

Cahier des Charges

P-ALES

Génération d'une représentation chromatique des systèmes phonologiques
pour l'enseignement des langues sinogrammiques

Sommaire

I. Présentation des parties	3
1. Maîtrise d'œuvre	3
2. Maîtrise d'ouvrage	3
II. Présentation du projet	4
1. Définition de P-ALES	4
2. État de l'art	4
3. Étude du besoin	5
4. Objectif	6
III. Future solution	6
1. Spécificités techniques	6
2. Scénarisation	9
3. Ressources	9
IV. Organisation et calendrier	10
Organisation	11
Tâches	11
Diagramme de Gantt	13
V. SWOT	15

I. Présentation des parties

1. Maîtrise d'œuvre

Anays MICOLOD anays.micolod@gmail.com

Célia MARION celia.marion2@gmail.com

Nous sommes deux étudiantes en 2ème année de Master Sciences du Langage parcours Industries de la langue. Ce projet, encadré par Monsieur Thomas Lebarbé, est réalisé dans le cadre de l'UE Projet Professionnel.

Sylvain COULANGE sylvain.coulange@univ-grenoble-alpes.fr

Sylvain COULANGE est ingénieur pédagogique. M.A. didactique du FLE. mis à disposition par le maître d'ouvrage. Mission spécifique sur le persan et la visualisation des tableaux chromatiques. Sa mission sera concentrée sur les trois premières semaines de travail effectives.

2. Maîtrise d'ouvrage

Yoann GOUDIN yoann.goudin@univ-grenoble-alpes.fr

Yoann Goudin est ingénieur spécialiste du mandarin, Docteur en Didactique des Langues et a collaboré au projet Kinéphones.

Guillaume LECHIEN gui.lee@gmail.com

Guillaume LECHIEN est ingénieur informatique. M.A. TAL de l'Université Paris Diderot. Il est le réutilisateur final de P-ALES. Chaque décision de solution technique sera discutée et validée avec lui à distance.

II. Présentation du projet

1. Définition de P-ALES

P pour Phonochrome, nom de l'outil qui permet de noter des sons en couleurs.

ALES pour *À l'Écoute des sinogrammes*, publication « la page et l'écran » prévue sur l'acquisition du lexique en langues sinogrammiques et le système sinographique.

P-ALES est donc un outil permettant d'encoder au moyen de couleurs les phonèmes d'une langue considérée.

2. État de l'art

Le projet P-ALES hérite d'une approche de l'enseignement appelée *Silent Way* et s'inscrit dans la lignée de projets antérieurs, eux-même héritiers de cette méthode.

Silent Way est donc une approche d'enseignement des langues étrangères et à commencer par la phonologie et de la conscientisation phonologique. Cette approche a été conçue par Caleb Gattegno dans les années 70. Elle se caractérise par le fait que l'enseignant reste silencieux mais valide les phonèmes bien réalisés par les apprenants, qui sont ainsi encouragés le plus possible à la production orale. Les tableaux de phonèmes chromatiques (où chaque phonème apparaît sous la forme d'un rectangle de couleur) et les tableaux de mots (où les lettres sont colorisées en fonction de leur prononciation) sont des outils généralement utilisés dans l'application de cette méthode.

Le premier projet sur lequel s'appuie directement le nôtre est **Kinéphones**. Kinéphones¹ propose de dématérialiser la représentation phonologique du français et de l'anglais (GB et US), en représentant les phonèmes par leur symbole phonétique ou par des rectangles de couleur sur des tableaux. Il permet à l'apprenant de faire une découverte autonome du système phonologique de la langue cible et à l'enseignant d'entraîner la prononciation et la discrimination phonologique de ses apprenants. Sa contribution majeure est l'univocité de l'association phonème = couleur = positionnement sur les tableaux en fonction des caractéristiques articulatoires.

Notre projet se base également sur **Phonochrome**, projet réalisé en 2017 dans le cadre du M2 IDL et s'inscrivant aussi dans la lignée de Kinéphones. L'objectif initial était de permettre un encodage des sons en couleurs en définissant une association phonème-couleur "universelle" ou par défaut et pour un maximum de langues possible. L'outil permet une colorisation des graphèmes du texte saisi par l'utilisateur, lui donnant ainsi une indication sur la manière de prononcer.

3. Étude du besoin

Les utilisateurs (apprenants et enseignants) doivent pouvoir avoir accès aux représentations chromatiques des phonèmes ainsi qu'à la possibilité de voir les sons en décodant les couleurs comme des sons. Pour les apprenants, l'enjeu est de comprendre le système phonologique et la prononciation de la langue cible pour la maîtriser. Pour les enseignants, l'enjeu est d'avoir accès à un support fiable et performant pour enseigner la phonologie.

Comme indiqué dans la définition du projet, P-ALES concerne les langues sinogrammiques. Les langues sinogrammiques sont des langues d'Asie Orientale dont tout ou partie du lexique s'écrit ou s'est écrit en sinogrammes et leurs lectures qui respectent le patron syllabique du chinois classique. Les langues sinogrammiques auxquelles nous nous intéressons

¹ disponible à l'adresse suivante : <http://kinephones.u-grenoble3.fr>
description : <https://kinephones.hypotheses.org/>

dans le cadre du projet sont les suivantes : le mandarin, le coréen, le japonais, le taïwanais et le cantonais.

Les couleurs utilisées pour les représentations doivent répondre à deux contraintes. Tout d'abord, elles doivent rendre compte des différentes fréquences des phonèmes des langues cibles , ainsi les couleurs les plus saillantes seront attribuées aux phonèmes les plus fréquents. De plus, les phonèmes apparaissant fréquemment ensemble (dans les chaînes de phonèmes attestées) doivent avoir des couleurs contrastées afin de faciliter leur discrimination par les apprenants. L'outil Phonochrome met déjà à disposition des fréquences d'adjacences que nous étudierons lors de la réalisation dans le but de les intégrer dans notre outil.

4. Objectif

Notre objectif est de mettre à disposition une représentation chromatique des systèmes phonologiques pour l'enseignement des langues sinogrammiques. Cette représentation prendra deux formes. En effet l'outil que nous proposons permet la génération de tableaux chromatiques représentant le système phonologique de chaque langue cible.

Nous produirons également un document explicitant le processus permettant d'arriver à ces tableaux.

III. Future solution

1. Spécificités techniques

Notre outil se décompose en deux modules.

- module 1 : calculateur des couleurs

Premièrement, ce module permettra de générer la liste des paires adjacentes de phonèmes avec leurs fréquences respectives (sous la forme d'un tableau associatif) à partir de lexiques phonétisés. Le tableau associatif produit par le projet Phonochrome sera évalué afin de voir s'il peut servir de support de base.

Deuxièmement, il effectuera également le calcul des couleurs à partir du tableau associatif préalablement mentionné plus haut (paires adjacentes de phonèmes et leurs fréquences), indépendamment de la langue. Cette portion du module s'appuiera sur l'algorithme "delta-e" créé lors du projet Phonochrome. Les couleurs seront représentées par leur code hexadécimal. La sortie prendra donc la forme d'un tableau associatif [phonème : code hexadécimal de la couleur]. Cette association sera unique : un phonème = une couleur pour toutes les langues.

Ce module sera réalisé en python.

- module 2 : tableaux chromatiques

Le deuxième module s'occupe de la génération des tableaux chromatiques.

On modélisera dans un premier temps un tableau "patron" dont les cases représenteront tous les phonèmes possibles pour toutes les langues, avec leur position absolue. Nous obtiendrons un tel tableau en exploitant les traits articulatoires de chaque phonème décrits dans le fichier ipa_all.csv². C'est donc ses caractéristiques articulatoires qui déterminent la position d'un phonème.

Puis, pour chaque langue, on générera un tableau intermédiaire avec l'ensemble de ses phonèmes. Les cases de ce tableau représenteront tous les phonèmes pour chacune des langues, dans leur position absolue (à partir du tableau patron). Pour cela, nous utiliserons la liste des phonèmes de chaque langue produite dans le module 1.

Enfin, nous produirons une version dite "Kinéphones" du tableau intermédiaire qui soit plus lisible et si possible didactique. Dans ce tableau seront délimités les espaces représentant les voyelles nasales, orales, les semi-voyelles et les consonnes. La position et la taille des rectangles devront

² voir paragraphe Ressources

être relatives au nombre d'éléments de l'espace dans lequel ils se trouvent. A l'heure actuelle, nous ne savons pas comment passer des tableaux intermédiaires aux tableaux "Kinéphones".

A noter que les couleurs sur les tableaux apparaîtront seulement sur l'interface web. C'est le script php qui effectuera l'association entre les rectangles et les couleurs, via la feuille de style.

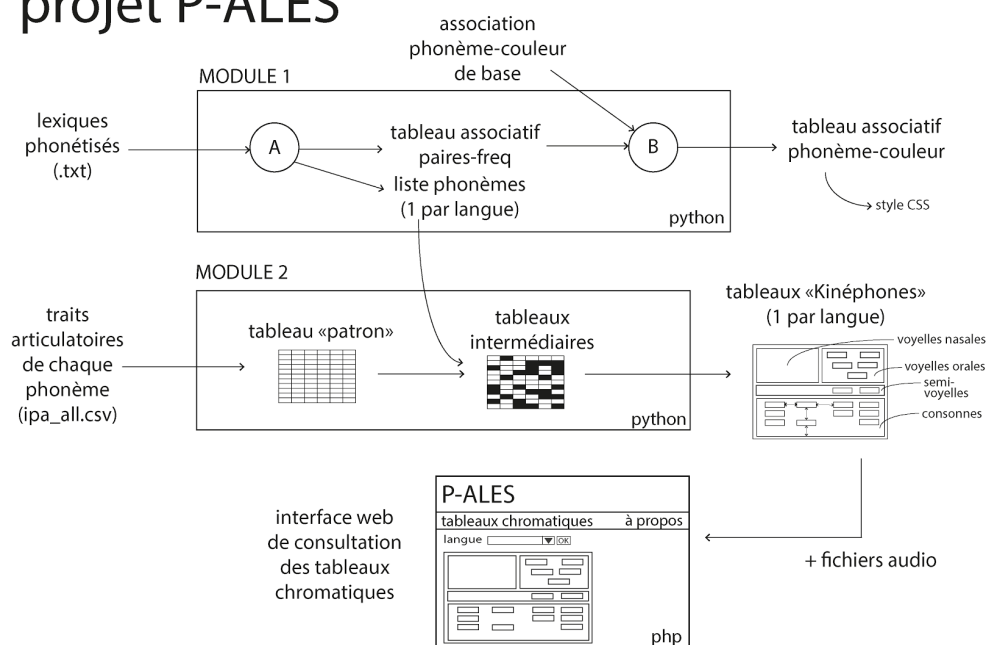
Nous proposerons également une option qui permettra à l'utilisateur d'écouter le son (phonème) correspondant à chaque couleur. Cette fonctionnalité permettra à l'outil d'être au plus proche de Kinéphones. Nous envisageons plusieurs scénarios d'interactions avec l'utilisateur dans lesquelles le lancement du fichier son se ferait :

- en cliquant sur l'un des rectangles de couleur du tableau ;
- depuis un "player" affiché en passant la souris (mouse-over) sur le rectangle de couleur ;
- directement durant le passage de la souris ;
- depuis une icône speaker cliquable que chaque rectangle de couleur posséderait.

Le script permettant de générer les tableaux sera réalisé en python et l'interface web de consultation des tableaux sera codée en PHP.

Ci-dessous un schéma représentant le fonctionnement de l'outil.

projet P-ALES



Il sera possible d'ajouter une nouvelle langue à l'outil à partir de son lexique phonétisé. Cet ajout mettra à jour la suite des paires adjacentes et de leurs fréquences, que l'on utilise pour re-calculer les couleurs (sur la base de l'association précédente). Nous testerons notamment cette généralité pour le cas du persan.

Nous prévoyons d'héberger notre plateforme sur le serveur dip01.u-grenoble3.fr. Cependant, ce serveur étant en fin de vie, il faudra héberger les données et codes sur un autre serveur la fin du projet. L'ensemble du travail sera disponible sur un github.

2. Scénarisation

Lorsque l'utilisateur (apprenant ou enseignant) se rendra sur la plateforme, il pourra accéder à différentes pages :

- page "Tableaux chromatiques"

Sur cette page, l'utilisateur aura la possibilité de choisir la langue cible dont il veut voir le tableau chromatique et, grâce à un bouton de confirmation, pourra lancer l'affichage du tableau. Une fois le tableau affiché nous souhaitons que l'utilisateur puisse entendre le son correspondant à chaque rectangle de couleur (phonème) en cliquant dessus ou grâce à un mouse-over.

- page "À propos" (à redéfinir lors de la conception de la plateforme)

Sur cette page, l'utilisateur aura accès à une description de la plateforme, au manuel d'utilisation, au lien github du projet et explicitera les différents choix possibles et les motifs qui ont conduit aux solutions finalement retenues.

3. Ressources

Les ressources dont nous avons besoin pour mener à bien ce projet sont des ressources phonologiques. Nous avons besoin d'un lexique phonétisé pour chaque langue cible (sous la forme d'un dictionnaire associant à chaque entrée sa forme phonétique) afin d'extraire les paires adjacentes de phonèmes et la liste des phonèmes de chaque langue. Nous avons également besoin de lexiques phonétisés pour le français et l'anglais, que nous prendrons en compte pour le calcul des couleurs.

De plus, nous aurons besoin de rassembler des fichiers audio pour les langues traitées afin de permettre leur mise à disposition via les tableaux chromatiques sur l'interface web. Nous nous chargerons de leur recherche et aurons besoin de la validation de leur qualité par la maîtrise d'ouvrage.

Nous n'avons pour l'instant pas accès aux ressources pour les langues sinogrammiques traitées, la recherche et la validation de ces ressources faisant partie de la phase de réalisation (à partir du 15 janvier 2018).

Le fichier nous permettant d'exploiter les traits articulatoires de chaque phonème de l'API (ipa_all.csv) provient du package panphon 0.5 (<https://pypi.python.org/pypi/panphon/0.5>). Ci-dessous une copie d'écran donnant un aperçu, avec en ordonnée les phonèmes en API et les traits en abscisse.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ipa	syl	son	cons	cont	delrel	lat	nas	strid	voi	sg	cg	ant	cor	distr	lab	hi	lo	back	round	velaric	tense	long
2	q	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-
3	q	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-
4	q	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-
5	g	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-
6	g	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-
7	g	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-
8	q	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	+
9	g	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	+
10	t	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-
11	t	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-
12	t	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-
13	t	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-
14	d	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-
15	d	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-
16	d	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-
17	d	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-
18	t	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	+
19	d	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	+
20	t	-	-	+	-	-	-	-	0	-	-	-	-	+	+	-	-	-	-	-	-	0	-
21	d	-	-	+	-	-	-	-	0	+	-	-	-	+	+	-	-	-	-	-	-	0	-

IV. Organisation et calendrier

1. Organisation

Nous informerons Guillaume LECHIEN par mail en destinataire principal (et Sylvain COULANGE et Yoann GOUDIN en copie) tous les deux jours pour faire un point sur la réalisation et la distribution des tâches, les problèmes rencontrés, les solutions trouvées, les problèmes irrésolus, questions, etc. Toutes les semaines, nous effectueront un back up sur le github pour indiquer l'avancement du projet, éviter les mauvaises surprises et permettre l'évaluation de la réalisation en cours.

2. Tâches

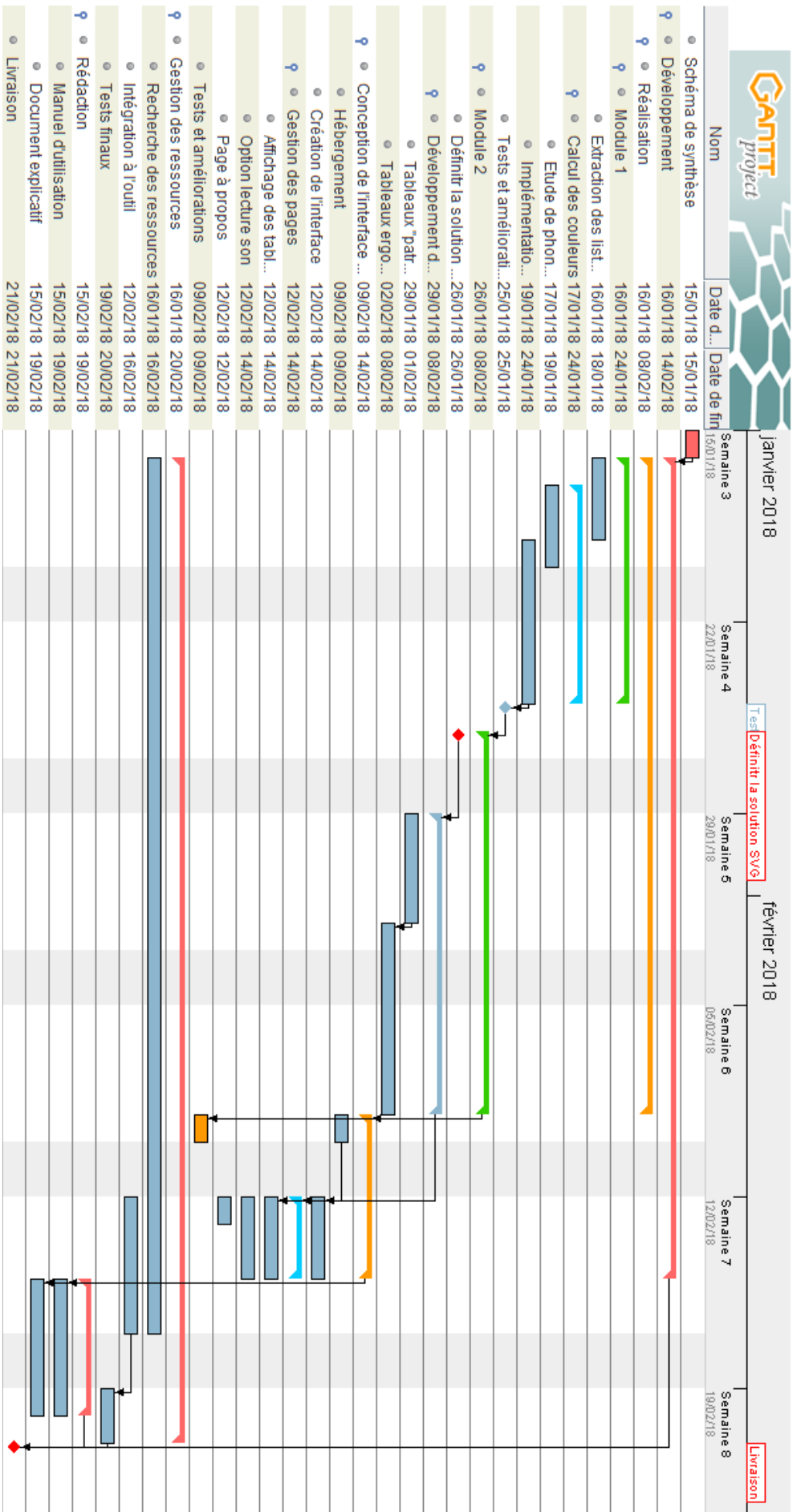
Nous disposons de 6 semaines pour développer notre outil et permettre son bon fonctionnement. Ci-dessous, la répartition des tâches et sous-tâches (prévisionnelle et susceptible de connaître des modifications en cours de réalisation).

Nom	Date de début	Date de fin
Schéma de synthèse	15/01/18	15/01/18
Développement	16/01/18	14/02/18
Réalisation	16/01/18	08/02/18
Module 1	16/01/18	24/01/18
Extraction des listes de paires adjacentes	16/01/18	18/01/18
Calcul des couleurs	17/01/18	24/01/18
Etude de phonochrome	17/01/18	19/01/18
Implémentation du calcul	19/01/18	24/01/18
Tests et améliorations	25/01/18	25/01/18
Module 2	26/01/18	08/02/18
Définir la solution SVG	26/01/18	26/01/18
Développement de la solution de génération	29/01/18	08/02/18
Tableaux "patrons"	29/01/18	01/02/18
Tableaux ergonomisés	02/02/18	08/02/18
Conception de l'interface web	09/02/18	14/02/18
Hébergement	09/02/18	09/02/18
Création de l'interface	12/02/18	14/02/18
Gestion des pages	12/02/18	14/02/18
Affichage des tableaux chromatiques	12/02/18	14/02/18
Option lecture son	12/02/18	14/02/18
Page à propos	12/02/18	12/02/18
Tests et améliorations	09/02/18	09/02/18
Gestion des ressources	16/01/18	20/02/18
Recherche des ressources	16/01/18	16/02/18
Intégration à l'outil	12/02/18	16/02/18
Tests finaux	19/02/18	20/02/18
Rédaction	15/02/18	19/02/18
Manuel d'utilisation	15/02/18	19/02/18
Document explicatif	15/02/18	19/02/18
Livraison	21/02/18	21/02/18

3. Diagramme de Gantt

Grâce au logiciel Gantt Project, nous avons réalisé un diagramme de Gantt pour présenter le calendrier prévisionnel de la réalisation du projet avec un détail des tâches et sous-tâches ainsi que le temps prévu pour chacune. Ce diagramme est non-exhaustif et susceptible d'être modifié au cours de la réalisation.

La livraison aura lieu en fin février et sera suivie d'une présentation du projet en présence de tous les acteurs.



V. SWOT

Nous allons maintenant présenter les forces et faiblesses de notre projet sous la forme d'une matrice SWOT.

<p>Strength : Force</p> <p>Ce qui fait la force de notre projet est sa généricité. En effet, on pourra utiliser cet outil pour toutes les langues que l'on souhaite, pas uniquement pour celle motivant ce projet.</p>	<p>Weakness : Faiblesse</p> <p>Les ressources que nous allons utiliser ne sont pas exhaustives, elles sont susceptibles de ne pas couvrir l'ensemble des paires adjacentes de phonèmes. Si tel est le cas, il faudra alors trouver de meilleures ressources, la conception et le fonctionnement de notre outil restera le même, seul le résultat final pourra être impacté.</p>
<p>Opportunities : Opportunités</p> <p>La constitution d'un corpus de fichiers sonores permettant l'écoute des phonèmes sera intéressante pour de nombreux chercheurs. C'est également l'opportunité pour nous de nous confronter à la réalité terrain.</p>	<p>Threats : Menaces</p> <p>La montée en compétence de notre équipe pourrait être un frein au projet si elle s'avère prendre plus de temps que possible compte tenu des délais.</p>