4<sup>ème</sup> phase ne sera pas mise en place dans cette étude.

Les entretiens individuels ont été enregistrés au moyen d'un Zoom H1 à condensateurs unidirectionnels, au format WAVE 16 Bit, 1536 Kbps, 48.0 KHz et en stéréo. Les quatre séances de formation ont été enregistrées de la même façon pour le son, et deux caméras ont été placées dans la salle (caméra arrière : plan sur l'enseignant et le tableau, SONY SR90, format vidéo MPEG 720\*576 (16 :9), 8961 Kbps et 25000 fps ; caméra avant : plan sur la classe, SONY SR11, même format et résolution, 8646 Kbps et toujours 25000 fps).

Notre objectif principal est d'observer la réalisation des phonèmes dans un environnement syllabique pluriconsonantique, *quasi-inexistant* en japonais<sup>2</sup>. Un certain nombre de clusters (séquences phonémiques pluriconsonantiques) ont été ciblés en fonction de leur éloignement avec le japonais (par exemple /ks/ a été évité, car il ne

---

<sup>2</sup>Si l'on omet les structures /NC/ et /QC/, et le fait que certains cas de dévocalisation interconsonantique amènent à une disparition physique de la voyelle (p.ex. /gakusei/ réalisé [gakuse:] ou [gakse:], cf. Chapitre 11, section 3).

présente vraisemblablement pas de difficulté (*cf.* note de bas de page précédente), mais aussi en fonction de leur fréquence estimée d'apparition en français (la séquence /kb/ de *snack-bar* par exemple ne semble pas très représentative des clusters français). De plus, les clusters cibles doivent pouvoir apparaître dans des mots du même champs lexical, pour permettre d'orienter la discussion sur celui-ci, et ainsi favoriser la production des mots contenant ces clusters. Le tableau 5.1 présente la liste des clusters (et mots ciblés) qui feront l'objet de nos observations. Les mots *classe* et *test* ont été choisis car ce sont des emprunts courants utilisés en japonais, malgré la difficulté qu'ils peuvent représenter pour les faire prononcer par les locuteurs (leur équivalent japonais sont [kūrasu] et [tesūto]).

<i>attaque</i>	<i>cluster cible</i>		<i>mot cible</i>
	<i>intervocalique</i>	<i>finale</i>	
/fʁ/			France
/gʁ/		/bl/	Grenoble
	/ʁs/		université
/st/			Stendhal
		/lp/	Alpes
/kl/			classe
		/st/	test
/tʁw/			trois
		/tʁ/	quatre
/tʁ/			treize
	/zɥ/	/ʁz/	quatorze
	/zn/		dix-huit
	/tk/		dix-neuf
/tʁ/	/tɥ/	/tʁ/	trente-quatre
	/tɥ/		quarante-huit
/sw/	/ts/		soixante-sept
/sw/	/td/ ; /zɥ/		soixante-dix-huit

TAB. 5.1 : Table des clusters ciblés

Il a initialement été prévu d'observer aussi la réalisation des consonnes seules en finale de mot, car elles sont aussi très rares en japonais<sup>3</sup>, cependant suite aux difficultés rencontrées pendant le traitement des données, l'analyse s'est finalement limitée aux clusters.

<sup>3</sup>Il n'existe que la nasale-more /N/ pouvant se retrouver en coda ; et la voyelle /u/ peut être dévocalisée voire totalement amuïe après certaines consonnes comme /s/ ou /Qk/.

Avant de continuer, gardons à l'esprit que nous travaillons ici sur des mots isolés. Dans le flux de la parole spontanée il arrive souvent que la finale d'un mot soit combinée avec l'attaque d'un autre, modifiant alors l'environnement phonémique de la consonne ou du cluster étudié, et rendant plus difficile la comparaison des données. De plus, certaines réalisations sont simplifiées en fonction de l'environnement : par exemple l'énoncé « quatre personnes » est généralement réalisé [katpɛχsɔn] et non [katχpɛχsɔn] par des francophones natifs. La rigueur scientifique nous inciterait ici à observer la réalisation des mêmes mots cibles par des locuteurs francophones natifs, dans les mêmes conditions, afin de comparer les réalisations spontanées et les stratégies de simplification phonologique en fonction de la langue maternelle. Plusieurs propositions sont faites en ouverture, mais la présente étude se centrera sur le traitement des données des locuteurs japonophones, avec pour référence la base de données du programme EasyAlign, dont l'utilisation est détaillée dans la section suivante.

Nous avons donc un certain nombre de mots cibles, et il faut à présent trouver le moyen de les faire dire aux locuteurs en évitant toute influence sur leur production (nous éviterons donc la répétition, la lecture, la prononciation de mots isolés etc.). Il est possible d'imposer un thème de discussion pour inciter l'utilisation de certains mots ; mais afin de comparer quantitativement les données, il est nécessaire d'obtenir au moins une version de chaque mot par tous les locuteurs. La décision prise pour cette étude a été de demander à chaque locuteur de parler d'eux-mêmes en intégrant obligatoirement 7 mots (Grenoble, Alpes, France, université, Stendhal, classe et test). Ces mots sont préalablement montrés au locuteur, écrits sur une feuille. Comme il s'agit de termes jugés courants pour notre public universitaire, et relatifs à leur environnement direct, on peut penser que leur lecture influence peu la prononciation, à la différence de mots moins familiers voire inconnus de certains locuteurs. La feuille sera retirée après quelques minutes afin de s'assurer que les mots ne sont pas lus pendant leur prononciation. La présentation du locuteur n'est pas limitée dans le temps, il faut seulement s'assurer que tous les mots sont utilisés au moins une fois. En cas d'oubli, il sera toujours possible de poser une question pour rediriger la conversation, ou éventuellement montrer à nouveau la fiche.

En fin d'entretien, il est proposé au locuteur de compter de 1 à 20, puis lire quatre nombres écrits en chiffres arabes sur une autre feuille de papier, et enfin compter de 20 à 1. Les chiffres sont généralement appris ensemble, dans l'ordre, et leur automa-

tisation rend probablement leur prononciation assez « naturelle » (pour le locuteur) même dans le contexte de cet entretien. Mais il est possible que cette automatisa-tion soit aussi la cause de la fossilisation d'une prononciation moins correcte, datant de l'époque où ces nombres ont été appris. Le locuteur est donc amené ensuite à recompter dans le sens inverse, ce qui nécessite une plus grande présence d'esprit (charge cognitive supérieure) car cet ordre est moins automatique. Nous pourrions alors comparer les réalisations et voir si une différence significative est audible, ou observable physiquement. Entre les deux énumérations, le locuteur sera amené à lire les nombres 34, 48, 67 et 78. Il s'agit ici d'une lecture, mais les chiffres arabes ne devraient pas influencer la prononciation, puisqu'ils ne donnent aucun indice phono-logique. Ces nombres ont été choisis pour les clusters intervocaliques /tk tɥ ts td/. Rappelons que l'omission du premier t dans « soixante-dix-huit » est courante chez les francophones natifs, et ne constitue donc pas une particularité due à un accent étranger. En revanche le deuxième t de « trente-quatre » est rarement omis chez les natifs.

Notre entretien avec les locuteurs est donc composé d'une présentation de celui-ci (monologue) et d'une séquence d'énumérations. Afin de « chauffer » un peu le locuteur, et détendre l'atmosphère en début d'expérience, il sera proposé dans les premières minutes une discussion la plus détendue possible sur le profil du locuteur. Plutôt que lui faire remplir un questionnaire papier en japonais, les questions seront posées à l'oral en français, et le locuteur sera libre de parler ou faire répéter la question en japonais s'il ne comprend pas. La discussion est amicale et permet de briser la glace de la première rencontre ou au moins commencer doucement par une prise de contact. Toutes les informations collectées permettront en outre de constituer une fiche d'identité du locuteur, qui permettent de caractériser plus précisément la population. Les entretiens seront intégralement transcrits et tous les mots présentant des clusters et prononcés par l'ensemble des locuteurs seront ajoutés à la liste des mots cibles pour les analyses. Ces mots, s'ils ne sont pas issus d'une répétition d'un mot prononcé par l'examineur, seront jugés « naturels » pour le locuteur, puisque totalement spontanés.

Une fois que tous les locuteurs ont passé l'entretien, la formation peut commen-cer. Par contrainte de temps et de moyens, la présente étude n'a proposé que 4 séances d'une heure, réparties sur deux semaines. Les séances ont lieu les mardis et jeudis soir.



Comme le nombre d'heures de formation est limité, tous les locuteurs doivent être présents à l'ensemble des séances, car même un seul quart de la formation en moins pourrait grandement influencer les observations de la phase 3. Malgré tout, trois locuteurs ne seront pas en mesure d'assister à l'ensemble des séances, et leurs données ne seront donc pas analysées. Pendant la formation, plusieurs champs lexicaux seront travaillés, mais il est important que les mots cibles ne soient pas – ou le moins possible – étudiés en cours, car les évolutions seraient évidentes dans les observations de la phase 3, et probablement pas généralisables pour tous les clusters du même type<sup>4</sup>. L'important est bien de corriger un problème et non la prononciation d'un mot isolé. Si la prononciation du mot *examen* est travaillée en classe, le mot sera peut-être mieux prononcé par la suite, mais cela ne veut pas dire que la prononciation du mot *exact* le sera aussi. L'objectif est bien-sûr qu'elle le soit, mais selon la façon dont *examen* est travaillé, le transfert sera plus ou moins bien effectué vers les autres énoncés présentant le cluster /gz/. Si ce cluster est mieux réalisé phonétiquement en T<sub>2</sub> qu'en T<sub>1</sub>, et qu'il n'a pas été directement travaillé pendant la formation intermédiaire, on peut penser que la prononciation est acquise et correctement transférée (étapes 3 puis 4 d'apprentissage selon Gattegno).

Après la formation, tous les locuteurs présents à l'ensemble des séances sont à nouveau conviés à un entretien individuel, dont l'objectif officiel est de discuter de leurs ressentis pendant la formation et me conseiller, moi en tant que potentiel enseignant de FLE au Japon. Les entretiens ont lieu dans une période d'une semaine à quinze jours après la formation. Là encore, la discussion se veut la plus détendue et amicale possible, en français ou en japonais si nécessaire. Le contenu de ces conversations a permis la rédaction du dernier chapitre de ce mémoire. Chaque locuteur est aussi amené à refaire le tableau phonologique de mémoire (sans les couleurs), et dans un deuxième temps je lui montre la version originale et je lui propose de la retravailler brièvement, dans l'objectif de m'en resservir en fin d'entretien. L'ensemble des mots cibles prononcés spontanément par le locuteur a été analysé et comparé avec les réalisations en T<sub>1</sub>, mais le champs lexical dominant étant assez éloigné de celui sollicité lors du premier entretien, il a fallu trouver un complément pour faire prononcer à nouveau l'ensemble des mots cibles du protocole. Il a donc été choisi ici

---

<sup>4</sup>Il pourrait toujours être intéressant de proposer plus de mots cibles avec les mêmes clusters, et comparer les réalisations de ceux qui ont été vus en cours et ceux qui ne l'ont pas été.

de demander aux locuteurs de se remémorer les 7 mots obligatoires qu'ils avaient dû utiliser un mois plus tôt<sup>5</sup>, et pour les mots oubliés, répondre à des questions du type « *quel pays c'est, ici ?* » ou « *c'est quoi le nouveau nom de cette université ?* » si *université* a déjà été prononcé. Puis les 4 nombres (34, 48, 67 et 78) écrits en chiffres arabes sont montrés à nouveau, le locuteur doit alors les prononcer ; et enfin, je leur demande de recompter de 20 à 1. Avant de terminer l'entretien, je propose aux locuteurs de pointer les nombres 4, 14, 19 et 34 sur le tableau phonologique, tout en les prononçant (les nombres sont montrés un par un sur une petite carte, toujours en chiffres arabes). Ces quatre derniers nombres ont été choisis car ils posaient particulièrement problème lors de la première phase d'observations. L'enregistrement audio ne suffisant pas, les gestes de pointage ont été soigneusement notés sur une feuille pendant l'entretien, et sont présentés dans le Chapitre VII.

### 5.3 L'analyse des données

Face au flux abondant et ininterrompu de la parole, il est nécessaire de séquencer, isoler et comparer ce qui est comparable. Les mots obligatoires de notre protocole permettent de cibler des séquences plus ou moins communes chez tous les locuteurs. L'isolation de ces séquences, tout en conservant leur contexte phonologique, sémantique et prosodique, permet une mise en parallèle des différentes réalisations du même item, et il est alors possible de les comparer.

#### 5.3.1 Isolation des mots utilisés par tous les locuteurs

À la fin de tous les entretiens de la phase 1, on a donc un minimum de 17 mots cibles dont l'analyse semble pertinente pour notre étude (*cf.* tableau 5.1, section précédente). Étant donné le petit nombre de locuteurs, une transcription intégrale de tous les entretiens a été effectuée, dans l'objectif de récupérer d'autres mots communs dans tous les enregistrements<sup>6</sup>. Les compétences globales de français des locuteurs étant très hétérogènes, la quantité d'énoncés en français est bien plus réduite pour le

---

<sup>5</sup>J'ai été surpris de voir à quel point les locuteurs se rappelaient des mots, malgré le mois les séparant du premier entretien.

<sup>6</sup>Il s'agit de l'ensemble des énoncés en français du locuteur japonophone natif qui a été transcrit, le reste n'étant pas pertinent pour cette étude.

locuteur débutant (N<sub>2</sub>) que pour le locuteur le plus avancé (N<sub>5</sub>). Le nombre de mots susceptibles d'être prononcés par tous les locuteurs étant plus réduit pour le locuteur débutant, c'est à partir de la transcription de son entretien que j'ai donc travaillé. Après avoir listé tous les mots présentant un cluster ou une finale consonantique, j'ai recherché automatiquement quels mots étaient aussi présents dans les autres transcriptions, qu'ils soient sous la même forme ou dérivés. Cependant, sur les 26 items présentant un cluster consonantique chez le locuteur débutant, et excepté les 17 items cibles du protocole, un seul est prononcé par les 7 autres locuteurs. Le mot *anglais* sera donc le seul à être ajouté à la liste des items cibles. De plus, le mot *France* n'a pas été prononcé par un des locuteurs, il sera tout de même conservé pour les analyses quantitatives automatiques, me ne pourra pas être comparé à sa réalisation en T<sub>2</sub> par ce même locuteur. Treize items à finale monoconsonantique et communs à tous les locuteurs sont aussi répertoriés.

Il est nécessaire de noter que je me suis uniquement basé sur les mots écrits dans les transcriptions d'entretien. Se limiter à la dimension du mot permet une sélection d'items cibles plus rapide, mais empêche la prise en compte de beaucoup d'occurrences de clusters consonantiques issus de la combinaisons de deux mots ou de la resyllabation naturelle de l'oral. Des occurrences fréquentes comme « est-ce que » /ɛskə/ ou « j'aime bien » /ʒɛmbjɛ/ sont donc ignorées. De plus, une réflexion plus approfondie devrait être menée sur la pertinence de se limiter à un même mot pour étudier un cluster. Si l'on utilise toujours le mot cible « anglais » pour observer la réalisation de /gl/, on est sûr que l'environnement phonologique ne varie pas – ou ne devrait pas varier – d'une occurrence à l'autre. Mais si l'on veut réellement se concentrer sur le problème qu'engendre la combinaison /g+l/, il faudrait observer d'autres occurrences dans d'autres environnements. Et puis la sécurité du mot n'est pas garantie avec un cluster en attaque ou en finale, puisque le mot précédent ou suivant va influencer sa réalisation. La définition du *mot* peut se limiter à l'espace graphique, mais lorsqu'il s'agit de parole, les frontières sont moins évidentes. Doit-on traiter « les Alpes » /lezalp/ ou « Grenoble Alpe » /gʁø̃nɔ̃blalp/ en un ou deux mots ? La différence dans les réalisations rend-elle la comparaison impossible ? Même s'il s'agit de la même séquence phonologique, bien des paramètres entrent en jeu et font varier le signal sonore final (le timbre, le contexte sémantique, l'intensité, le niveau de français et bien-sûr la région d'origine du locuteur), la variation est inévitable et

la question est comment la prendre en compte dans les analyses.

### 5.3.2 Alignement automatique avec EasyAlign

Si l'analyse manuelle au cas par cas, en passant par exemple par une transcription fine en API, est toujours plus précise, l'analyse automatique permet de traiter plus de données et selon des paramètres invariants selon le locuteur, la période d'analyse, ou l'observateur. Mais l'objectivité résultant de cette invariance pose un problème majeur : des paramètres considérés comme non-pertinents dans une analyse manuelle peuvent grandement influencer les résultats. Il est donc important de toujours vérifier les résultats, les observer attentivement pour détecter les éventuels problèmes. Si l'on se contente d'interpréter les résultats obtenus automatiquement, on peut se retrouver à faire des conclusions bien éloignées de celles qui auraient pu être faites après une analyse manuelle.

Dans la présente étude, l'objectif est d'observer comment sont réalisés des mots cibles (et surtout les clusters consonantiques qu'ils contiennent) par différents locuteur sur les deux périodes de temps  $T_1$  et  $T_2$ . Cette observation se concentrera sur la segmentation phonémique de ces mots, afin de savoir comment sont réalisés les clusters consonantiques. Le chapitre IV nous a permis de définir l'hypothèse selon laquelle le locuteur japonophone a tendance à insérer une voyelle d'appui (épenthèse) entre deux consonnes normalement concaténées.

Pour ce faire, nous utiliserons le logiciel libre néerlandais Praat, accompagné de l'extension EasyAlign. Praat permet la manipulation et le traitement du son vocal, EasyAlign permet quant à lui d'aligner automatiquement des phonèmes sur et en fonction du signal sonore. Grâce à ces outils il est donc possible de transcrire automatiquement les mots cibles, pour observer quels phonèmes les constituent, selon le programme. Les analyses automatiques de cette étude se cantonneront à la transcription, mais un grand nombre d'analyses différentes sont envisageables et quelques unes seront proposées en ouverture.

#### Utilisation d'EasyAlign

EasyAlign propose une série de scripts permettant à Praat d'aligner automatiquement des phonèmes à partir d'une transcription orthographique d'un enregistrement

sonore. Les prérequis sont les suivants : ouvrir un fichier son au format wave dans Praat, et annoter cet enregistrement pour obtenir une transcription orthographique dans un fichier .textgrid. On obtient le résultat suivant :

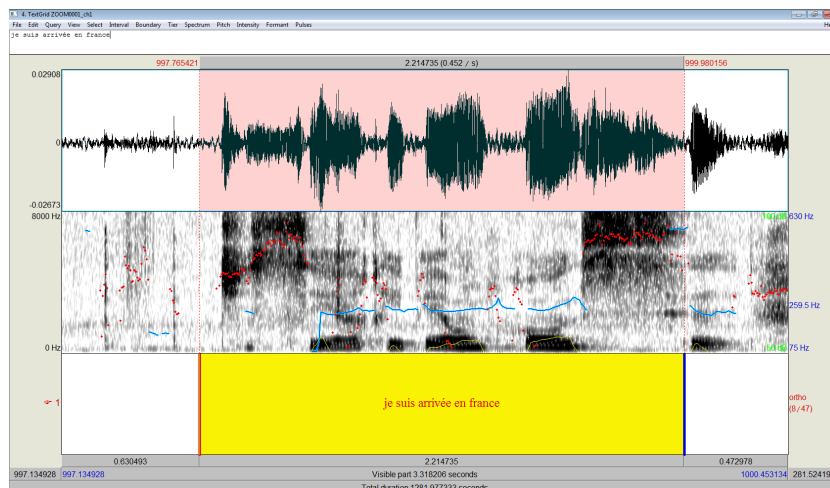


FIG. 5.3 : Préparation de l'enregistrement et de son annotation avant de lancer EasyAlign

EasyAlign est constitué d'un dictionnaire phonétique en plusieurs langues, proposant une transcription par défaut de tous les mots de la transcription orthographique, s'ils font partie du dictionnaire. En lançant le premier script d'EasyAlign, on obtient la transcription phonétique intégrale du fichier .textgrid, resyllabisée en fonction de l'environnement phonémique de chaque mot (*cf.* illustration 5.4). À ce stade il s'agit d'une transcription « à l'aveugle » sans aucun lien avec la réalité phonétique et phonologique de l'enregistrement, EasyAlign ne se base là que sur le fichier texte et sur sa base de données. Remarquons que les symboles phonétiques utilisés sont ceux du SAMPA (*cf.* tableau de correspondances avec l'API en annexe ??).

Si l'on lance directement le script suivant, le programme va découper la transcription phonétique en syllabes, et aligner ces dernières sur le spectre du fichier son. Puis dans la foulée, il va décomposer les syllabes en phonèmes pour les aligner. On obtient alors le résultat présenté dans la figure 5.5.

EasyAlign aligne donc ici ce qu'il pense « entendre », soit la transcription phonétique pré-enregistrée dans sa base de données. Son observation du signal se limite ici à poser des frontières, d'abord de syllabes puis de phonèmes, sur le signal sonore en

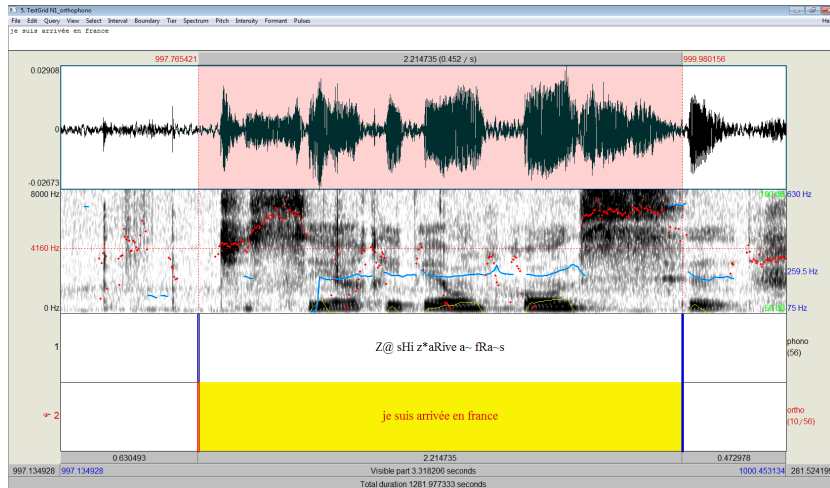


FIG. 5.4 : Étape 1 : Phonétisation de la transcription orthographique

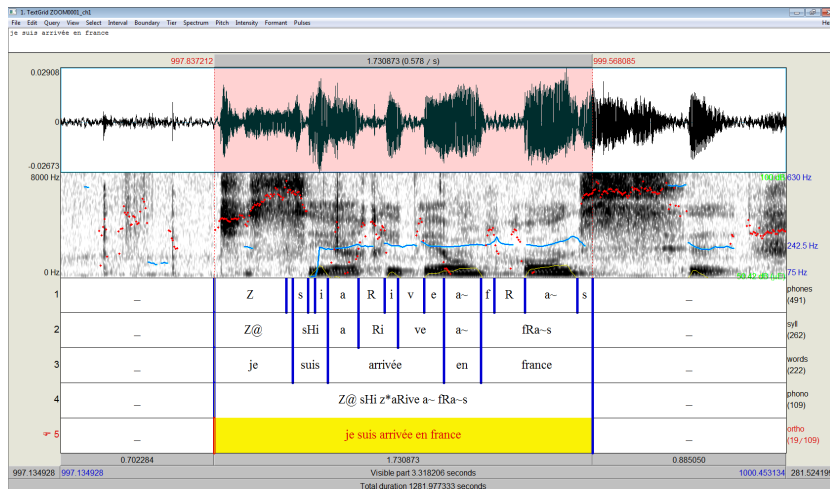


FIG. 5.5 : Étape 2 : Alignement des syllabes & phonèmes sur le signal sonore

fonction de son « relief » (fréquences, durées, silences et explosions pour les occlusives etc.). Cet alignement de la transcription automatique ne nous intéresse pas réellement ici, puisque nous souhaitons observer si certains phonèmes sont ou non réalisés par le locuteur (ou du moins détectés ou non par le programme). Vous aurez peut-être remarqué l'astérisque dans la transcription phonétique automatique. Celui-ci indique que le phonème précédant est optionnel, et EasyAlign va juger, d'après la réalité du signal, si celui-ci est perceptible ou non, et l'aligner le cas échéant. C'est donc en proposant des options à EasyAlign que nous allons vérifier automatiquement si les locuteurs prononcent ou non des voyelles d'appui entre les consonnes des clusters, ou après une consonne en finale.

Grâce aux observations de la partie théorique, un ensemble d'options seront donc proposées dans les transcriptions orthographiques de tous les locuteurs, afin d'évaluer automatiquement les types de réalisations produites.

## Chapitre 6

# Déroulement de la formation & observations

Les quatre séances de la formation ont été réparties sur deux semaines, les mardis et jeudis après-midi à partir de 17h30. Huit participants japonais ont été présents aux quatre séances, dont un qui n'a pas passé l'entretien pré-formation. Le locuteur N<sub>3</sub> a dû s'absenter à partir de la troisième séance pour des raisons familiales. Seuls les sept locuteurs présents à l'enregistrement pré-formation seront invités à l'entretien de la phase 3 (post-formation). L'enseignant et les participants se sont rencontrés pour la première fois lors de la séance 1.

### 6.1 Découverte de l'outil Kinéphones

La première séance était consacrée à la prise de contact et à la découverte du tableau phonologique. Dès le début, l'enseignant a su s'adapter aux différents niveaux des participants, en évaluant dans les premiers échanges les niveaux de compréhension et production orale approximatifs. Malgré une certaine discrétion de la part des participants, le contact avec l'enseignant était bon et les échanges dynamiques. Pour les deux premières séances, le tableau phonologique ne comportait ni diphtongues, ni affriquées, ni semi-voyelles, afin de présenter un outil épuré et moins complexe. Après avoir échangé brièvement avec les participants et détendu l'atmosphère avec un peu d'humour, l'enseignant s'est tourné vers le tableau et à pointé la couleur blanc.



Puis, face à la classe, il ouvre grand la bouche, langue en position antérieure et sans un bruit, symbolise la poussée d'air depuis les poumons par un geste de la main. Après quelques secondes d'hésitation un locuteur prononce /a/, suivi immédiatement par le reste de la classe. L'enseignant acquiesce ; le voile commence à se lever sur ce tableau mystérieusement coloré.

L'enseignant réitère le pointage du blanc et le geste, en désignant du doigt un à un les participants, qui doivent alors prononcer le son. Puis il pointe de manière prolongée la couleur, en symbolisant une roue qui tourne avec la main. Un autre participant prononce alors plusieurs /a/ à la suite, mais l'enseignant fait non de la tête et recommence jusqu'à ce qu'on prononce un /a/ continu. Ça y est, après moins d'une minute de cours, la classe réagit machinalement au pointage de la couleur et au type de pointage de l'enseignant.

L'enseignant pointe ensuite le rouge, en haut à gauche du tableau, étire ses lèvres, bouche presque fermée, et symbolise à nouveau une poussée d'air de la main. Tout le monde prononce alors /i/. Il refait différents pointages, pointe tantôt le blanc, tantôt l'une puis l'autre couleur, tantôt les deux enchaînées. La classe comprend vite le fonctionnement de l'outil.

## 6.2 Pointage des premières difficultés

S'en suit alors le pointage du vert foncé, en haut à droite du tableau. D'une bouche quasi-fermée, les lèvres allongées vers l'avant, l'enseignant réitère son geste de la main. On entend alors des /o/ et des /ɔ/ et après quelques répétitions du geste, on entend plutôt [u]. L'enseignant pointe ici une grande difficulté phonologique des apprenants : le /u/, qui n'existe en japonais que sous sa réalisation centralisée, moins labiale et plus détendue qu'en français (*cf.* chapitre III). Certains prononcent /ø/, bien qu'ils semblent avoir compris qu'il s'agit du /u/ français. L'enseignant prononce alors pour la première fois la voyelle. Par mimétisme les participants se rapprochent du /u/ puis reviennent naturellement sur une variante centrale détendue. Le problème est qu'il ne semblent pas percevoir de réelle distinction entre ces sons. Nous francophones, entendons tantôt l'un ou l'autre, mais pour une oreille japonaise, il s'agit du même phonème car le japonais ne nécessite pas de les distinguer.

L'enseignant, qui a repéré un participant prononçant un [u] très clair, ce sert

de lui pour aider les autres. Il le fait prononcer, puis désigne un autre apprenant qui prononce alors avec enthousiasme /ø/. L'enseignant infirme alors de la tête et désigne à nouveau le premier participant. Il demande alors qui est en mesure de le prononcer, et l'enseignant donne un *feedback* (confirmation ou infirmation) à chaque proposition. Il s'assure que toute la classe participe.

Le /u/ ne s'acquerra pas si vite. Mais il ne faut pas trop insister, et détendre la mâchoire et l'atmosphère en prononçant les voyelles plus familières aux participants. En demandant à la classe de prononcer des suites de voyelles pointées, l'enseignant tente de leur faire prendre conscience de la position de la langue, tantôt en avant, tantôt en arrière, la bouche tantôt mi-fermée, tantôt ouverte, et enfin les lèvres allongées ou non.

### 6.3 Suite de la présentation du tableau

La voyelle /u/ étant décidément centralisée par la plupart des participants, l'enseignant présente alors le /y/ qui est représenté en orange, en haut et au centre du tableau<sup>1</sup>. S'en suivront /e/, /ø/ et /o/. Les autres voyelles ne sont pas présentées tout de suite. Pour l'instant, les participants se sont familiarisés avec les sept voyelles les plus fréquentes. Découvrons maintenant quelques couleurs de la moitié basse du tableau.

De la même manière que pour les voyelles, l'enseignant pointe une première couleur du bas du tableau puis le blanc des voyelles, touche ses alvéoles de la langue puis ouvre grand la bouche. Les participants ont compris qu'il s'agit du /a/ précédé d'une consonne et produise à l'unanimité /ka/. Pourquoi la consonne /k/ ? Serait-ce parce que c'est la première consonne de leur syllabaire ? L'enseignant infirme et les participants font d'autres tentatives (/ta/, /ma/) jusqu'à ce que l'un d'eux prononce /na/, confirmé par l'enseignant. Ce dernier le fait alors répéter aux autres, puis change la voyelle avec les six autres connues. Il pointera ensuite trois couleurs à la suite, de structure (V<sub>1</sub>CV<sub>2</sub>), compris sans difficulté. Une fois le /m/ présenté, il s'agit de diversifier les structures : CVCV, VCVC, CVC etc. Le /n/ en coda n'étant pas une

<sup>1</sup>La dernière version du tableau phonologique de Kinéphones présente le /y/ à côté du /i/, à gauche du tableau. La vieille version du tableau utilisée en cours était peut-être un peu moins adaptée à la classe, qui avait tendance à centraliser et délabialiser légèrement le /y/, peut-être sous son influence. Il pouvait y avoir aussi confusion avec le /ø/. Cependant les couleurs de l'ancienne version sont un peu plus espacées, et le pointage/repérage de celles-ci est légèrement plus simple.

difficulté de par son existence dans le système japonais, le /m/ en coda étant quant à lui purement étranger.

Chaque présentation d'une nouvelle consonne est précédée d'une révision de toutes les couleurs connues, à travers des structures et combinaisons diverses, pour bien automatiser l'utilisation du tableau, et bien faire prendre conscience de l'*unité phonologique* et surtout l'unité consonantique qui n'existe pas en japonais. Les couleurs sont comme des pièces de puzzle qui sont combinables à volonté. Le fait de pointer des pseudo-mots n'existant pas en français peut paraître étrange et inapproprié, mais il est important de passer par cette phase de combinaison libre pour séparer le phonème consonantique de la voyelle à laquelle il est forcément attaché en japonais. C'est un passage essentiel pour combiner par la suite deux ou trois consonnes et former les clusters consonantiques du français.

Après les nasales /m/ et /n/, ce sont les fricatives /f/ et /s/ qui sont présentées. Le /v/ est rapidement deviné, grâce à sa proximité avec /f/, et le geste de l'enseignant, identique à celui de /f/. Le /z/ devient alors évident. On constate que les fricatives vont par paire, et que la seule différence entre chaque élément d'une paire, c'est la vibration des cordes vocales. Jusqu'à là toutes les consonnes peuvent être tenues.

Suivra ensuite la latérale /l/ avant un petit exercice où les participants doivent faire deviner le nom d'une ville ou d'un pays à la classe en pointant les couleurs au tableau. L'enseignant intervient alors si un phonème vient à manquer. L'exercice était certainement plus difficile que prévu à cause des connaissances régionales qu'il requérait, en plus des seules sept consonnes et sept voyelles connues. Lorsque c'est l'enseignant qui pointe, les réactions sont plus nombreuses : le phonème /t/ étant pourtant encore inconnu, le pointage du mot "Italie" n'a pas posé problème. On pourra noter qu'un des participants pointe /s/ /i/ /n/ pour "Chine". Est-ce par manque du phonème /ʃ/, ou est-ce un réel problème de distinction ? On peut rappeler que lors de la première combinaison /s/ /i/, plusieurs participants ont prononcé [çi]. Cette production ne justifie pas la difficulté de perception, mais elle y est certainement liée.

On peut constater que le phonème [ʒ] est aussi source d'importantes difficultés chez certains participants. Sa réalisation ne semble pas un problème pour les locuteurs japonophones, mais il est souvent précédé d'une occlusive apico-dentale voisée, amenant à l'affrication [dʒ]. La difficulté était donc de faire prendre conscience d'une unité /ʒ/, qu'il est possible de faire précéder par une autre unité /d/. On a vu que le

#### 6.4. UTILISATION D'AUTRES SUPPORTS EN PARALLÈLE DU TABLEAU<sub>91</sub>

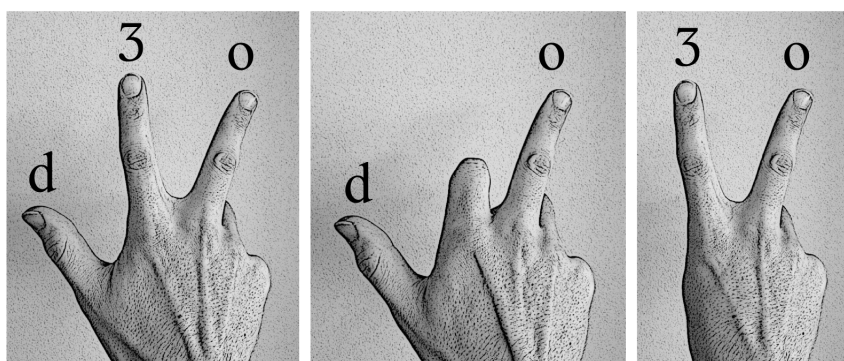


FIG. 6.1 : Distinction phonologique entre /ɜ/ et /dɜ/

japonais standard ne présentait que les unités /d/ ou /z/, dont l'allophone palatalisé commun à ces deux phonèmes se réalise tantôt [ʝ], tantôt [ɟʝ], sans constituer d'opposition phonologique (KINDAICHI et MAËS 1978). Si aucune difficulté n'est présente sur la plan de la production, c'est bien au niveau perceptif que doit opérer la prise de conscience, car c'est à la distinction entre /ɜ/ et /dɜ/ que se sont rendus sourds la plupart des japonais. L'enseignant va alors représenter chaque phonème avec ses doigts : le pouce représentant /d/, l'index /ɜ/ et le majeur une voyelle quelconque. C'est alors en montrant tantôt le pouce et le majeur (/dV/), tantôt l'index et le majeur (/ɜV/) ou les trois doigts en même temps (/dɜV/), que l'enseignant fait prendre conscience de l'indépendance des phonèmes.

### 6.4 Utilisation d'autres supports en parallèle du tableau

Dès la fin de la deuxième séance, et après s'être familiarisé avec la plupart des couleurs du tableau, d'autres supports sont présentés à la classe afin de travailler des énoncés concrets. Une illustration de plan d'une ville et une carte des régions de France permettront d'aborder les champs lexicaux des lieux courants de la ville, auxquels sont quotidiennement confrontés les participants à la formation, ainsi que les régions et villes de France, ou encore les pays limitrophes.

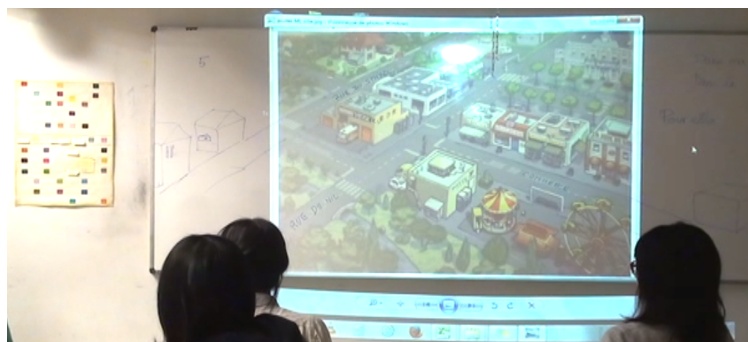


FIG. 6.2 : Plan de la ville avec extensions dessinées de part et d'autre

### 6.4.1 Un plan de ville

Le plan utilisé est une illustration que propose LAURENT (2004) et LAURENT (2009) et qui présente une ville simple avec les principaux commerces, une école, une poste etc. Afin d'introduire d'autres noms communs constitués de clusters, l'enseignant a l'idée de dessiner autour du plan, et ajouter un restaurant, un super-marché ou une banque. Les possibilités sont quasiment illimitées.

Après avoir fait décrire la ville à la classe, en demandant le nom des lieux ou en faisant lire le nom des rues, l'enseignant propose à la classe la structure suivante : "Dans la rue..., il y a... (et...)". À tout moment l'enseignant fait appel au tableau phonologique pour décomposer les mots ou les phrases. La phrase type permet de mettre dans un énoncé complet les mots qui étaient jusqu'alors isolés. Par la suite, une autre phrase sera proposée : "Pour aller de... à..., on passe par...". Les phrases types sont écrites de cette façon lacunaire au tableau. Les différents noms utilisés ne sont ni écrits ni transcrits.

### 6.4.2 Une carte des régions de France

L'utilisation de la carte de France a permis plusieurs choses. D'une part elle a apporté un nouveau champs lexical à travailler, mais elle a aussi introduit la graphie et la lecture dans la formation. En effet, la carte présentait les régions de France ; les principales agglomérations ainsi que les pays limitrophes étaient aussi écrits. Avec cet outil, encore une fois, l'enseignant a utilisé l'énoncé lacunaire "Pour aller de... à..., on passe par...".

## 6.5 Conclusion

On aurait pu penser que les participants aient été plus pudiques et gênés de prononcer des sons et grimacer en public – et surtout devant la caméra. Une telle expérience réalisée au Japon n’aurait probablement pas eu le même effet. La force, la brièveté et la précision du pointage incite à prononcer fort et distinctement. Lorsque les participants prononcent trop doucement, l’enseignant repointe plus énergiquement la couleur, et la classe prononce alors naturellement plus fort et distinctement. On ressent un intérêt particulier du tableau phonologique pour /u/ (qui pose vraiment problème) et pour les clusters et codas consonantiques. Les choix de l’outil et de l’approche pédagogique semblent avoir été totalement adaptés à l’objet d’étude de ce mémoire.

On constate qu’il y a bien différents profils de participants, certains ayant plus de difficultés que d’autres, et ce indépendamment du nombre d’années d’apprentissage du français ou de la durée de séjour en France. Même si quelques participants prennent globalement moins la parole que les autres (N<sub>2</sub>, N<sub>3</sub>, N<sub>10</sub>), tout le monde participe dans les 4 séances. L’attention est focalisée sur le tableau et l’enseignant, mais on sent aussi une écoute mutuelle entre les apprenants, pour s’autocorriger en fonction des productions du groupe.

Les productions pendant le cours sont très satisfaisantes : les problèmes sont surmontés par les apprenants grâce aux *feedbacks* de l’enseignant et à la réaction du groupe. Il s’agit maintenant de savoir si des évolutions sur le long terme sont observables dans la prononciation des participants. Des enregistrements une à deux semaines après le dernier cours auront lieu, pour comparer les réalisations, ainsi que pour faire une évaluation qualitative de la formation en parlant individuellement avec tous les participants.



# Chapitre 7

## Analyse des données

Après avoir enregistré les participants avant et après la formation, il est maintenant possible d'analyser les signaux à l'aide du logiciel Praat, et d'utiliser EasyAlign pour transcrire phonétiquement les énoncés des locuteurs. Il sera d'abord présenté dans ce chapitre les différents essais d'alignement, les problèmes rencontrés ainsi que les solutions trouvées pour obtenir des données analysables. Seront ensuite détaillés le traitement des données et la présentation des résultats pour les deux périodes d'enregistrements  $T_1$  &  $T_2$ .

### 7.1 Alignement automatique avec EasyAlign

#### 7.1.1 Conception d'une première liste de lois de pose d'options

Il a initialement été constitué pour cette étude, une liste de lois pour poser des options dans la transcription phonétique des fichiers textgrid, en fonction des hypothèses issues de l'étude de cas du chapitre IV. Ces options ont été posées sur l'ensemble des énoncés français prononcés par les locuteurs, une fois les enregistrements annotés des transcriptions orthographiques. Les lois étaient les suivantes :

- @\* entre les consonnes de clusters (voyelle centrale relâchée proche du  $\ddot{u}$  observé chez le locuteur du chapitre IV) ;
- $d^*$  devant  $Z$  et  $z$  ;
- @\* après les codas consonantiques (sauf  $n$ ,  $t$ ,  $d$ ) ;
- $o^*$  après les codas consonantiques  $t$  et  $d$  ;



- tous les phonèmes du cluster sont optionnels, car certains ne sont parfois pas réalisés :  $C_1^* @ C_2^*$  ;
- $tS^*y$  au lieu de  $ty$  ;
- $s^*S^*i$  et  $s^*S^*Hi$  à la place de « si » et « sH », de même pour les équivalents voisés ( $z^*Z^*i$ ,  $z^*Z^*Hi$ ) ;
- $R^*l^*$  au lieu de  $R$  et  $l$ .

Seulement, après maintes tentatives, il s'avère que le programme échoue systématiquement dans l'alignement d'un grand nombre de segments, et ignore l'ensemble des options, à de rares exceptions près. Les premiers tests sont effectués sur les enregistrements des locuteurs N1 et N8 (couvrant les deux genres, avec une prononciation vraisemblablement influencée par la langue maternelle). On assiste notamment à une fréquente redéfinition des frontières par EasyAlign, qui déplace les frontières manuellement positionnées dans le textgrid original, et semble les décaler arbitrairement en plein milieu des segments, alignant alors tous les phonèmes sur une partie seulement de l'énoncé correspondant. De nombreux phonèmes étant alignés sur une partie du signal qui ne leur correspond pas, l'analyse ne peut être poursuivie.

Ces échecs d'alignement et déplacements de frontières ne semblent pas liés à la longueur des segments de la transcription orthographique, ni au nombre important d'options par segment. En effet, certains segments particulièrement longs sont correctement alignés (jusqu'à 8 secondes 42 pour N1, il s'agit d'une énumération de neuf à un, avec hésitation, 30 phonèmes à aligner), alors que l'alignement échoue, ou est incorrect sur plusieurs segments plus courts. Plusieurs tests ont été réalisés afin d'identifier le problème. Les segments contenant 8 à 9 options ont été recoupés, l'alignement semble fonctionner jusqu'à 6 options par segment, l'alignement échoue cependant à partir de trois options consécutives ; les silences supérieurs à une seconde en début ou fin de segment semblent perturber l'alignement lorsqu'il y a des options. Les échecs d'alignement et les problèmes de redéfinition de frontière peuvent donc être évités, cependant un problème persiste malgré les différentes tentatives, toutes les options, ou presque, sont ignorées.

On constate que dans beaucoup de segments, même les phonèmes initialement obligatoires (dans la phonétisation automatique<sup>1</sup>), et mis en option par la suite, sont ignorés. Observons par exemple le mot « université », transcrit automatiquement

---

<sup>1</sup>Étape 1 de l'alignement – cf. chapitre v section 3.

*ynivERsite* et manuellement *ynivERs\*S\*ite*, à l'étape 2 il est aligné *ynivERite*, alors que la fricative est clairement perceptible et visible sur le spectrogramme.

Certains phonèmes semblent aussi poser problème à EasyAlign, sans parler d'option, on voit notamment que le glide *j* est presque toujours aligné sur une durée excessive, ce qui décale alors l'alignement ; c'est aussi le cas des occlusives en attaque, dont le silence initial tend à décaler les frontières du textgrid ; les occlusives en finale sont souvent coupées par EasyAlign et se retrouvent à l'extérieur des segments (*Ronaldp* devient *Ronal* ; *kaRa-tHit* devient *kaRa-tHi* etc.) et ce même lorsque l'occlusive est clairement perceptible. Dans tous les cas, est aligné coûte que coûte l'ensemble des phonèmes n'étant pas suivis d'un astérisque.

L'illustration 7.1 montre un cas classique d'option ignorée (*d\**) devant la fricative *Z* malgré sa présence flagrante dans le spectrogramme. On distingue en effet une attaque très claire, caractéristique des *bursts* des occlusives, et on peut même voir l'emprunte du /d/ juste avant la fricative sur une fréquence plus basse. S'il une fricative non-affriquée était prononcée, celle-ci laisserait seulement une empreinte de haute fréquence, continue, avec un début et un fin plus progressives.

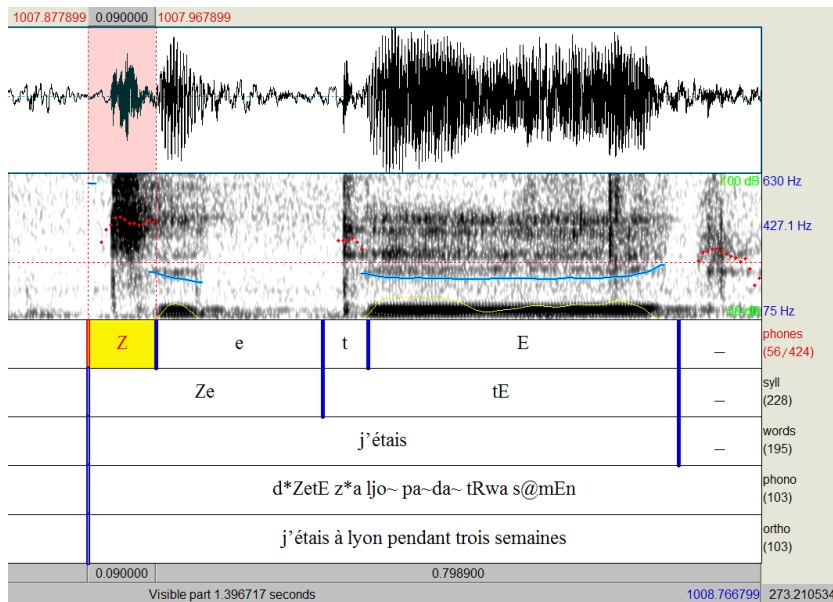


FIG. 7.1 : Cas d'option ignorée malgré la présence du phonème dans le signal

Enfin quelques problèmes supplémentaires proviennent de la transcription phonétique automatique du textgrid (étape 1), où certains phonèmes viennent à manquer,

ou sont obligatoires quand il devraient être optionnels. Ces erreurs doivent être corrigées avant le lancement du script d'alignement. Parmi les mots cibles on trouve notamment « dix-neuf » transcrit *di nɔf* au lieu de *di zɔf*, ou la combinaison « huit neuf » transcrite *Hi nɔf* ; « Alpes » n'est pas reconnu par le dictionnaire du programme et se retrouve épelé *a El pe ɔ Es*, il faut alors l'éditer manuellement. Remarquons que toutes les hésitations transcrites dans le textgrid, si elles ne correspondent pas à un mot de la base de données du programme, sont épelées. Par exemple le locuteur N1 prononce « qua quatre », écrit de cette manière dans le textgrid il devient alors *ky y a katR*, là encore, l'édition manuelle résout le problème mais ne doit pas être oubliée sous peine de corrompre l'alignement<sup>2</sup>.

La détection automatique des épenthèses se révèle donc plus compliquée que prévu. L'alignement automatique avec options est un échec. Lorsque l'on compare sur un même écran l'alignement sans option ajoutée, et l'alignement avec options, on constate que ce dernier présente des phonèmes souvent décalés par rapport au signal, et de nombreux segments ne sont pas alignés. L'alignement sans option est quant à lui satisfaisant, la gestion des silences et des hésitations est plutôt réussie, tant que la transcription est fidèle ; et tous les phonèmes obligatoires sont alignés sur le signal, même lorsque la prononciation est assez éloignée de celle d'un natif.

EasyAlign est un programme performant, et il semble bien fonctionner avec des locuteurs dont la prononciation est éloignée de celle à partir de laquelle il a été conçu (ou un français « *standard* ». Mais notre ambition d'utiliser les options pour détecter les épenthèses est-elle vaine ? Les options sont-elles réellement appropriées à la production des locuteurs ? Ou EasyAlign suit-il des règles d'alignement selon lesquelles, par exemple, il ne pourrait pas aligner d'occlusive devant une fricative (les affriquées étant plutôt rares en français) ?

### 7.1.2 Transcription fine & nouveau choix d'options

Afin de vérifier si les options précédemment utilisées sont adaptées aux locuteurs, trois transcriptions phonétiques fines vont être réalisées. Tout d'abord, il semble plus raisonnable de limiter les analyses aux mots cibles uniquement, et poser les options

---

<sup>2</sup>D'autres bugs mineurs ont été relevés par la suite : « vingt » et transcrit avec un *t* obligatoire, l'accent circonflexe sur « août » n'est pas reconnu, alors qu'il l'est sur « être », il doit donc être enlevé.

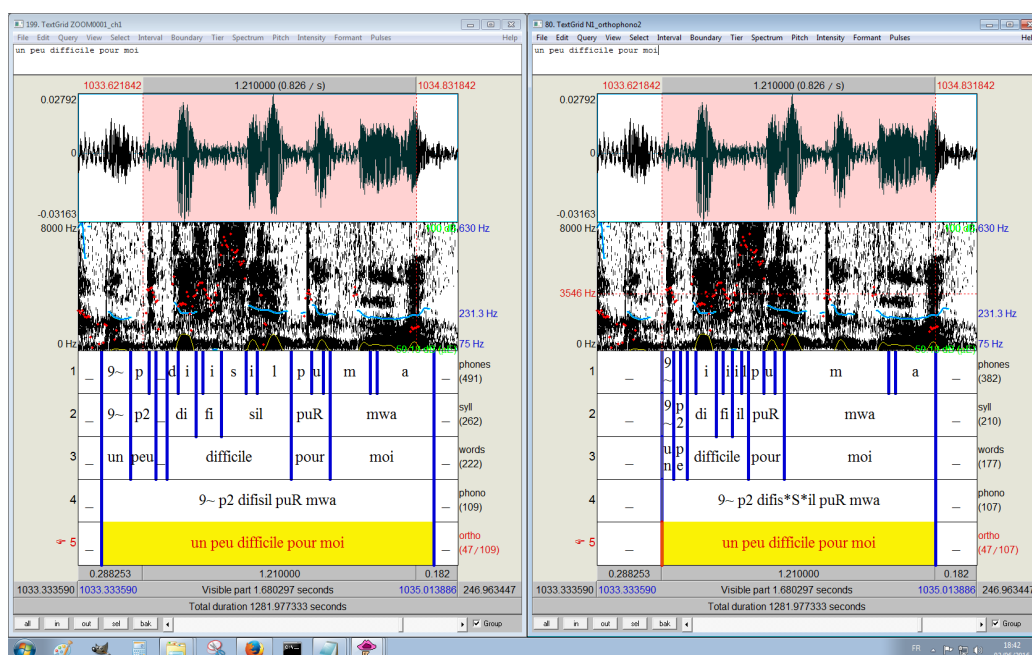


FIG. 7.2 : Comparaison de l'alignement avec (droite) et sans (gauche) options

seulement sur ces mots-là. De plus, tous les segments devront contenir au moins un mot cible, le reste des énoncés sera ignoré pour cette analyse quantitative. Les locuteurs N1, N6 et N8 semblent être ceux avec le plus de difficultés sur le plan de la prononciation. Les mots cibles sont donc répartis dans un tableau, chaque occurrence du mot est prise en compte, ainsi que les syllabes précédant et suivant le mot, pour conserver le contexte phonologique immédiat. Le caractère # signale que le mot est précédé ou suivi d'un silence. Les transcriptions sont présentées en annexe.

Grâce à quelques observations au cas par cas, on se rend compte que les épenthèses sont finalement une stratégie assez rarement utilisée par ces locuteurs, ou de manière isolée. En réalité, même avec trois locuteurs on se retrouve face à une grande diversité de productions, et il n'est pas évident de dégager des tendances générales. On peut percevoir une certaine fréquence d'utilisation de la stratégie de substitution d'une des consonnes du cluster par une consonne différente, voire l'omission de certains éléments pour simplifier la prononciation. On observe par exemple une tendance du /ʁ/ à être sourd et à assourdir sont environnement ([gʁ] devient [kχ]); dans « université » par contre, le /ʁ/ a plutôt tendance à disparaître, ou le /t/ parfois, accompagné du /i/ pour simplifier la structure ; les /f/ de « France » et « dix-neuf » sont souvent très

faibles, voire absents ; « trois » est souvent prononcé [tʁa] ou [twa]. Tout cela pourrait expliquer l'impossibilité d'aligner les options initialement proposées, puisque ce ne sont pas les bons phonèmes qui ont été mis en option.

En fonction des réalisations observées, de nouvelles options ont été définies pour chaque mots cibles, et vont être posées systématiquement dans les transcriptions de tous les enregistrements en T<sub>1</sub> et T<sub>2</sub>. Les nouvelles options sont présentées dans le tableau 7.1.

Mot cible	Transcription par défaut	Transcription complétée
Grenoble	gR@nObl	gR*@nObl*
université	ynivERsite	ynivER*site
Stendhal	ste-dal	ste-dal*
Alpes	alp	al@*p
classe	klas	k@*las
France	fra-s	f*Ra-s
anglais	a-gle	a-g@*le
trois	tRwa	tR*w*a
quatre	katR	kj*atR
treize	tREz	tR*Ez
quatorze	katORz	katR*OR*z
dix-huit	dizHit	diz*Z*Hit
dix-neuf	diznɔf	diz*nɔf*
trente-quatre	tRa-tkatR	tR*a-t*kj*atR*
soixante-sept	swasa-tsEt	swasa-t*sEt
soixante-dix-huit	swasa-tdizHit	swasa-t*diz*Z*Hit

TAB. 7.1 : Liste des nouvelles options

Précisons que les mots « Grenoble » et « Université » devaient d'abord être transcrits  $g^*k^*R^*@nObl^*$  et  $yniv^*b^*ER^*site$ , mais après plusieurs échecs d'alignement, ils ont été simplifiés par contrainte technique. Ajoutons aussi que les mots « test » et « quarante-huit » n'ont posé aucune difficulté de production, ils ont donc été jugés non-pertinents pour cette étude et aucune option ne sera posée.

Cette fois-ci les résultats sont concluants, et bien que de légères imperfections dans l'alignement demeurent, on obtient une transcription tout à fait proche de la réalité. À chaque alignement, une écoute rapide des segments permet de vérifier si la transcription et l'alignement correspondent au signal, puis les résultats sont récupérés directement dans le fichier textgrid, via un éditeur de texte.

### 7.1.3 Traitement des données

Lorsqu'on ouvre un textgrid avec un éditeur tel que Notepad++, le contenu est affiché de la manière suivante :

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 902.136
tiers? <exists>
size = 5
item []:
  item [1]:
    class = "IntervalTier"
    name = "phones"
    xmin = 0
    xmax = 902.136
    intervals: size = 270
    intervals [1]:
      xmin = 0
      xmax = 49.08195191749949
      text = ""
    intervals [2]:
      xmin = 49.08195191749949
      xmax = 49.116951917499485
      text = "_"
    intervals [3]:
      [...]
```

Chaque item correspond à une tire de transcription : la première est la tire « phones », dans laquelle chaque phonème apparaît isolément, c'est le résultat de la dernière étape d'alignement. Cette tire pourrait être utilisée, mais la seconde tire, celle des syllabes alignées, est plus pratique car elle regroupe plusieurs phonèmes par segment (elle est donc plus courte dans le textgrid, car il y a moins d'*intervals* (segments)). Il suffit donc d'isoler le contenu de l'item [2] puis isoler le contenu textuel de chaque segment, puisque c'est lui qui nous intéresse. Le texte contenu dans chaque segment est signalé par `text = "X"` et le texte est affiché à la place de X. Pour isoler ce contenu on peut utiliser des expressions régulières. Ces expressions sont des chaînes de caractères suivant une syntaxe bien définie, permettant de faire des recherches précises dans un texte. Les deux expressions qui vont nous intéresser ici seront les suivantes : `(["^"]+)(["^"]*)` et `$2\n`, la première va sélectionner

automatiquement tout ce qui est contenu entre des guillemets, et la deuxième va supprimer tout le reste. On se retrouve alors qu'avec les chaînes de caractères contenues dans les segments de la tire. On peut éventuellement rechercher toutes les chaînes " \_ " pour les supprimer (il s'agit des segments vides), puis supprimer toutes les guillemets pour avoir seulement le texte de la tire. Il est alors très simple de reporter dans un tableau les phonèmes qui ont été alignés (cf. illustration 7.3). Cette manipulation peut paraître fastidieuse, mais elle permet un gain de temps particulièrement appréciable.

The image shows two windows side-by-side. The left window is a spreadsheet titled 'Resultats\_alignements.ods - LibreOffice Calc'. It contains a table with words in the first column and counts for phonemes N1 through N7 in the subsequent columns. The right window is a text editor titled 'D:\Personnel\Memoire\PEAA1\_suit\_aventure\traitement\_donnees\T1N7.txt - Notepad++'. It contains a list of phonemes and their corresponding words, with some words crossed out.

	N1	N2	N4	N5	N6	N7
gR*@nObl*						
g@nOb		2				
gR@nOb					1	
g@nobl						1
gR@nObl	1		1	2		
anglais	N1	N2	N4	N5	N6	N7
a~g@*IE						
a~gIE	1		2	3		
a~g@IE		1	1		1	
université	N1	N2	N4	N5	N6	N7
ynivER*site						
ynivEsite	1	3	2		1	1
ynivERsite	1	1		2		1
stendhal	N1	N2	N4	N5	N6	N7
ste~dal*						
ste~da	1	2	1	1		
ste~dal	1	1			2	
classes	N1	N2	N4	N5	N6	N7
k@*las						
klas	1	1	2	1	1	
k@las	1	1				
France	N1	N2	N4	N5	N6	N7

The text editor window shows the following content:

```

2 Intervallier
3 gull
4
5
6 di
7 zHi
8 ta~
9
10
11
12 gê
13 nObl
14
15
16 pa~
17 da~
18 tRa
19
20 za~
21
22
23
24 y
25 di
26 vE
27 si
28 ce

```

FIG. 7.3 : Traitement des données

Afin de savoir comment est aligné chaque mot cible pour chaque locuteur, et dans l'objectif de présenter les informations de manière à obtenir facilement des statistiques, tous les alignements possibles de chaque item seront inscrits dans la première colonne, et il n'y aura plus qu'à compter le nombre d'occurrences de chaque alignement pour chaque locuteur. Ainsi le mot « Grenoble » (transcrit  $gR^*@nObl^*$ ) pourra être aligné de  $2^2$  manières différentes puisqu'il a deux phonèmes optionnels :  $g@nOb$ ,  $gR@nOb$ ,  $g@nobl$  ou  $gR@nobl$ . Toutes les occurrences du mot cible sont traitées. Il arrive que l'alignement soit toujours le même pour toutes les occurrences chez la même personne, ou qu'il y ait des variations.

Dans le cas des mots « trente-quatre » ( $tR^*a-t^*kj^*atR^*$ ) et « soixante-dix-huit » ( $swasa-t^*di z^*Z^*Hit$ ), on se retrouve avec  $2^4$  (16) et  $2^3$  (8) possibilités d'alignement. Pour limiter la taille du tableau je n'écrirai que les occurrences attestées au moins une fois pour un locuteur.

Afin d'identifier la stratégie de simplification utilisée par chaque locuteur et pour les différents mots, trois colonnes supplémentaires seront insérées à gauche des tableaux de données. Dans la première sera indiqué le mécanisme de réalisation de chaque occurrence, dans la deuxième la position du phonème concerné et dans la troisième le phonème en question. Si un phonème doit être aligné mais ne l'est pas, il s'agira d'une disparition ; si un phonème est ajouté alors qu'il ne le devrait, il s'agira d'une insertion, et enfin si un phonème est remplacé par un autre (deux options côte à côte, parmi lesquelles une seule doit être alignée), il s'agira alors d'une substitution. Lorsque l'occurrence est correctement réalisée, un « ok » sera indiqué dans la première colonne, et les autres seront laissées vides. Il est important d'ajouter que, pour quantifier les erreurs, il est nécessaire de dédoubler les mots contenant plusieurs options. Le mot « dix-neuf » par exemple en comporte deux ( $diz^*nɔf^*$ ), s'il est aligné  $di nɔ$  dans cinq occurrences, il sera compté cinq disparitions de /z/ et cinq disparition de /f/; soit dix réalisations incorrectes.

Nous obtenons finalement deux tableaux présentant le nombre d'alignements de chaque mot cible, en fonction des réalisations et des locuteurs, dans les enregistrements avant et après la formation intensive.

## 7.2 Analyse des productions à T<sub>1</sub>

Dans l'ensemble des productions à T<sub>1</sub>, 229 occurrences des 16 mots cibles ont été alignées. Parmi elles, seules 89 occurrences sont correctement prononcées, soit 39% seulement. Quatre mécanismes de réalisations se distinguent : 1. disparition d'une consonne ; 2. insertion d'une voyelle d'appui entre deux consonnes, ou d'un glide ou d'une consonne entre une consonne et une voyelle ; 3. substitution d'un phonème par un autre ; et enfin 4. réalisation correcte. Les cas de substitution ne concernent que le phonème /z/ (substitué par Z) dans les mots « dix-huit » et « soixante-dix-



huit », ils seront donc traités séparément<sup>3</sup>. L'insertion d'une consonne concerne le mot « quatorze », qui a souvent été réalisé *katROz* ou *katRORz*. Des pointages en T<sub>2</sub> montreront qu'il s'agit plutôt d'un problème lexical et il sera aussi analysé de manière indépendante.

Le nombre d'occurrences des mots cibles intégrées aux analyses quantitatives s'élève donc à 198. Parmi elles, 76 seulement sont correctement réalisées (38%). Dans les productions erronées (122 occurrences, 136 phonèmes concernés), sont observées 123 disparitions de consonne et 13 insertions de voyelle ou glide. Le tableau 7.2 présente le nombre d'observations de chaque mécanisme en fonction de la position du phonème en option dans la syllabe. Le phonème observé peut être la première consonne d'un cluster (C<sub>1</sub>), la deuxième (C<sub>2</sub>) ou la troisième (C<sub>3</sub>); ou une voyelle d'appui entre deux consonnes (C<sub>1</sub>\_ C<sub>2</sub>), ou encore un glide *j* entre une consonne et une voyelle (C<sub>1</sub>\_ V).

Somme - Total	Position						
Mécanisme	C <sub>1</sub>	C <sub>1</sub> _C <sub>2</sub>	C <sub>1</sub> _V	C <sub>2</sub>	C <sub>3</sub>	coda	Total Résultat
disparition	17			20	24	62	123
insertion		6	7				13
<b>Total Résultat</b>	<b>17</b>	<b>6</b>	<b>7</b>	<b>20</b>	<b>24</b>	<b>62</b>	<b>136</b>

TAB. 7.2 : Nombre de disparitions et d'insertions en fonction de la position du phonème

Les disparitions de phonème ont principalement lieu en coda (64 occurrences), les consonnes en C<sub>3</sub> ont plus tendance à disparaître que celle en C<sub>2</sub>, qui elles aussi disparaissent plus souvent que les consonnes en début de cluster (respectivement 24, 20 et 17 disparitions). Si l'on croise maintenant les phonèmes dont la réalisation est incorrecte, avec les positions de ces derniers dans la syllabe (*cf.* tableau 7.3), on peut quantifier le nombre d'erreurs par position pour chaque phonème. Il apparaît par exemple que /ʒ/ est presque autant ignoré en C<sub>2</sub> (20), qu'en coda (18); que /f/ l'est légèrement plus en C<sub>1</sub> (17), qu'en coda (14).

Si l'on prend en compte le nombre de fois que chaque phonème en option est correctement réalisé<sup>4</sup>, on peut voir que 38 /ʒ/ disparaissent alors que 57 sont cor-

<sup>3</sup>Pour « soixante-dix-huit », seul la réalisation du phonème /t/ sera intégrée aux analyses quantitatives.

<sup>4</sup>Des réalisations correctes du phonème peuvent apparaître dans les mots jugés incorrects à cause

Somme - Total	Phonème								Total Résultat
Position	@	f	j	l	R	t	w	z	
C <sub>1</sub>		17							17
C <sub>1</sub> _C <sub>2</sub>	6								6
C <sub>1</sub> _V			7						7
C <sub>2</sub>					20				20
C <sub>3</sub>							24		24
coda		14		14	18	15		1	62
<b>Total Résultat</b>	<b>6</b>	<b>31</b>	<b>7</b>	<b>14</b>	<b>38</b>	<b>15</b>	<b>24</b>	<b>1</b>	<b>136</b>

TAB. 7.3 : Nombre d'erreurs en fonction du phonème et de sa position

rectement réalisés, 31 /f/ disparaissent contre 2 cas seulement où il est présent, 24 /w/ disparaissent contre 7 réalisations correctes, 15 /t/ disparaissent contre 4 réalisations correctes, 14 /l/ disparaissent contre 13 réalisations correctes et enfin un seul /z/ disparaît alors qu'il est réalisé dans 14 occurrences.

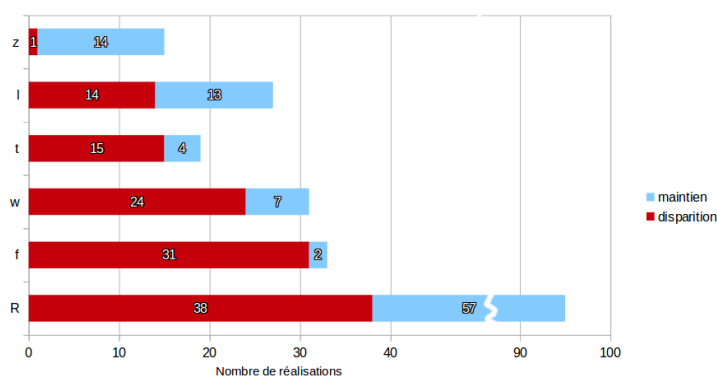


FIG. 7.4 : Nombre de disparitions et maintiens en fonction du phonème optionnel

Ces rapports entre réalisations correctes et incorrectes des mêmes phonèmes, permettent de constater que le /f/ n'est pratiquement jamais aligné, tandis que le /z/ l'est presque toujours. Excepté pour ce dernier phonème, on observe un grand nombre de disparitions ou insertions des phonèmes mis en option, ce qui montre qu'ils ont bien été posés aux endroits où il y avait des difficultés.

En ce qui concerne les insertions, elles ne concernent dans notre corpus que la voyelle @ (schwa) entre deux consonnes dans les mots « Alpes » et « anglais », et le glide j entre l'attaque et la voyelle de « quatre ». On constate dans le tableau 7.3 que d'autres options.

les deux phonèmes sont insérés pour 6 et 7 occurrences respectivement. C'est une difficulté qui ne concerne cependant que certains locuteurs. Afin de voir plus facilement qui est à l'origine de chaque réalisation, des graphiques ont été générés pour chaque mot cible, présentant en ordonnées les différentes réalisations, en abscisses le nombre d'occurrences ; et pour lesquels chaque couleur représente un locuteur. On peut ainsi voir d'un coup d'œil qui a prononcé quoi, comment et combien de fois<sup>5</sup>.

Le graphique 7.5 présente les alignements du mot anglais en fonction des 7 locuteurs. N'ayant qu'une seule option, il peut être aligné de deux manières : avec ou sans la voyelle d'appui. Il apparaît que N2, N4 et N6 l'ont insérée. On voit aussi que N5 l'a produit 3 fois sans épenthèse, et N4 deux fois. On a donc 3 réalisations du mots pour N4, dont une seule avec l'épenthèse. N2 et N6 en revanche ne l'ont prononcé qu'une fois, avec voyelle d'appui.

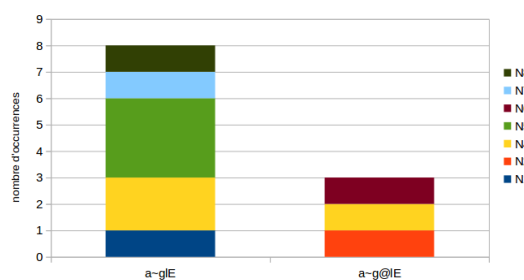


FIG. 7.5 : Alignements de « anglais » à T<sub>1</sub>

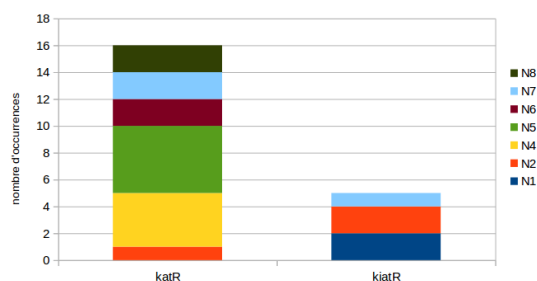


FIG. 7.6 : Alignements de « quatre » à T<sub>1</sub>

Pour l'insertion du glide, on peut voir sur le graphique 7.6 qu'il est réalisé par N1, N2 à nouveau et N7. N1 et N2 l'ont produit deux fois avec le glide, mais N2 l'a aussi produit une fois sans. N7 l'a quant à elle produit deux fois sans, et une seule fois avec. Il semble que N2 et N7 soient en cours d'acquisition de la réalisation correcte, N7 étant à un stade légèrement plus avancé que N2. N5 a prononcé 5 fois le mot sans glide, elle semble avoir totalement acquis sa réalisation.

Ces graphiques présentant les alignements en fonction des locuteurs, permettent d'identifier qui fait quelles erreurs, et si ces erreurs sont constantes ou mélangées parmi des réalisations correctes. Sur les graphiques de la figure 7.7, il apparaît que « soixante-sept » n'est prononcé avec un /t/ en première coda que par N7, tandis que

<sup>5</sup>Tous les graphiques sont présents en annexe.

tous les autres prononcent *swasa-sEt*. Dans le cas de « trente-quatre », chaque locuteur prononce d'une manière différente, mais le /t/ mis en option n'est réalisé que par N<sub>1</sub>, N<sub>5</sub> et N<sub>8</sub>. Le nombre limité d'occurrences ne nous permet pas ici d'observer de régularité.

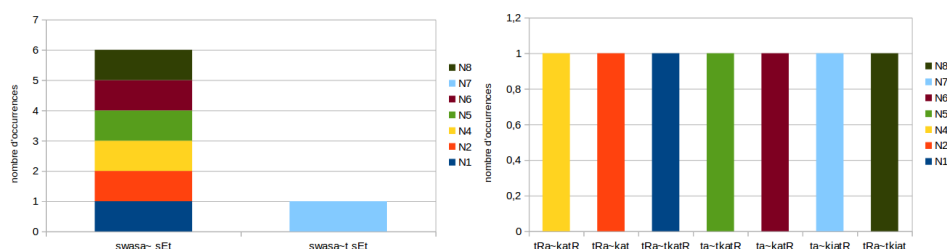


FIG. 7.7 : Alignements à T<sub>1</sub> de « soixante-sept » (gauche) et « trente-quatre » (droite)

Le mot cible « quatorze » présente un cas d'insertion d'une consonne (*R*) entre une consonne et une voyelle (*t* et *O*). Ce phénomène pourrait être dû à la difficulté dont est à l'origine la coda biconsonantique *Rz*, mais les pointages réalisés par les participants lors de l'entretien post-formation montrent qu'il s'agit plutôt d'un problème lexical. Probablement influencé par « quatre », le mot devient alors « quatorze », et il est pointé ainsi chez N<sub>1</sub> et N<sub>2</sub><sup>6</sup> (cf. chapitre VIII section 8). Les alignements du mot sont présentés dans le graphique 7.8. N<sub>2</sub>, N<sub>6</sub> et N<sub>7</sub> prononcent deux fois chacun *katROz*, N<sub>5</sub> le prononce une fois correctement et une fois *katRORz*, et N<sub>1</sub> le prononce tantôt *katOz*, tantôt *katRoRz*.

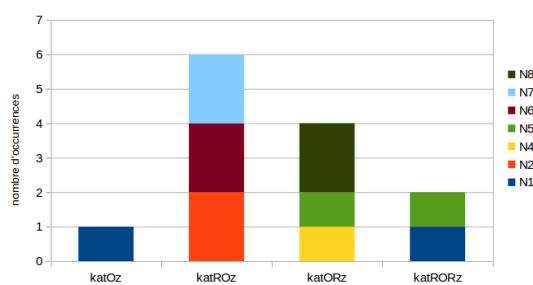


FIG. 7.8 : Alignements de « quatorze » à T<sub>1</sub>

Pour les mots « dix-huit » et « soixante-dix-huit », les graphiques de la figure 7.9 présentent une réalisation systématiquement chuintante du /z/ (en *Z*) pour les locuteurs N<sub>1</sub>, N<sub>2</sub> et N<sub>5</sub>. EasyAlign aligne parfois les deux phonèmes. Des analyses avec d'autres mots permettront d'observer les réalisations de /z/ et /zɥ/ dans d'autres contextes.

Grâce à la base de données qui a été constituée, il est possible par exemple de

<sup>6</sup>Alignements manquant pour N<sub>6</sub> et impossible pour N<sub>7</sub>.

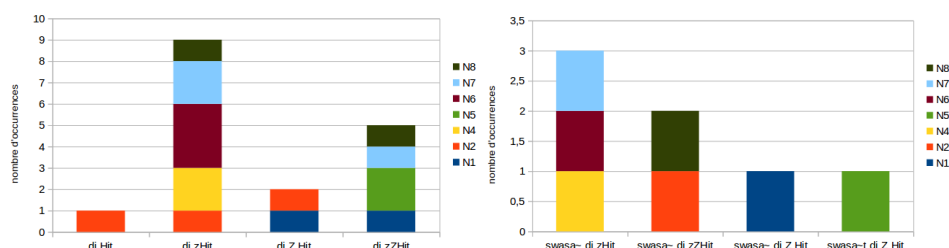


FIG. 7.9 : Alignements à T<sub>1</sub> de « dix-huit » (gauche) et « soixante-dix-huit » (droite)<sup>7</sup>

savoir quels sont les codas qui ont causé le plus de difficultés, ou encore combien d'erreurs en coda ont été commises par chaque locuteur. Le tableau 7.4 présente tous les problèmes détectés en coda. Dans notre corpus il ne s'agit que de disparition de consonne, mais différents phonèmes et mots cibles sont concernés. On peut alors savoir grâce à la colonne de total, le nombre d'alignements de chaque occurrence et chez qui l'alignement a lieu. On peut savoir grâce à la ligne de total le nombre de codas disparues pour chaque locuteur et voir quels mots et quels phonèmes ont posé problème pour chacun d'entre eux. Ici on peut voir que le *R* d'« université » a été très souvent supprimé par les locuteurs, de même pour le *f* de « dix-neuf ». On constate que N<sub>2</sub>, N<sub>7</sub> et N<sub>8</sub> sont ceux qui ont eu le plus de problème en coda (13, 11 et 13 disparitions), et N<sub>5</sub> beaucoup moins (4).

Si l'on se penche à présent sur la réalisation générale du *R*, quelle que soit sa position dans la syllabe, on observe les mêmes tendances : plus de difficultés pour N<sub>2</sub>, N<sub>7</sub> et N<sub>8</sub> (38%, 40% et 41% de réalisations incorrectes) mais beaucoup moins pour N<sub>5</sub> (7%), mais aussi très peu pour N<sub>1</sub> et N<sub>6</sub> (9%). Pour obtenir ces résultats, il a fallu tenir compte du nombre total de réalisations des mots contenant un *R* en option, produits par chaque locuteur. Le nombre total de réalisations correctes du *R* en option ( $N_{R.correct}$ ), y compris les occurrences traitées comme fausses à causes d'autres options, a ensuite été divisé par le nombre total d'occurrences de mots contenant un *R* en option par chaque locuteur ( $N_{R.total}$ ), puis multiplié par 100 pour obtenir un score de réalisation correcte en pourcentage.

$$Score = 100 \left( \frac{N_{R.correct}}{N_{R.total}} \right)$$

De la même manière, en comptant le nombre total d'alignements corrects par locuteur, et en divisant ce nombre par le total d'erreurs produites par chacun d'eux,

mécanisme	position	phoneme	mots	N <sub>1</sub>	N <sub>2</sub>	N <sub>4</sub>	N <sub>5</sub>	N <sub>6</sub>	N <sub>7</sub>	N <sub>8</sub>	Total
disparition	coda	l	g@nob		2					1	3
disparition	coda	l	gR@nOb					1	1	1	3
disparition	coda	R	ynivEsite	1	3	2		1	6	3	16
disparition	coda	l	ste-da	1	2	1	1			3	8
disparition	coda	z	di n9						1		1
disparition	coda	f	di n9						1		1
disparition	coda	f	di zn9	3	2	1	2	2	1	2	13
disparition	coda	z	di n9f								0
disparition	coda	t	tRa-katR			1					1
disparition	coda	R	tRa-kat		1						1
disparition	coda	t	tRa-kat		1						1
disparition	coda	t	ta-katR					1			1
disparition	coda	R	tRa-tkjat							1	1
disparition	coda	t	swasa-sÈt	1	1	1	1	1		1	6
disparition	coda	t	swasa-di zHit			1		1	1		3
disparition	coda	t	swasa-di zZHit		1					1	2
disparition	coda	t	swasa-di ZHit	1							1
			<b>Total</b>	<b>7</b>	<b>13</b>	<b>7</b>	<b>4</b>	<b>7</b>	<b>11</b>	<b>13</b>	<b>62</b>

TAB. 7.4 : Erreurs en coda par occurrence et par locuteur

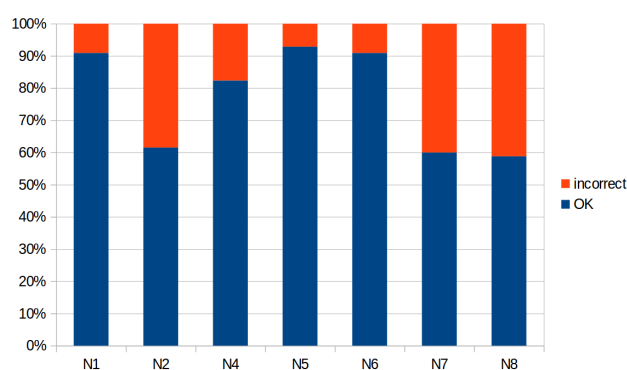


FIG. 7.10 : Score obtenu par locuteur pour la prononciation du R

	N1	N2	N4	N5	N6	N7	N8	Total
<b>Total occ. correctes</b>	10	8	14	13	10	12	10	77
<b>Total occ.</b>	11	13	17	14	11	20	17	103

TAB. 7.5 : Nombre de réalisations correctes et total de R\* en fonction des locuteurs

on peut envisager un score global de prononciation par participant (cf. figure 7.11)<sup>8</sup>. La pertinence d'évaluer globalement la prononciation pour chaque locuteur est discutée en fin de chapitre. On peut voir que le score global n'est pas proportionnel au nombre d'années d'apprentissage.

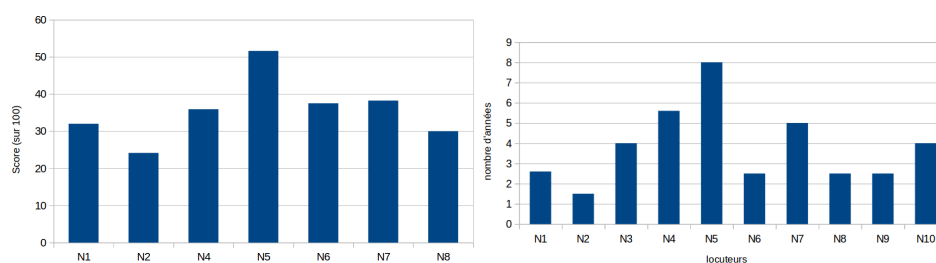


FIG. 7.11 : Score global de prononciation par locuteur (gauche) et nombre d'années d'apprentissage du français par locuteur (droite)

### 7.3 Comparaison des productions à T<sub>1</sub> & T<sub>2</sub>

Afin de comparer les productions de mots cibles à T<sub>1</sub> & T<sub>2</sub>, les mêmes traitements de données ont été réalisés sur les enregistrements à T<sub>2</sub>.

Un total de 95 occurrences des mots cibles a été aligné en T<sub>2</sub><sup>9</sup>. Parmi elles, 34 occurrences sont correctement prononcées, soit 36%. Les quatre heures de formation n'ont pas suffi pour observer une évolution significative dans la production des locuteur à moyen terme, et trop peu d'occurrences ont été enregistrées pendant l'entretien post-formation. Cette section va tout de même présenter les résultats à T<sub>2</sub>.

<sup>8</sup>Calculé ainsi, l'alignement *g@nOb* est plus pénalisant que *gR@nOb*, puisqu'on a deux erreurs dans le premiers et seulement une dans le second. Ce n'est donc pas seulement le nombre d'alignements corrects qui détermine le score.

<sup>9</sup>Par erreur de l'expérimentateur, les données sont incomplètes en T<sub>2</sub> : les nombres « trente-quatre », « soixante-sept » et « soixante-dix-huit » sont manquants pour les locuteurs N2, N4, N5 et N7, ainsi que « trois » et « treize » pour N6. Le mot « alpes » pour N5 n'a pas pu être aligné pour cause de bruits parasites.

Après avoir isolé à nouveau les occurrences de « quatorze » et de substitution du /z/ par Z dans « dix-huit » et « soixante-dix-huit », on observe 54 phénomènes de simplification. Une consonne disparaît dans 49 cas, et une voyelle est insérée dans 5 (cf. tableau 7.6).

Somme - Total	Position							
Mécanisme	C <sub>1</sub>	C <sub>1</sub> _ C <sub>2</sub>	C <sub>1</sub> _ V	C <sub>2</sub>	C <sub>3</sub>	coda		Total Résultat
disparition	8			13	3	25		49
insertion		o	5					5
<b>Total Résultat</b>	<b>8</b>	<b>o</b>	<b>5</b>	<b>13</b>	<b>3</b>	<b>25</b>		<b>54</b>

TAB. 7.6 : Nombre à T<sub>2</sub> de disparitions et d'insertions en fonction de la position du phonème

Cette fois-ci on observe qu'aucune voyelle d'appui n'a été alignée, mais cinq glides ont été détectés dans « quatre » et « trente-quatre ». Le /f/ disparaît toujours autant en C<sub>1</sub> (8) pour « France », qu'en coda (7) pour « dix-neuf ». On observe cependant moins de disparitions du R en coda (5) qu'en C<sub>2</sub> (13), alors qu'ils étaient autant sujets à disparaître en T<sub>1</sub> (18 et 20, respectivement).

Si l'on intègre l'ensemble des occurrences dans lesquelles les phonèmes en option sont maintenus, on constate que les proportions sont assez similaires avec les résultats à T<sub>1</sub>. Dans la figure 7.12, on retrouve un /z/ ne causant aucune difficulté (0 disparition, 7 maintiens); le /l/ n'est toujours présent que dans la moitié des occurrences (7 disparitions, 8 maintiens); le /f/ est constamment non aligné (15 disparitions, 0 maintien). Le /t/ et le /w/ sont quant à eux plus maintenus qu'à T<sub>1</sub> (6 disparitions et 6 maintiens du /t/ à T<sub>2</sub> pour 15 disparitions contre 4 maintiens seulement à T<sub>1</sub>; 3

Somme - Total	Phonème									
Position	@	f	j	l	R	t	w	z		Total Résultat
C <sub>1</sub>		8								8
C <sub>1</sub> _ C <sub>2</sub>		o								o
C <sub>1</sub> _ V			5							5
C <sub>2</sub>					13					13
C <sub>3</sub>							3			3
coda		7		7	5	6		o		25
<b>Total Résultat</b>	<b>o</b>	<b>15</b>	<b>5</b>	<b>7</b>	<b>18</b>	<b>6</b>	<b>3</b>	<b>o</b>		<b>54</b>

TAB. 7.7 : Nombre à T<sub>2</sub> d'erreurs en fonction du phonème et de sa position



disparitions et 3 maintiens de /w/ à T<sub>2</sub> pour 24 disparitions contre 7 maintiens à T<sub>1</sub>). R est cependant omis dans plus de 50% des cas (18 disparitions, 16 maintiens), bien plus qu'à T<sub>1</sub> (38 disparitions contre 57 maintiens).

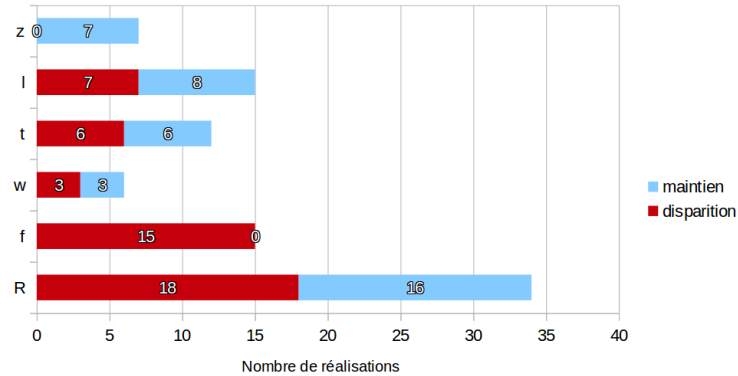


FIG. 7.12 : Nombre à T<sub>2</sub> de disparitions et maintiens en fonction du phonème optionnel

La comparaison des graphiques d'alignement à T<sub>1</sub> et T<sub>2</sub> présente plusieurs cas d'alignements relativement similaires (par les mêmes locuteurs), mais aussi quelques cas inexplicables de variation. Pour « quatre », on constate que N<sub>1</sub>, N<sub>2</sub> et N<sub>7</sub> palatalisent systématiquement le /k/, et N<sub>6</sub> le palatalise seulement à T<sub>2</sub>. N<sub>4</sub>, N<sub>5</sub> et N<sub>8</sub> quant à eux n'insèrent jamais le glide (cf. figure 7.13). Dans le cas de « université », c'est un peu moins évident. N<sub>2</sub>, N<sub>4</sub> et N<sub>8</sub> font constamment disparaître le R, mais les autres locuteurs tantôt le maintiennent, tantôt l'omettent. On constate cependant que quelque soit la période d'observation, le R est absent dans la majorité des occurrences (cf. figure 7.14).

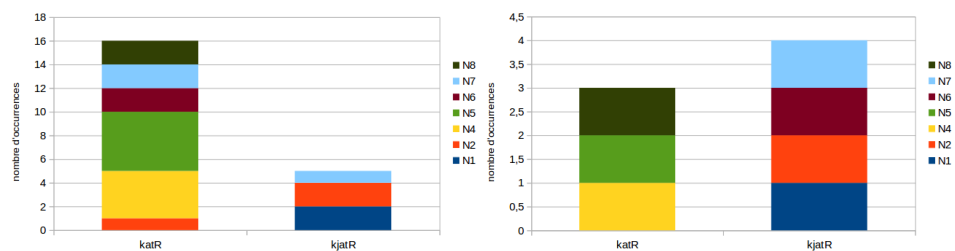
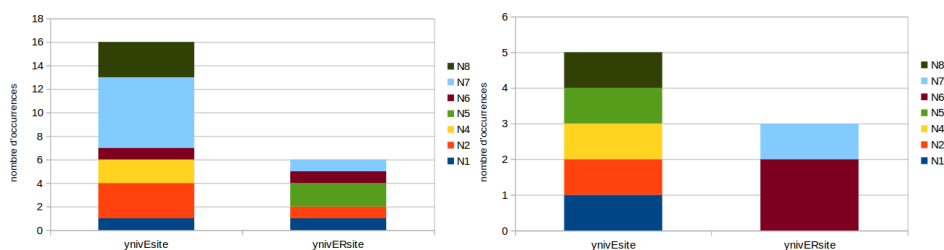


FIG. 7.13 : Alignements de « quatre » à T<sub>1</sub> (gauche) et T<sub>2</sub> (droite)

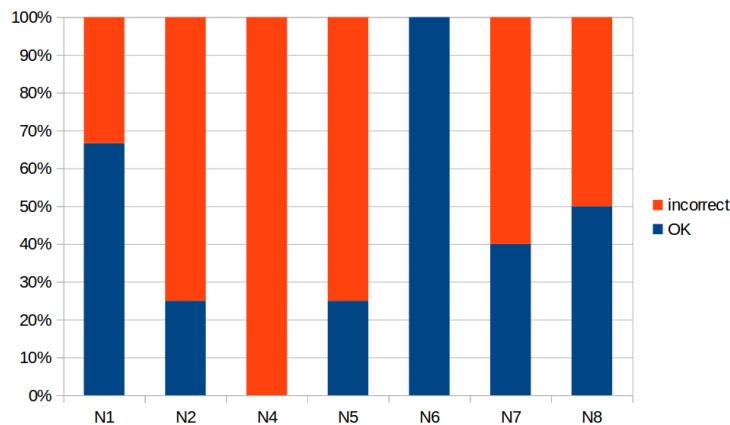
Quant à la réalisation de R, toutes positions confondues, il est maintenu 16 fois sur 34, dont 5 fois sur 5 par N<sub>6</sub> et 4 fois sur 6 par N<sub>1</sub>. N<sub>8</sub> maintient la consonne

FIG. 7.14 : Alignements de « université » à  $T_1$  (gauche) et  $T_2$  (droite)

	N1	N2	N4	N5	N6	N7	N8	TOTAL
<b>Total occ. correctes</b>	4	1	0	1	5	2	3	<b>16</b>
<b>Total occ.</b>	6	4	4	4	5	5	6	<b>34</b>

TAB. 7.8 : Nombre à  $T_2$  de réalisations correctes et total de  $R^*$  en fonction des locuteurs

dans 3 occurrences sur 6 ; pour les autres locuteurs la consonne disparaît plus souvent qu'elle est maintenue (cf. tableau 7.8 et figure 7.15).

FIG. 7.15 : Score obtenu à  $T_2$  par les locuteurs pour la prononciation du R

Enfin, le score global de prononciation par locuteur n'est pas significativement différent entre  $T_1$  et  $T_2$ . Le score à  $T_2$  est tout de même supérieur à  $T_1$  pour tous les locuteurs exceptés  $N_4$  et  $N_5$  (cf. figure 7.16). Lors d'une expérience sur une plus longue durée, et avec plus d'occurrences enregistrées à  $T_2$ , une évolution plus nette devrait être observable.

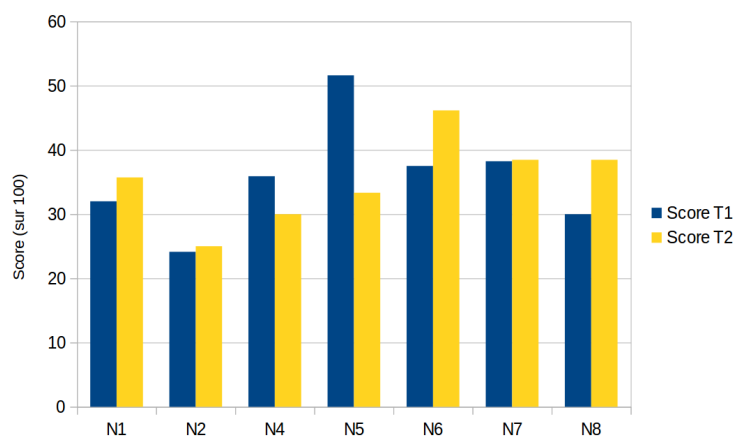


FIG. 7.16 : Score global de prononciation à T<sub>1</sub> et T<sub>2</sub> par locuteur

## 7.4 Discussion

Il est probablement impossible de s'affranchir totalement de l'analyse manuelle des données, mais l'automatisation que permet les technologies actuelles est sans nul doute la clé vers l'analyse à grande échelle. Nous avons pu voir ici que, utilisé avec modération, et adapté à la réalité de la population étudiée, l'alignement automatique des phonèmes permet d'observer des tendances générales de prononciation, de mettre en évidence les formes réellement problématiques et d'évaluer les compétences des participants. Il reste encore à mesurer l'évolution de ces tendances sur une durée plus importante, afin de pouvoir quantifier les progrès réels des apprenants.

Les analyses fines ont permis de constater que l'élision est vraisemblablement la stratégie de simplification la plus fréquente chez ces locuteurs. L'analyse quantitative automatique a quant à elle permis de quantifier la fréquence d'utilisation de cette stratégie, ainsi que celle d'insertion de voyelle et celle de substitution. Ce chapitre a présenté comment traiter les données enregistrées lors des entretiens afin d'observer les tendances de réalisation générales et par locuteur ; en fonction des phonèmes concernés et de leur position dans la syllabe.

Ce chapitre a aussi proposé des perspectives d'évaluation et de correction automatique. Grâce aux résultats détaillés obtenus, il est possible de donner un *feedback* approprié à chaque locuteur en fonction de sa prononciation. La présente étude a permis le calcul de "scores" de prononciation de certains phonèmes par participant, et de scores globaux de prononciation. La perspective de "noter" les apprenants n'est

pas notre objectif ici, mais les résultats obtenus peuvent permettre de discriminer les locuteurs et de les positionner pour adapter leur formation ; ou encore d'observer une évolution globale de leur prononciation dans le temps. Le score peut être considéré comme source de motivation pour l'apprenant : il peut constater une évolution visible de sa prononciation, et a les moyens de se tester et évaluer sa prononciation en autonomie, sans être jugé par un enseignant ou un proche.

Si l'alignement et le traitement des données de cette étude sont totalement automatisés, il deviendra possible de proposer une évaluation/correction en temps réel de la prononciation des apprenants. Il serait tout à fait envisageable de concevoir une application permettant à l'utilisateur de s'enregistrer, et traitant automatiquement le signal pour finalement donner un *feedback* approprié. L'application serait alors comparable à celle de KAWAI et HIROSE (1997) sur le fonctionnement (alignement automatique), mais adaptée au public japonophone apprenant le français, et calibrée sur un grand nombre de difficultés phonétiques. Comme le remarquait DIGALAKIS et MOUSTROUFAS (2007), la principale difficulté est de concevoir un programme capable d'évaluer du discours spontané, sans faire lire un texte pré-analysé à l'utilisateur.

On pourrait imaginer une application dotée d'un outil de reconnaissance vocale, chargé d'analyser la parole de l'utilisateur, et d'identifier des mots cibles qui auraient été préalablement intégrés à la base de données, avec des phonèmes en options, par le concepteur. Une fois le discours analysé, le programme aligne alors automatiquement les mots compris dans sa base de données, et aligne ou non les phonèmes en option en fonction du signal. Pour chaque option aura été associée une description de l'erreur, avec le mécanisme de simplification, le phonème concerné et sa position ; et en fonction de la fréquence des erreurs et de leur type, le programme donne alors un *feedback* à l'utilisateur pour l'aider à conscientiser le problème et à le résoudre.



## Chapitre 8

### Entretiens avec les participants

À l'issue de la formation intensive organisée dans le cadre de cette étude, des entretiens individuels ont été réalisés pour obtenir un enregistrement T<sub>2</sub> des mots cibles à analyser. En cette occasion, une discussion ouverte avec le participant a aussi été proposée pour obtenir des informations plus précises sur son apprentissage du français ainsi que ses ressentis pendant la formation. Les cinq questions suivantes ont été abordées pendant la discussion, en français ou en japonais, selon la volonté du participant.

- Comment as-tu appris la prononciation du français jusqu'à maintenant ?
- Quelles difficultés tu as en ce qui concerne la prononciation du français ?
- À quel genre de formation tu t'attendais avant de venir ?
- Qu'as-tu ressenti pendant les premières minutes devant le tableau de rectangles ?
- Quels conseils aurais-tu à me donner si je veux faire un cours similaire au Japon ? (qu'est-ce qui t'a plu ou déplu pendant la formation ? Comment cela se passerait-il au Japon ?)

Après quoi il a été demandé au participant de refaire le tableau sur une feuille blanche, de mémoire, sans aucun indice. Puis pointer quelques mots après une brève revue du tableau original. Comme il apparaît dans le corps du chapitre, les avis et les ressentis sont divers et il est important de prendre du recul sur les commentaires avant toute généralisation. L'entretien se voulait dialogal et compréhensif : l'objectif étant d'avoir un aperçu des sentiments du participants sans l'influencer dans ses commentaires. Rien n'était donc justifié de ma part, ni contredit, mais je réagissais

aux commentaires de manière à faire progresser la conversation dans le sens du participant.

Les entretiens étaient intégralement enregistrés et ont été étudiés et comparés a posteriori, pour en faire le contenu du présent chapitre.

## 8.1 Apprentissage de la phonologie au Japon

En ce qui concerne notre population, il ressort globalement un sentiment de manque de formation sur l'oral et la prononciation du français, au Japon comme en France. N1 reconnaît ne jamais avoir appris à prononcer les mots du français jusqu'à son arrivée en France, car il y avait très peu de correction phonétique à l'université au Japon. Sur les 3h de cours hebdomadaires, elle dit ne jamais avoir eu d'entraînement à la prononciation, et pour passer l'entretien – en français – pour obtenir sa mobilité à l'étranger, elle raconte avoir rédigé un texte qu'elle a fait lire à une collègue française de son université en l'enregistrant, puis elle a appris par cœur l'enregistrement. N4 raconte avoir eu des cours de prononciation mélangés à la grammaire quand elle était au lycée. Lorsque les étudiants lisaient des textes, l'enseignant (japonais) les corrigeait parfois en leur demandant de répéter après lui. Il y avait aussi un cours de théâtre en français, avec un enseignant francophone natif, où de même il fallait répéter après lui. Cependant elle dit n'avoir eu aucune formation à l'université (en tant que non-spécialiste du français). Selon N8 et N9, on est habitué au Japon à transcrire les langues étrangères à l'aide du syllabaire katakana (*cf.* chapitre III), en particulier au collège et au lycée. Il n'est pas rare que les premières pages voire la totalité des manuels de langues ou des guides de conversation proposent la transcription en katakana en dessous de la phrase orthographiée en français. N9 a eu un cours spécifique à la phonétique française, mais centré sur la réalisation des nasales. Là encore il s'agissait de lire des textes et répéter la correction de l'enseignant lorsque c'est nécessaire. Cependant N9 dit aussi avoir étudié en cours le triangle vocalique. Quant à N6, n'ayant pas eu de cours spécifiques sur la prononciation, il a appris seul dans des livres de phonétique, il connaît la différence sourd/voisé, la variation d'aperture de la bouche pour les voyelles ou les différents lieux d'articulation. N8 a appris des éléments de phonétique dans la chorale qu'il fréquentait.

N10 explique que les enseignants au Japon ont tendance à abandonner rapide-

ment l'enseignement de la prononciation. Il y a déjà une certaine difficulté pour les non-natifs à l'enseigner, et les enseignants natifs qu'elle connaît sont souvent restés plusieurs dizaines d'années au Japon, connaissent bien les difficultés des Japonais, mais sont aussi habitués à leur prononciation et corrigent rarement les apprenants. De plus, il y a au Japon une habitude de ne pas parler beaucoup en classe, et elle ajoute qu'il faut vraiment se forcer pour répondre à une question. S'il faut faire en plus attention à notre prononciation, alors on préfère se taire. On a vraiment peur de se tromper ou de se faire corriger devant les autres. Pour N10, la grammaire et le vocabulaire sont plus importants que la prononciation. Lorsqu'elle était en première année d'université, elle raconte que ses cours de français étaient exclusivement des cours de grammaire et de vocabulaire : « on lisait seulement ce qui était écrit ». À cela s'ajoute souvent le problème du nombre d'étudiants dans la classe, entre 20 et 40 étudiants pour un seul enseignant. Et il y a peu d'occasions d'utiliser le français au Japon, c'est donc difficile de pratiquer ou vérifier la prononciation est de qualité : « personne n'est là pour nous aider ». N10 continue en ajoutant qu'il y a des gens qui peuvent bien prononcer, et d'autres qui ne peuvent pas. « La prononciation, ça fait ressortir des différences individuelles et personnelles qui discriminent, et qui sont mauvaises pour la classe ». Et comme les apprenants parlent peu, il y a peu de choses à corriger.

Quant à l'apprentissage en France, N1 dit avoir eu des cours de prononciation à Lyon, où il fallait discriminer des sons en disant si on entendait plutôt A ou B, notamment avec les nasales. N2 apprend le français dans une association à Grenoble, elle y apprend la prononciation à partir de l'alphabet, elle apprend à lire et écrire en même temps, et lit des mots et se fait corriger quand c'est nécessaire. Mais N2 précise que l'objectif de cette association est d'enseigner le minimum de survie : « qu'importe la prononciation, si on est intelligible, c'est suffisant ».

## 8.2 Difficultés perçues dans l'apprentissage de la phonologie du français

Lorsque l'on demande aux participants les difficultés particulières qu'ils ont ou qu'ils ont eu dans leur apprentissage de la prononciation du français, c'est d'abord



les voyelles ouvertes, le /u/, le /ø/ puis la distinction des consonnes /r/ et /l/, et les consonnes /ʃ/ /ʒ/. N2 explique que ce sont les sons où la langue ne touche pas le palais comme ʒ (par rapport à dʒ en japonais), ou ʁ (contre r en japonais) qui sont difficiles, ainsi que le l (contact plus long qu'en japonais). N2 reconnaît qu'il y a aussi une certaine honte à ouvrir la bouche en grand, exagérer la prononciation, même si on sait que c'est mieux pour notre apprentissage. L'apprentissage d'autres langues peut aussi aider, l'estonien, dans le cas de N2, pour le [u] par exemple, qui existe en estonien (elle prononce cependant une voyelle plus centrée et moins arrondie). Pour N10, le ʁ n'a pas été une difficulté. Elle s'amusait de voir que tous les enseignants au Japon leur disaient que cette consonne était la particularité du français et qu'elle constituait la difficulté principale de la phonologie française. En revanche, les paires /s/-/ʃ/, /z/-/ʒ/ et /u/-/ø/ lui posent encore problème. Elle ajoute « je sais que je prononce mal mais je ne sais pas ce qui va pas, je ne sais plus ce qui est correct ou incorrect. C'est trop fossilisé en moi ». N7 insiste sur le fait que la prononciation n'est pas importante, et que c'est d'abord la communication qui importe. Plusieurs locuteurs affirment que les autres participants sont plus forts qu'eux sur le plan de la prononciation, et ils semblent l'accepter comme un fait irrémédiable, ou comme une partie de leur identité langagière.

N10 dit avoir l'impression de mettre partout des voyelles qui n'existent pas, mais elle ne sait pas lesquels sont correctes ou incorrectes. Cependant, à l'exception de ce locuteur, personne ne mentionne la difficulté que pose les enchaînements de consonnes. Peut-être que les participants ont plutôt compris la question comme « quels sont les phonèmes problématiques ».

### 8.3 Retours sur les premiers ressentis devant le tableau

J'ai trouvé intéressant de demander aux participants de se remémorer les premiers sentiments qu'ils ont éprouvés devant le tableau, lors de la première séance de formation. Les réponses étaient diverses : « ça a l'air dur ! » (N9), « c'est bizarre ; je comprends rien » (N4), « ça a l'air intéressant, mais ça doit être un peu dur » (N1). N2, qui est enseignante, connaissait l'existence du Silent Way, elle y a pensé en voyant le tableau, mais ne savait pas ce que cela représentait. N1 se réjouissait de connaître les autres participants, cela la mettait plus en confiance devant cet outil inconnu.

On sent globalement une certaine appréhension due à la situation : une formation *extraordinaire* sur la prononciation du français, avec une classe de 10 personnes, se connaissant plus ou moins pour certains, avec des niveaux de français et des âges variés. L'enseignant leur était inconnu, la formation était filmée, et ils se trouvaient devant un tableau totalement inconnu et très abstrait à première vue. Cependant, notons qu'ils étaient curieux et motivés par l'opportunité de travailler leur prononciation avec une méthode différente de ce qu'ils avaient connu jusqu'alors.

## 8.4 Retours sur l'utilisation du tableau

En ce qui concerne l'utilisation du tableau phonologique, il y a eu beaucoup de commentaires positifs sur le fait de se concentrer sur la prononciation en faisant abstraction du reste (N9, N2). N9 reconnaît ne pas trop prendre le temps d'y penser quand elle parle, mais à cette occasion elle a pu faire l'effort de se poser des questions. N1, N2 et N9 ont apprécié le fait de ne pas écrire. N9 dit avoir tendance à lire des katakanas dans sa tête, même lorsque les mots sont orthographiés en français ; et le fait de ne pas passer par l'écrit évite de se remémorer intuitivement une prononciation « japonaise ». N2 reconnaît qu'il peut être intéressant de commencer l'apprentissage d'une langue par une telle activité, sans rentrer immédiatement dans un objectif communicatif. Elle apprécie notamment le fait que les sons soient décrits entre autre par leur position dans le tableau. N1 s'exclame : « enfin un cours où on ne fait pas qu'écrire ! ». N5 explique qu'il faudrait utiliser ce tableau au début du cours mais après quelques temps il est nécessaire d'utiliser les lettres de l'alphabet ou l'API « parce qu'il est important d'apprendre le système avec les lettres latines ».

N4 dit avoir enfin compris la différence entre ʒ et ʤ, et reconnaît aussi l'utilité de ce tableau pour travailler les nasales. Cependant elle ajoute que c'est assez difficile si on ne connaît pas les mots utilisés dans l'activité, et que si l'on prononce des mots que l'on ne connaît pas, ou qui n'existent pas, alors cela n'a pas de sens. Elle avoue avoir ressenti un vrai besoin de traduction, notamment avec le plan de la ville ou la carte de France.

Le tableau est intéressant, mais il y a des personnes qui peuvent ne pas aimer travailler avec cet outil, selon N2, N7, N9 et N10. Tout est très différent, il n'y a pas de support écrit pour se rassurer avec quelque chose de tangible. N7 précise

que chacun a sa propre façon de mémoriser. Elle préfère personnellement écrire et mémoriser le son avec sa ou ses formes graphiques, comme N<sub>5</sub> et N<sub>10</sub>. Elle dit que c'est comme ça qu'elle travaille depuis qu'elle est au collège, et elle se sent démunie, et mise dans une situation désagréable devant un tel tableau. Elle ajoute qu'elle veut se souvenir des sons de la langue par les mots. Par des mots qui existent et qui lui sont utiles. Elle s'énerve lorsqu'elle se rappelle le nombre de fois qu'elle a dû répéter le mot « clou » pour travailler le son [u], alors qu'elle ne savait ni ce que cela voulait dire, ni à quel point c'était un mot inutile pour elle. « C'est n'importe quoi ! », dit-elle à ce propos. « Moi je préfère me souvenir des règles de prononciation, et c'est comme ça que je fonctionne », ajoute-t-elle enfin. N<sub>1</sub> et N<sub>5</sub> avouent aussi qu'il y avait trop de répétition, et que leur motivation chutait à cause de cela.

## 8.5 Retours sur la formation et l'approche pédagogique

D'après N<sub>2</sub>, il est impératif de proposer ces activités une fois que tous les participants se connaissent. Il faut que l'atmosphère soit détendue, sinon il pourrait y avoir un blocage. Dans cette formation, elle dit que tout le monde se connaissait plus ou moins donc il n'y avait pas de problème. N<sub>1</sub> parle d'une bonne expérience, sans stress, car elle connaissait tous les participants, elle n'avait pas trop honte de prendre la parole. Au niveau du public cible, N<sub>4</sub> propose débutants-intermédiaires, mais la formation ne lui semble pas pertinente pour un public de niveau avancé. Pour N<sub>2</sub> cette formation est intéressante pour tous les niveaux, y compris les débutants complets. D'après elle, c'est aussi intéressant pour ceux qui ont un certain niveau, mais c'est plus dur pour eux de se remettre en question, surtout devant des apprenants de moins bon niveau. Pour N<sub>10</sub> en revanche, ce cours est intéressant quand on connaît déjà un peu la langue, et il n'est pas très adapté aux débutants complets. L'utilisation du plan de la ville et de la carte de France a été appréciée, N<sub>5</sub> se réjouit d'avoir travaillé sa culture générale. N<sub>9</sub> a trouvé ces supports un peu trop difficiles parce qu'elle ne connaissait pas bien le nom des villes et des régions de France, ni le vocabulaire détaillé relatif à la ville. N<sub>1</sub> aussi précise que le vocabulaire était difficile, d'après elle le plan de la ville ne marcherait pas bien au Japon car cela ne fait pas partie de la réalité des étudiants (ils ne sont pas en contact avec ces termes en français). Elle ajoute que le vocabulaire de la carte de France n'est pas un vocabulaire très familier, et qu'il serait plus intéressant de

travailler sur du vocabulaire que l'on utilise tous les jours. Elle conclue cependant que le fait de travailler sur des images et non du texte est vraiment intéressant et original (mais il se peut que des étudiants japonais soient perturbés par cette originalité). Pour N7, cette formation n'a pas d'intérêt car elle n'est pas communicative. De plus elle relève une asymétrie qui lui déplait particulièrement en tant qu'enseignante de japonais, il s'agit du fossé hiérarchique qu'elle perçoit entre le professeur et la classe. Elle qualifie même l'enseignant de « dictateur », qui ne fait qu'(or)donner la parole aux participants, lui, étant seul juge de la qualité de la prononciation des locuteurs. Le fait de devoir répéter tant qu'une version acceptable n'est pas prononcée est stressant et démotivant. N7 ajoute qu'un fossé se creuse aussi entre les participants qui prononcent bien, et ceux qui ont plus de difficultés, car c'est toujours les mêmes qui doivent répéter, et c'est toujours les mêmes qui sont pris pour exemple. Elle dit même avoir ressenti une stigmatisation de l'un des participants. Ceci étant dit, la majorité des participants m'ont dit qu'ils avaient vraiment apprécié cette petite formation, certains m'ont même demandé s'il était possible de continuer.

## 8.6 Retours sur le groupe

Pour ce qui est de la classe d'apprenants pour cette formation intensive, elle a été qualifiée de dynamique et agréable. Le fait d'être avec des participants de même culture, se connaissant déjà plus ou moins était décidément un paramètre non négligeable. N2 ajoute que la présence de N5 permettait au groupe de bien avancer, car elle donnait souvent la bonne réponse quand tout le monde était perdu. « ah mais oui ! C'est ça ! », pensait-elle souvent après avoir entendu la proposition de N5 pendant le cours. Elle précise que c'est plus facile de parler quand quelqu'un a déjà pris la parole. Mais peu après ce commentaire N2 avoue que le cours avançait trop vite pour elle, peut-être parce que les niveaux étaient trop hétérogènes. Elle explique : « quand je n'y arrive pas, je veux essayer encore, mais le professeur passe rapidement à la suite et on ne peut pas s'entraîner, à chaque fois c'est comme ça et du coup la motivation baisse ».

Si l'on écoute maintenant les commentaires de N5, elle dit effectivement que c'était un peu trop facile pour elle, et que la progression était globalement lente. Elle ajoute : « si tu fais ça au Japon, au bout d'un moment on en a marre ». Plusieurs participants

m'ont effectivement alerté du fait que cette formation ne serait certainement pas vécue de la même manière au Japon. N2 et N10 expliquent que les Japonais ne parlent pas trop en cours, ils ne sont pas habitués à beaucoup participer, et qu'une formation de ce genre risquerait de ne pas bien fonctionner à cause de cela. Ici, tous les participants vivent en France depuis un certain temps, ils sont un peu francisés, et ils n'ont plus tout à fait le même caractère que les Japonais du Japon, selon N2. N1 ajoute qu'ici, les participants sont plus motivés qu'au Japon, ce ne sont pas les mêmes conditions d'apprentissage, et pas le même rapport à la langue.

N10 dit qu'il est intéressant d'avoir une telle formation quand on est jeune, mais à partir d'un certain âge (25-30 ans selon elle), il vaut mieux aborder d'autres aspects de la langue en parallèle, ou en même temps, et éviter d'isoler la prononciation pour ne travailler que ça. Elle ajoute qu'il y a des gens qui sont à l'aise à l'oral, et d'autres qui ne le sont pas. D'après elle, ceux qui ne le sont pas ne le deviennent pas non plus. Et pour ce genre d'apprenants, ce cours n'est pas très intéressant. Elle conclue enfin que la difficulté de prise de parole en classe est plus un problème culturel qu'un problème linguistique ; on apprend pas de la même manière en France et au Japon.

## 8.7 Ancrage du tableau dans la mémoire à moyen terme

Il a ensuite été demandé à chacun des participants de dessiner de mémoire le tableau phonologique, sans les couleurs, mais en indiquant à quoi correspond chaque rectangle dessiné. L'objectif était d'observer ce qui avait été bien et moins bien mémorisé, et quelles traces le tableau avait laissées en mémoire après 4 séances de cours. La transcription était libre, je proposais d'écrire la ou les premières lettres qui leur passaient par la tête pour écrire le son en question, ou écrire un mot en exemple. Lorsque le participant dit avoir terminé, je lui demande de prononcer chaque son – je note alors la transcription au crayon à côté du rectangle, si la prononciation diffère de ce qui est écrit ou si aucun caractère n'a été trouvé pour transcrire le phonème. Jusque là aucun indice n'est donné, et le tableau est tel qu'il apparaît dans la mémoire du participant. Les tableaux sont placés en annexe de ce mémoire.

On observe que sur les 8 participants ayant dessiné le tableau, 4 s'en souviennent parfaitement (N1, N5, N8 et N9), 2 s'en souviennent bien mais oublient plusieurs phonèmes (N2 et N6) et les deux autres s'en souviennent très peu (seulement quelques

phonèmes présents pour N4 et N7). La qualité de mémorisation ne semble pas liée au niveau de langue, ni à la qualité de prononciation en T<sub>1</sub>, ni encore à l'appréciation du cours.

Si N7 était particulièrement réticente à l'approche de la formation, et n'a dessiné qu'un tableau très vague dans lequel il manque la plupart des éléments, N4, qui semble avoir bien plus apprécié le cours, elle aussi se souvient de très peu d'éléments du tableau. N5 est le participant avec l'expérience d'apprentissage du français et la durée de séjour en France les plus longues, elle a effectivement dessiné un tableau très complet ; pourtant N1 et N8, dont l'apprentissage n'a duré que deux ans et demi, le séjour en pays francophone seulement six mois, et dont les difficultés de prononciation sont flagrantes en T<sub>1</sub>, ont eux aussi proposé un tableau pratiquement parfait.

Les voyelles sont globalement mieux restituées que les consonnes. Le triangle vocalique et la logique d'aperture et de profondeur (avant-arrière) sont assez bien restitués<sup>1</sup>, excepté pour N2, N4 et N6. Pour ces derniers la disposition des phonèmes ne reflète pas la réalité articulatoire.

Si l'on regroupe le *e* muet et *æ*, et si l'on ne tient pas compte du *a* postérieur, des voyelles nasales et du *ŋ*, qui n'ont été que très brièvement abordés en cours, et si l'on compte seulement les rectangles dont le son est retenu par le participant, on constate que N8 n'a oublié aucun phonème, N9 et N5 en ont oublié qu'un (*ø* et *æ*), N1 et N2 en ont oublié deux (*k*, *g* et *æ*, *g*), N6 en a oublié 7 (*æ*, *o*, *ɔ*, *p*, *b*, *f* et *ʒ*), N4 en a oublié 10 et N7 en a oublié 11<sup>2</sup>. La figure 8.1 présente les phonèmes colorés en fonction de leur mémorisation par les participants. Plus le phonème est foncé, plus les participants sont nombreux à s'en être rappelé. Le nombre de cercles sous chaque phonème indique le nombre de participants ayant correctement placé le phonème par rapport aux autres. Les voyelles nasales ont été placées entre les voyelles et les consonnes, mais dans un ordre aléatoire.

On constate que les cinq voyelles *i e a u y* sont constamment dessinées, toutes à leur position initiale, sauf deux inversions de *i* et *e* (N2, N9). Le *o* a été oublié par N6, ce qui est assez étrange puisque tous les autres participants ont constamment pensé aux 5 voyelles du japonais ; le *y* est quant à lui toujours présent. Pour les consonnes on trouve sur tous les tableaux les phonèmes *t f v s* et *l*, puis *m n d z ʁ* pour 7

<sup>1</sup>Le fait que la disposition des rectangles soit respectée ne signifie pas qu'elle est comprise.

<sup>2</sup>Respectivement *ε*, *ɔ*, *ø*, *æ*, *d*, *v*, *k*, *g*, *f*, *ʒ* et *ɔ*, *æ*, *m*, *n*, *p*, *b*, *z*, *ʒ*, *k*, *g*, *ʁ*.

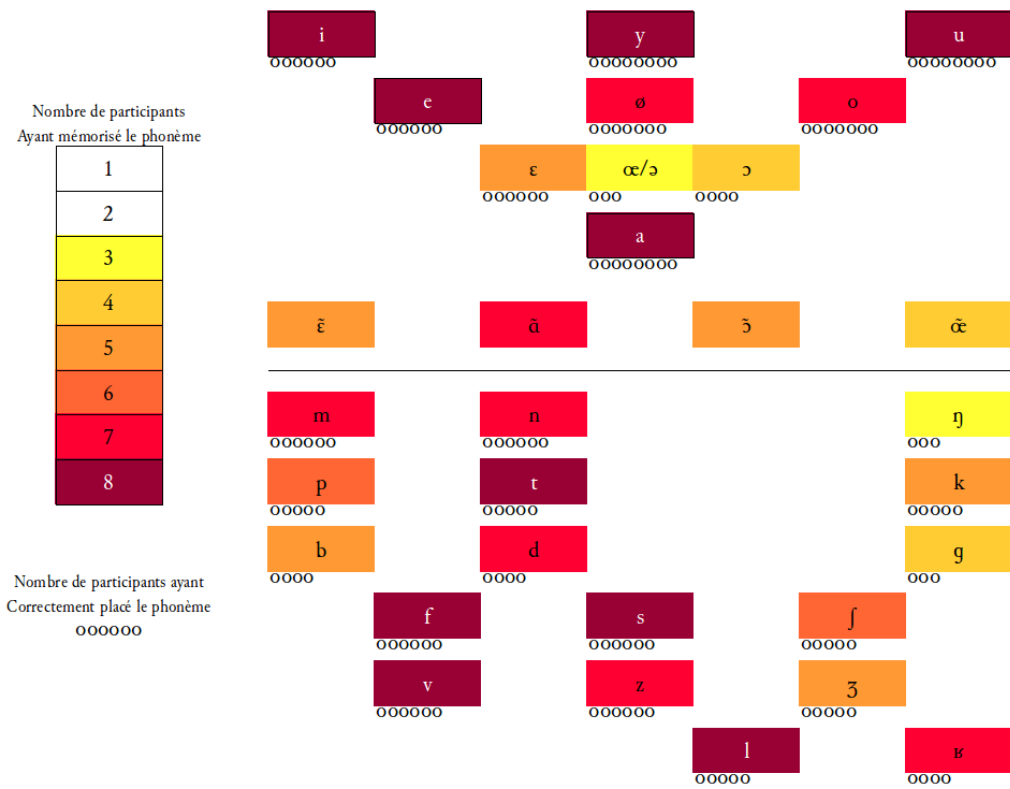


FIG. 8.1 : Mémorisation du tableau par les participants

participants sur 8. Tous les participants excepté N7 ont placé des voyelles nasales, 3 à 4 selon les locuteurs, avec  $\tilde{a}$  dans tous les cas, puis  $\tilde{\epsilon}$  et  $\tilde{\imath}$ , et enfin  $\tilde{\alpha}$  dans quatre tableaux. Plusieurs paramètres nous empêchent pourtant de penser que les phonèmes les plus présents sont les plus acquis : d'une part leur degré d'utilisation pendant le cours (les consonnes  $k$  et  $g$  ont été peu travaillées, car vraisemblablement moins problématiques que les autres), mais encore leur fréquence d'apparition en français ( $\tilde{\alpha}$  ou  $\eta$  sont plus rares, et sont facilement oubliés). Cependant, on peut penser que le  $\alpha$  (ou le  $\text{ə}$ ) sont une réelle difficulté phonétique pour nos locuteurs qui peinent encore parfois à différencier le  $\emptyset$  du  $u$  (cf. 8.2).

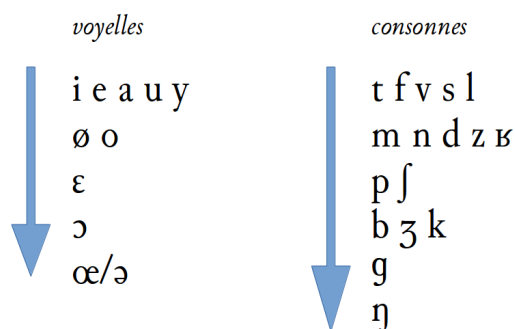


FIG. 8.2 : Fréquences d'apparition des phonèmes dans les tableaux (du plus au moins fréquent)

Le tableau semble donc globalement bien mémorisé. Cependant, est-ce la mémoire visuelle de certains qui leur permet de dessiner le tableau, ou leur conscientisation du rapport entre les phonèmes ?

J'ai demandé à chaque locuteur si la disposition des phonèmes leur semblait logique. N1 explique que, de ce qu'elle a compris pendant la formation, plus on descend dans le tableau des voyelles, plus la bouche s'ouvre, quand on va vers la gauche, les lèvres sont étirées, quand on va vers la droite les lèvres s'allongent vers l'avant. Pour les consonnes, N1 comprend le rapport sourde/sonore, mais ne relève pas spécialement de logique gauche-droite. N8 fait une comparaison entre les paires consonantiques sourde/sonore avec le trait de voisement du japonais (le *dakuten*  $\grave{\text{}}$ , /ka/ ㄎ → /ga/ ㄎ $\grave{\text{}}$ ). D'après N10, il semblerait que les japonais ne font pas ce lien en général : le dakuten est pour eux seulement un diacritique qui modifie la consonne, le fait que la différence



ne vient que de la vibration des cordes vocales n'est pas une évidence<sup>3</sup>. Selon N9, les consonnes qui se ressemblent sont proches, mais en ce qui concerne le rapport entre les voyelles, c'est plus difficile. Elle dit ne pas avoir compris jusqu'à une discussion des participants dans le tramway, après la première séance de cours. À ce moment-là, N8 aurait expliqué à plusieurs personnes ce que représentait la disposition des voyelles du tableau. N4 dit aussi ne pas avoir deviné le sens de cette disposition avant que N8 ne lui explique. D'après N6, le tableau des voyelles est plus difficile que celui des consonnes mais il comprend mieux la logique de la disposition pour les voyelles.

Plusieurs participants disent ne pas avoir bien compris la logique des couleurs du tableau, et n'avoir retenu les phonèmes que par la position des rectangles dans le tableau.

Enfin, en analysant de plus près les tableaux et les enregistrements, on constate que /y/ est souvent transcrit *yu* ou *iu* (N2, N6, N9) et/ou prononcée [jy] (N1, N2, N9). Les voyelles /u/, /ø/ et /œ/ sont prononcées [u] par N2 et N4. Lors des révisions du tableau original, N6 prononce [ɕʒ] pour la fricative palato-alvéolaire, puis puis se corrige ([ʒ]) et enfin donne un exemple « je » qu'il prononce [ɕʒə]. On remarque aussi que N4 transcrit ce qu'elle prononce [v] avec la lettre b.

## 8.8 Remarques concernant les pointages de mots

En fin d'entretien, chaque locuteur a été amené à pointer les nombres suivants : quatre, quatorze, dix-neuf et trente-quatre. Ces nombres ont été choisis pour éclairer les observations faites en T<sub>1</sub>, concernant la palatalisation du /k/ devant /a/ (EasyAlign a aligné plusieurs *kjatR*), la position de l'uvulaire dans « quatorze » (plusieurs *katROz* alignés), le /f/ en coda (presque toujours ignoré par EasyAlign) et enfin le cluster /tk/ de « trente-quatre ». Les pointages ont été effectués sur une copie du tableau original, révisé après le dessin de mémoire. N4 et N7 se souvenaient trop peu du tableau, et même après révision ne se sentaient pas capables de pointer. Pour des raisons externes, N9 n'a pas pointé non plus.

N6 a cherché comment pointer /kya/ pour « quatre » et « trente-quatre », il voulait

<sup>3</sup>On peut voir d'ailleurs dans certains mangas depuis les années 90, des voyelles (a, i, u et e) avec le dakuten pour exprimer un sentiment déplaisant. Elles sont lues généralement avec une laryngalisation (ou *creaky voice* en anglais : ʔ). Le /u/ + dakuten ne transcrit plus alors le nouveau phonème /v/.

un rectangle pour *y*, et a précisé « ce n'est pas /ka/ ». Il a essayé /kia/ puis /køa/ avant de pointer /ka/, peu convaincu. N<sub>2</sub> a cherché le <q>, influencée par l'orthographe, puis a pointé *k* suivi de *a*. Tous les autres ont pointé /ka/. La première syllabe de « quatorze » quant à elle a été pointé /ka/ à l'unanimité, comme s'il s'agissait d'un autre phonème. Cependant le mot a été effectivement pointé /katɔz/ par N<sub>1</sub> et N<sub>2</sub>, cela justifiant leur prononciation. La prononciation de N<sub>6</sub> reste inexpliquée (/katɔz/ pointé /katɔz/), ou s'agit-il d'une difficulté physique de prononciation ? Un *f* en coda a constamment été pointé pour « dix-neuf », appuyant l'hypothèse selon laquelle EasyAlign a plus de mal à détecter ce phonème, et qu'il s'agit donc bien d'un problème technique. Le *æ* de « dix-neuf » a cependant été pointé sur *u* (N<sub>1</sub>), *ø* (N<sub>4</sub>), *ə* (N<sub>5</sub>) et *ɛ* (N<sub>8</sub>). Les deux consonnes du cluster /tk/ de « trente-quatre » ont été pointées *tk* par tous les locuteurs sauf N<sub>1</sub> qui n'a pointé que le *k*. Les autres phonèmes pour l'ensemble des mots n'ont posé aucun problème, ni même le *ɸ*. Les clusters /tɸ/, /zn/ et /ɸz/ ont été correctement pointés.

## 8.9 Discussion

Cet entretien a permis de faire ressortir plusieurs points sur le ressenti des apprenants japonophones face à cet outil et à cette approche pédagogique. Tout d'abord on constate que la présence des autres et les relations interindividuelles, y compris avec l'enseignant, jouent un rôle fondamental pour le bon déroulement du cours. S'il s'agit d'un paramètre fréquemment pris en compte par les didacticiens et les enseignants, une difficulté supplémentaire apparaît ici pour ces apprenants qui doivent remettre en question leurs habitudes culturelles d'apprentissage et leur attitude en classe. Le statut de l'écriture dans l'apprentissage de chacun peut aussi constituer un obstacle à la présente approche, et si certains sont prêts à l'ignorer dans un premier temps pour se concentrer sur l'oral, d'autres se sentent démunis et incapables de tirer profit d'un tel enseignement. C'est un certain challenge à accepter de changer ses pratiques, et la difficulté qu'y éprouvent certains est tout à fait compréhensible. On sent une volonté générale de transcrire les sons d'une manière ou d'une autre pour mieux se les approprier, mais plusieurs apprenants sont aussi conscients du problème que cela peut engendrer, notamment en ce qui concerne la transcription en katakana d'une langue comme le français. On constate enfin des préjugés profondément ancrés en certains

apprenants en ce qui concerne les capacités à prononcer une langue étrangère.

Rappelons que ce que nous avons proposé là était un cours intensif sur la phonologie du français, où rien d'autre que la prononciation n'a été réellement abordé pendant les quatre séances. En temps normal il serait plus pertinent de proposer cette formation sur le long terme, de manière moins intensive et en parallèle à d'autres apprentissages relatifs à la langue pour diversifier les contenus de cours. De plus, le tableau utilisé dans cette étude est aussi et surtout un médiateur vers la l'orthographe de la langue étudiée. La couleur des phonèmes est notamment utilisée pour transcrire le son par colorisation des différentes graphies possibles (*cf.* chapitre 11). De nombreuses remarques ici appuient le fait qu'il est important d'éviter l'apprentissage intensif et isolé d'un seul aspect linguistique.

De plus nous avons dans cette étude un groupe d'apprenants particulièrement hétérogène, en âge, en occupation ou en compétences globales de français. Cette hétérogénéité était à l'origine d'une dynamique de progression ne permettant pas toujours aux apprenants moins expérimentés d'assimiler les traits phonétiques étudiés. On peut réfléchir à la pertinence d'une telle activité pour un groupe aux difficultés de prononciation trop hétérogènes. La diversité des participants permet une progression dynamique grâce à une riche mémoire collective ; cependant, nous observons dans notre étude que le même locuteur est souvent pris pour exemple car sa prononciation est satisfaisante, et d'autres se retrouvent finalement à répéter ce qu'il dit. Il semblerait donc plus efficient de proposer cette activité à des groupes dont les difficultés sur le plan phonologique sont à peu près de même niveau.

En ce qui concerne plus spécifiquement la population japonophone, certains paramètres sont peut-être à adapter, comme la façon de donner la parole à l'apprenant (qui n'y est pas habitué), ou encore la manière de lui faire comprendre si sa prononciation est satisfaisante ou non (ne pas être trop direct de peur de le froisser ou le vexer), éviter autant que possible la stigmatisation car il n'est pas évident de savoir à quel degré la prennent réellement les apprenants.

Quels que soient les commentaires post-formation, on constate qu'il y a eu une très bonne participation de l'ensemble des apprenants pendant la formation, tout le monde était attentif et présent à l'activité proposée, et c'est ce que l'on peut espérer de meilleurs pour nos apprenants.

# Conclusion

Cette étude a su répondre à deux problématiques. La première est pédagogique, et concerne la difficulté d'enseigner la prononciation en langue étrangère sans se laisser influencer par l'écriture. La population cible de ce travail a été choisie pour l'éloignement phonologique de sa langue par rapport au français, et pour la complexité de son système graphique. Dix apprenants japonophones multi-niveaux ont donc été conviés à une formation intensive à la phonologie du français, à travers une approche didactique multimodale permettant d'appréhender la langue sans avoir recours à l'écriture, et facilitant la conscientisation des phonèmes, de leur indépendance et de leurs traits phonétiques, grâce à un tableau les présentant en rectangles de couleur disposés logiquement.

La seconde problématique est technique, et consiste à évaluer de manière automatique la prononciation des apprenants en privilégiant le discours libre et spontané. L'automatisation permet de réitérer l'expérience avec un groupe plus important d'individus, sans augmenter significativement la charge de travail dans le traitement des données. Le discours spontané permet d'éviter l'influence de la lecture dans l'oralisation d'un texte écrit, ou la répétition par mimétisme après l'écoute du mot ou de l'énoncé cible.

Le protocole mis en place a permis d'enregistrer une production spontanée de chaque participant, et de traiter les données pour quantifier les principales stratégies de simplification phonétiques employées par les locuteurs. Cette analyse permet d'une part d'évaluer la qualité de prononciation des apprenants et d'identifier précisément leurs difficultés, et d'autre part d'observer en détails l'évolution de celle-ci dans le temps. De nombreuses perspectives s'offrent à présent pour optimiser le système d'évaluation automatique, et notamment la possibilité d'en faire un outil de remédiation phonétique en temps réel.

Enfin, une réflexion qualitative a été menée avec les participants, quant à la pertinence de l'approche didactique suivie et des outils pédagogiques utilisés en cours, avec une population aux habitudes et à la culture d'apprentissage tout à fait différentes de celles présentes en Occident. Il en est ressorti qu'une adaptation du comportement de l'enseignant et de la gestion du groupe pourraient être réalisées afin de faciliter l'acceptation de la méthode et atténuer le malaise dû au bouleversement des habitudes d'apprentissage. Bien que parfois déstabilisante, l'approche choisie se révèle être un outil puissant pour faire prendre conscience aux apprenants des réalités phonologiques de la langue étudiée.

# Bibliographie

- AOYAMA, Katsura et al. (2003). « Foreign accent in English words produced by Japanese children and adults ». In : *Proceedings of the 15th International congress of phonetic sciences*, p. 3201-3204.
- BENSTEIN, Patricia (1996). « Explaining Concepts Behind The Silent Way ». In : *The Science of Education in Questions. Une Education Pour Demain* 14.
- BLANC, Georgette (2011). « L'enseignement de la phonétique du français langue étrangère (FLE) ». In : *Babylonia* 2, p. 33-37.
- BLANCHET, Philippe (2016). *Discriminations : combattre la glottophobie*. Paris : Textuel. 192 p. ISBN : 978-2-84597-544-6.
- BLOOMFIELD, Leonard (1935). *Language*. Allen et Unwin.
- COLLETTA, Jean-Marc (2009). *Le geste de pointage au service des acquisitions métaphonologiques*. Atelier.
- COSTE, Daniel et Robert GALISSON (1976). *Dictionnaire de didactique des langues*. Hachette.
- DEHAENE-LAMBERTZ, Ghislaine et Brit VAN OOIJEN (2002). « Le langage dans les premières années de la vie, nouvelles perspectives ». In : *Médecine et enfance* 7.22, p. 393-399.
- DETEY, Sylvain (2007). « Transcription, translittération et didactique de l'oral en FLE au Japon : katakana, rômaji et orthographe française ». In : *Revue japonaise de didactique du français* 2.1, p. 19-36.
- DIGALAKIS, Vassilis et N. MOUSTROUFAS (2007). « Automatic pronunciation evaluation of foreign speakers using unknown text ». In : *Computer, Speech and Language*.
- EIMAS, Peter D. et al. (1971). « Speech perception in infants ». In : *Sciences* 171, p. 303-306.

- FLAMENT-BOISTRANCOURT, Daniele (2006). « Enseignement/apprentissage des langues à l'aube du XXI<sup>e</sup> siècle : enjeux et tendances ». In : *Revue française de linguistique appliquée* XI.1, p. 5-6.
- FLEGE, James E. (1980). « Phonetic approximation in second language acquisition ». In : *Language Learning* 30, p. 117-134.
- FLEGE, James E., David BIRDSOING et al. (2006). « Degree of foreign accent in English sentences produced by Korean children and adults ». In : *Journal of phonetics* 34, p. 153-175.
- FLEGE, James E., Murray MUNRO et Ian MACKEY (1995). « Factors affecting degree of perceived foreign accent in a second language ». In : *Journal of the acoustical society of America* 97, p. 3125-3134.
- HATTORI, Shirō (1960). *言語学の方法 (Méthodes en linguistique)*. Iwanami Shōten.
- HIRSCHFELD, Ursula (2011a). « Integration der Ausspracheschulung in den DaF-/DaZ-Unterricht ». In : *Babylonia* 2, p. 53-57.
- (2011b). « Phonetik im Kontext mündlicher Fertigkeiten ». In : *Babylonia* 2, p. 10-17.
- KAMATA, Tadashi et Torataro YONEYAMA (2001). *新版漢語林 (Dictionnaire de kango nouvelle édition)*. 第 2 版. 東京 : 大修館書店.
- KAMIYAMA, Takeki et Jacqueline VAISSIÈRE (2009). « Perception and production of French close and close-mid rounded vowels by Japanese-speaking learners ». In : *Acquisition et interaction en langue étrangère (Aile... Lia 2)*. Avec la coll. de Jean-Yves DOMMERGUES, p. 9-41.
- KAWAI, Goh et Keikichi HIROSE (1997). « A CALL system using speech recognition to train the pronunciation of Japanese long vowels, the mora nasal and mora obstruents ». In : *Proceedings of EuroSpeech 1997*.
- KINDAICHI, Haruhiko et Hubert MAËS (1978). *Phonologie du japonais standard*. Paris : : Université de Paris 7. 1 vol. (59 p.)
- KUFFER, Jean-Louis (2006). *Le creaveu bmauin lit dnas le dsérorde*. Carnets de JLK.
- KUHL, Patricia K. (2004). « Early language acquisition : cracking the speech code ». In : *Nature Reviews* 5, p. 831-844.
- LABRUNE, Laurence (2006). *La phonologie du japonais*.

- LAURENT, Maurice (2004). *Les jeunes, la langue, la grammaire : d'une grammaire implicite à une grammaire explicite. 1, Catégories de mots, constituants de la phrase*. Larnod : UEPD. 2 vol. (508-479 p.)
- (2009). *Les jeunes, la langue, la grammaire : d'une grammaire implicite à une grammaire explicite. 2, Orthographe grammaticale, expression du temps, conjugaison*. Larnod : UEPD. 1 vol. (479 p.)
- LE PAGE, Noreen (2015). « Kinéphones : un outil pour la prononciation en langue étrangère et le sentiment d'efficacité personnelle des enseignants du primaire ». *Mém.de mast.*, p. 150.
- LIU, Ming et Kiyoshi WASHIO (2010). « 中国語の日本語への影響 (The Influence of Chinese on Japanese) ». In : *The Bulletin of Chuo-Gakuin University : man & nature* 31, p. 165-176.
- MATTER, Johan F. (2006). « La prononciation authentique en langue étrangère : un problème négligé ». In : *Revue française de linguistique appliquée* XI.1, p. 21-32.
- MORAZ, Myriam (2011). « Parler comme un francophone ou comment s'approprier l'intonation ». In : *Babylonia* 2, p. 38-43.
- MOROHASHI, Tetsuji (2000). *大漢和辞典 (Daikanwa Dictionary)*. 修訂第2 鎌田版. Tokyo : Taishūkan shoten.
- NAZZI, Thierry, Josiane BERTONCINI et Jacques MEHLER (1998). « Language discrimination by newborns : towards an understanding of the role of rhythm ». In : *Journal of Experimental Psychology : Human Perception and Performance* 3.24, p. 756-766.
- NHK (2015). 紫式部・清少納言～国風文化の誕生～ (*Murasaki Shikibu - naissance de la culture nationale*). NHK for School, 歴史にドキリ (Étonnante Histoire). URL : <http://www.nhk.or.jp/syakai/dokiri/> (visité le 02/04/2016).
- NINOMIYA, Chika (2015). « Attention et prises de conscience dans l'apprentissage : une étude des comportements enseignant/apprenants dans l'approche Silent Way ». *Mém.de mast.*, p. 91.
- NOTSU, Hiroshi et Jean-Pierre LEVET (1997). *Manuel d'initiation au japonais : les caractères linguistiques généraux, les premières leçons*. Presses Univ. Limoges. 232 p.
- PISTORIUS, Hannelore et Bettina WETZEL-KRANZ (2011). « L'enseignement de la prononciation en langues étrangères ». In : *Babylonia* 2, p. 4-89.
- RACINE, Isabelle, Sylvain DETEY et Yuji KAWAGUCHI (2012). « Les voyelles /y-u/ dans IPFC : évaluation perceptive de productions natives, hispanophones et japono-



- phones ». In : *Actes de la conférence conjointe JEP-TALN-RECITAL 2012*. ATALA & AFCP 1, p. 385–392.
- ROSLYN, Young et Piers MESSUM (2011). *How we learn and how we should be taught*. London : Duo Flumina. 1 vol. (x-202 p.) ISBN : 978-0-9568755-0-1.
- TELLIER, Marion (2010). « Faire un geste pour l'apprentissage : Le geste pédagogique dans l'enseignement précoce ». In : *L'enseignement des langues vivantes étrangères à l'école*. L'Harmattan, p. 31–54.
- THOMAS, Alain (2015). « L'impact d'une séjour linguistique à l'étranger sur la prononciation d'étudiants anglophones de FLS ». In : *Les carnets de l'atelier de socio-linguistique* 10.
- TRAN, Thi Thuy Hien (2011). « Processus d'acquisition des clusters et autres séquences de consonnes en langue seconde : de l'analyse acoustico-perceptive des séquences consonantiques du vietnamien à l'analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE. » Thèse de doct. Université Stendhal - Grenoble III. (Visité le 23/04/2016).
- TRUBETZKOY, Nikolai S. (1939). *Principes de phonologie (Grundzüge der Phonologie)*. Klincksieck.
- VANCE, Timothy J. (2008). *The Sounds of Japanese with Audio CD*. Cambridge University Press. 263 p.

# Table des figures

1.1	Catégorisation des voyelles d'après KUHL (2004, p. 834) . . . . .	10
1.2	Les transcriptions <i>bopomofo</i> et <i>pinyin</i> . . . . .	17
1.3	Interface de l'utilisateur traduite du programme de Kawai et Hirose . . . . .	24
2.1	Illustration du processus d'apprentissage chez un apprenant d'anglais . . . . .	30
2.2	Tableau phonologique du français de Gattegno, avec transcription . . . . .	34
2.3	Le fidèle pour l'anglais des États-Unis . . . . .	35
3.1	Extrait du <i>Nihon-syoki</i> (720) . . . . .	43
3.2	Extrait du manuscrit du <i>Dit du Genji</i> (XI <sup>e</sup> siècle) . . . . .	45
3.3	Les <i>kana</i> japonais et une transcription phonétique possible . . . . .	47
3.4	F <sub>0</sub> , F <sub>1</sub> , F <sub>2</sub> des voyelles et triangle vocalique du japonais (Sugito, 1995, cité par KAMIYAMA et VAISSIÈRE (2009)) . . . . .	50
5.1	Lieux de vie principaux des locuteurs . . . . .	75
5.2	Nombre d'années d'apprentissage du français par locuteur . . . . .	75
5.3	Préparation de l'enregistrement et de son annotation avant de lancer EasyAlign . . . . .	84
5.4	Étape 1 : Phonétisation de la transcription orthographique . . . . .	85
5.5	Étape 2 : Alignement des syllabes & phonèmes sur le signal sonore . . . . .	85
6.1	Distinction phonologique entre /z/ et /dz/ . . . . .	91
6.2	Plan de la ville avec extensions dessinées de part et d'autre . . . . .	92
7.1	Cas d'option ignorée malgré la présence du phonème dans le signal . . . . .	97
7.2	Comparaison de l'alignement avec (droite) et sans (gauche) options . . . . .	99
7.3	Traitement des données . . . . .	102
7.4	Nombre de disparitions et maintiens en fonction du phonème optionnel . . . . .	105
7.5	Alignements de « anglais » à T <sub>1</sub> . . . . .	106

7.6	Alignements de « quatre » à T <sub>1</sub> . . . . .	106
7.7	Alignements à T <sub>1</sub> de « soixante-sept » (gauche) et « trente-quatre » (droite) . . . . .	107
7.8	Alignements de « quatorze » à T <sub>1</sub> . . . . .	107
7.9	Alignements à T <sub>1</sub> de « dix-huit » (gauche) et « soixante-dix-huit » (droite) . . . . .	108
7.10	Score obtenu par locuteur pour la prononciation du R . . . . .	109
7.11	Score global de prononciation par locuteur (gauche) et nombre d'années d'apprentissage du français par locuteur (droite) . . . . .	110
7.12	Nombre à T <sub>2</sub> de disparitions et maintiens en fonction du phonème optionnel . . . . .	112
7.13	Alignements de « quatre » à T <sub>1</sub> (gauche) et T <sub>2</sub> (droite) . . . . .	112
7.14	Alignements de « université » à T <sub>1</sub> (gauche) et T <sub>2</sub> (droite) . . . . .	113
7.15	Score obtenu à T <sub>2</sub> par les locuteurs pour la prononciation du R . . . . .	113
7.16	Score global de prononciation à T <sub>1</sub> et T <sub>2</sub> par locuteur . . . . .	114
8.1	Mémorisation du tableau par les participants . . . . .	126
8.2	Fréquences d'apparition des phonèmes dans les tableaux (du plus au moins fréquent) . . . . .	127

# Liste des tableaux

1.1	Les 6 niveaux de difficulté dans l'acquisition de la phonologie en langue étrangère, d'après BLANC (2011) . . . . .	13
2.1	Les étapes d'apprentissage selon Gattegno . . . . .	29
3.1	Tableau des phonèmes consonantiques d'après KINDAICHI et MAËS (1978, p. 11)	52
3.2	Quelques réalisations de /N/ à l'intervocalique inspiré de HATTORI (1960) . .	56
3.3	Listes des structures syllabiques admises en japonais d'après KINDAICHI et MAËS (1978, p. 12) . . . . .	59
3.4	Structures syllabiques rares, d'après KINDAICHI et MAËS (1978, p. 12) . . . . .	59
5.1	Table des clusters ciblés . . . . .	77
7.1	Liste des nouvelles options . . . . .	100
7.2	Nombre de disparitions et d'insertions en fonction de la position du phonème	104
7.3	Nombre d'erreurs en fonction du phonème et de sa position . . . . .	105
7.4	Erreurs en coda par occurrence et par locuteur . . . . .	109
7.5	Nombre de réalisations correctes et total de R* en fonction des locuteurs . .	110
7.6	Nombre à T <sub>2</sub> de disparitions et d'insertions en fonction de la position du phonème . . . . .	111
7.7	Nombre à T <sub>2</sub> d'erreurs en fonction du phonème et de sa position . . . . .	111
7.8	Nombre à T <sub>2</sub> de réalisations correctes et total de R* en fonction des locuteurs	113



# Annexes

- Annexe 1. : Transcriptions fines des mots cibles à T<sub>1</sub> pour N1, N6 et N8
- Annexe 2. : Graphiques d'alignement des mots cibles à T<sub>1</sub> et T<sub>2</sub>
- Annexe 3. : Tableaux de correspondance SAMPA - API
- Annexe 4. : Tableaux dessinés de mémoire par les participants à T<sub>2</sub>

Transcription fine des mots cibles en T<sub>1</sub> pour N1, N6 et N8

N1			N6			N8		
ortho	phono	tmpps	ortho	phono	tmpps	ortho	phono	tmpps
à Grenoble pour	akχənobləpυrɣu	1041	qui est à Grenoble#	kjetəgənoblə	548	à Grenoble#	agənob	384
						#Grenoble est le	g(ɤ)ənob(1)elə	709
						#Grenoble est	gənobɣle	714
l'anglais#	lāgle	519	l'anglais# (lent)	lāgle	318	l'anglais#	lāgle	435
l'anglais#	lāgle	528						
cours d'université#	lūkuɣ.dynivɛɣsɣte	438	l'université de	lynβɛɣsed	541	l'université de Tokyo	lynivessedəto:kjo:	220
l'université Stendhal	lynivɛɣs.testādal	1048	#université Stendhal#	ynivɛɣstestēdal	584	#université de Tokyo	jy(n)ivɛsɣtedəto:kjo:	693
l'université Stendhal	lynivɛɣsɣtest.ādal	1057				l'université de Grenoble	lynibestedəkχənobl	707
université Stendhal#	lynivɛɣs.testādal	1048	de Stendhal#	ɔstēdal	542	j'ai cours de Stendhal mais	zɛku.dəstēdajme:	49
université Stendhal#	lynivɛɣsɣtest.ādal	1057				cours de Stendhal#	ku:hdəstēdah	124
à classe de	a.klasdə:	1015	dans la classe#	dālklasə	577	des classes de Stendhal	deklas.dəstēda(l)	738
pour classer#	pυrɣklase	1024						
j'ai un test#	ɔɣeētɛsth	1027	final test#	finaltɛsth	575	un test#	ētɛst	751
Rhône-Alpes#	χənalphu	1074	#alpes#	alps:	553	la Rhône-Alpes#	lalənalp <sup>h</sup>	719
en France#	āpχā.sə	998	et après France	eapχe:pχās	163	je viens en France#	zəvjē.āhχās	702
			je suis en France#	əzəsɣizāpχāsə	526			
			étudier le france#	etydjələpχās:s:ə	539			
deux ans ou trois ans	dəzā.utχ <sup>o</sup> azā	262	tr trois cours	tχ tχ <sup>o</sup> akwɣ	293	depuis trois ans	dəpɣithwazā	326
pendant trois semaines	pādātɣ <sup>o</sup> asmən	463	deux trois quatre	dətχwakatɣə	642	depuis trois ans	dəpɣitwazā	329
pendant trois semaines	pādā:tχwasəmən	1010	#trois#	tχa	720	#trois semaines	tχwasmen	390
deux trois quatre	dətχwakatɣə	1121				deux trois#	dətχwa	812
#trois#	tχa	1187				#trois#	twa	857
trois quatre cinq	tχwakatχəsēk	1121	trois quatre cinq	tχwakatχəsēk	642	#quatre cinq	katχəsēk	813
Trente-quatre#	tχākatɣ	1147	#quatre#	katɣə	719	#quatre#	katɣ	857
#quatre#	katɣ (palatalisation?)	1185						
douze treize#	dustɣes	1127	#treize#	tχɛ:s	649	douze treize quatorze	dustɣɛskatɔɣzə	818
quatorze treize douze	katɣəstɣesdus	1173	#treize#	tχɛ:s	710	#treize#	t(χ)ɛsu	851
#quatorze quinze	katɣəskēs	1130	#quatorze#	katɣos	650	douze treize quatorze#	dustɣɛskatɔɣzə	818
quinze quatorze treize	kēskatɣəstɣes	1173	#quatorze#	katɣos	708	#quatorze#	katɔɣs	850
dix dix-huit ans	di:dizɣitā	384	jusqu'à dix-huit ans	zyskadizɣitā	147	#dix-huit#	dizɣit	823
dix-sept ou dix-huit#	disɛteudizɣit	387	#dix-huit#	dizɣit	654	#dix-huit#	dizɣit	847
dix-sept dix-huit dix-neuf	disɛt'dizɣitdiznə	1133	#dix-huit#	dizɣit	702	soixante-dix-huit	swasādizɣit	838
soixante-dix-huit#	s <sup>o</sup> asādizɣith	1154						
#dix-huit dix-sept	dizɣithdisɛt	1169						
dix-neuf dix-neuf vingt	diznəf.vē	1135	#dix-neuf#	diznəf	655	#dix-neuf#	diznəf	824
dix dix-neuf#	disn.di:znəf	1167	#dix-neuf#	diznəf	701	#dix-neuf#	diznəf	847
trente-quatre	tχākatɣ	1147	trente-quatre	tātkatɣh	669	trente-quatre	tχātkatɣh	830
quarante-huit	kaχātɣith	1150	quarante-huit	kaɪātɣith	671	quarante-huit	kχātɣit	833
soixante-sept	swasāsɛth	1152	soixante-sept	swasāsɛt	673	soixante-sept	swasāsɛt	837
soixante-dix-huit	s <sup>o</sup> asādizɣith	1154	soixante-dix-neuf	swasādizənə	674	soixante-dix-huit	swasādizɣit	838





Correspondances SAMPA - API, tirées du projet Rhapsodie

<http://projet-rhapsodie.fr/>

SAMPA	API	Exemples
p	p	<i>pont</i>
b	b	<i>bon</i>
t	t	<i>temps</i>
d	d	<i>dans</i>
k	k	<i>coût, quand, koala</i>
g	g	<i>gant</i>
f	f	<i>femme</i>
v	v	<i>vent</i>
s	s	<i>sans, dessus, cerise</i>
z	z	<i>zone, rose</i>
S	ʃ	<i>champ</i>
Z	ʒ	<i>gens, jambon</i>
j	j	<i>ion [jo~]</i>
m	m	<i>mont</i>
n	n	<i>nom</i>
J	ɲ	<i>oignon</i>
N	ŋ	<i>camping</i>
l	l	<i>long</i>
R	ʁ	<i>rond</i>
w	w	<i>quoi [kwa]</i>
H	ɥ	<i>juin [ZHe~]</i>

SAMPA	API	Exemples
i	i	<i>si</i>
e	e	<i>Blé</i>
E	ɛ	<i>seize</i>
a	a	<i>patte</i>
A	ɑ	<i>pâte</i>
O	ɔ	<i>comme</i>
o	o	<i>gros</i>
u	u	<i>doux</i>
y	y	<i>du</i>
2	ø	<i>deux</i>
9	œ	<i>neuf</i>
@	ə	<i>justement</i>
e~	ē	<i>vin</i>
a~	ā	<i>vent</i>
o~	ō	<i>bon</i>
9~	œ̃	<i>brun</i>

N9

[ɛ] <sup>e œ œ</sup>  
|  
[i]

[y] <sup>i u</sup> <sup>ix</sup>

[~~œu~~] <sup>u</sup>

|  
[œ] <sup>ø</sup>

|  
[o] <sup>o</sup>

|  
[ɔ] <sup>ɛ</sup>

|  
[ao] <sup>o</sup>

[a]

<sup>ü</sup> [œn] <sup>ø œ</sup> [œn] <sup>an</sup> [œn]



[m] [n]

[ŋ]

[p] [t]

[k] [g]

[b] [d]

[f] [s] [ch]

[v] [z] [j]

[l] [r]

i

"u"

"ou"

e

ø

o

e

ø

entre O et A

a

?

?

"au"

"ou"

é

œ

m

n

ng

k

p

t

g

b

d

f

s

ch

☺

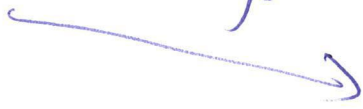
v

z

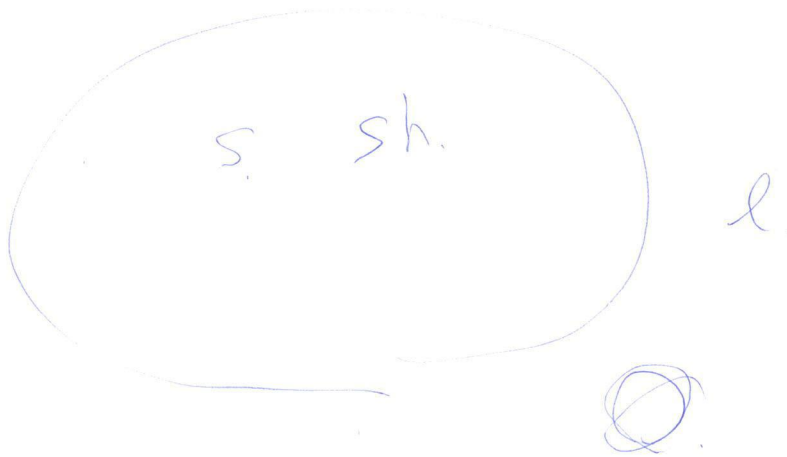
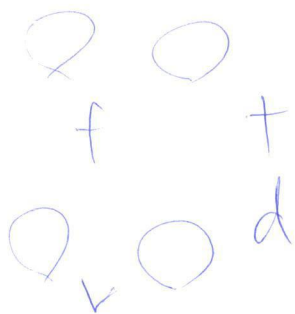
Composé "e"

l

r



N7



[r]	[k]	[g]	[d]	[t]	[f]	[v]	[z]	[m]	[n]
[l]									

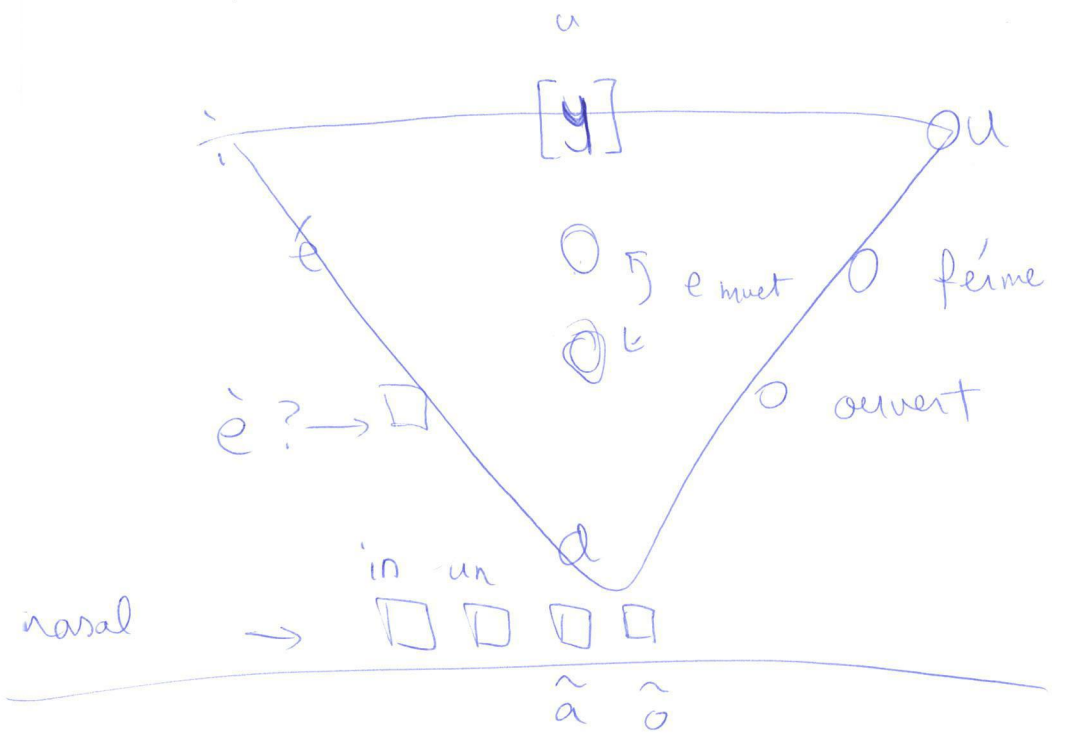
[ɔ]	[æ]	[ɛ]	[ɜ]
-----	-----	-----	-----

[a]

[3]	[e]	[i]	[u]	[ɔ]	[ɜ]	[aɪ]	[e]	[i]

a i

N5



} m n  
 } p t  
 } b d  
 }

gn n k g

? f s ch  
 v z j

d  
 e b

i.

u [y]

ウ [u]

オ [o]

~~φ~~  
I [e]

a. ẽ h ãn.

m. n.

s f. p.

t.

z. b.  
[v]

[k]

r. l.

N2

<sup>e</sup>  
 [e]      <sup>i</sup>  
 [i]      <sup>yu</sup>  
 [õ]      <sup>w</sup>  
 [ö]

<sup>ε</sup>  
 [ ]      [ä]      <sup>ü</sup>  
 [ü]      <sup>u</sup>  
 [u]      [o]

<sup>a</sup>  
 [a]      ~~[ ]~~  
 [a]

<sup>2</sup>

<sup>ĩ</sup>  
 [ēn]      <sup>ã</sup>  
 [ãn]      <sup>õ</sup>  
 [õn]

[m]      [n]

[p]      [t]

[b]      [d]      [f]      [s]      [ ] + sh      [k]

[v]      [z]      [j]      [r]



N1

i □

u □  
~~u~~

□ ~~ou~~

e □

□ ø

□ °

e □

□ ə

□ ɔ

a □

ẽ  
□ in

ã  
□ an

ã  
□ en

õ  
□ on

□ m

□ n

□ b

□ p

□ d

□ t

□ f

□ s

□ ch

□ v

□ z

□ j

□ r

□ l