UNIVERSITÉ DE GENÈVE
COMMISSION INFORMATIQUE

# DEMANDE DE COFINANCEMENT A LA COMMISSION INFORMATIQUE

Merci de transmettre votre demande pour préavis au président de la Commission Informatique de votre Faculté (CIF):
M. Bastien.Chopard@unige.ch / Sciences, M. Antoine.Geissbuhler@unige.ch / Médecine,
Mme Genoveva.Puskas@unige.ch /Lettres, M. Dimitri.Konstantas@unige.ch / GSEM,
M. Eugen.Horber@unige.ch / SdS, M. Alexandre.Flueckiger@unige.ch / Droit,
M. Jean-Daniel.Macchi@unige.ch / Théologie, Mme Mireille.Betrancourt@unige.ch / PSE,
M. Kilian.Seeber@unige.ch / FTI.

Toutes les demandes de cofinancement doivent parvenir à votre CIF **au plus tard le vendredi 30 janvier 2015** pour être validées et transmises au Bureau de la COINF (Thierno Diallo). Les projets de type HPC seront préavisés par M. Bastien Chopard, coordinateur du groupe HPC de la COINF, qui en fera une analyse plus approfondie. Pour les structures (centres interfacultaires) ne bénéficiant pas d'une commission informatique de type CIF, veuillez adresser directement vos demandes à Thierno.Diallo@unige.ch. La sélection finale aura lieu à la séance COINF du 05 mars 2015.

| | |
|---|---|
| Date de réception : | Demande no : |
| Faculté/Département : Lettres/ELCF | Tél. : 022 379 74 46 |
| Nom/Prénom : RACINE Isabelle | E-mail :isabelle.racine@unige.ch |

**Exposé des motifs** (Explication et justification de l'importance du projet / Raison d'un financement spécial, brève description du matériel souhaité).

Conception et mise à disposition d'une base de données informatique permettant le stockage et l'interrogation des données de français oral d'apprenants non-natifs. Une quantité conséquente de données sont collectées depuis 2008, dans le cadre du projet « Interphonologie du français contemporain » (IPFC, Direction : I. Racine, UniGE, S. Detey, Waseda University, Japon et Y. Kawaguchi, TUFS, Japon, financement FNS no 132144/1, de 2011 à 2013). L'objectif de ce projet collaboratif, qui regroupe actuellement 15 équipes de chercheurs dans 15 pays différents, est de constituer et de mettre à disposition des chercheurs et des enseignants une large base de données, permettant des analyses de la parole non-native en français oral, notamment au niveau phonétique et phonologique. Une base de données informatisée est, à ce stade du projet, indispensable pour rendre, à terme, ces données pérennes et accessibles.

Voir document annexé.

**Indiquer si ce projet est de type HPC** (High Performance Computing) : NON

**Indiquer si ce projet est dans le domaine de l'enseignement** : NON

**Coût total net : CHF 20'370.-**

**Cofinancement demandé à la COINF : CHF 9'400.-**
*(voir règlement plus bas).*
**Type de demande :** investissement ☐ autres ressources ☒

| | |
|---|---|
| Préavis de la CIF ou du bureau de la COINF : | Priorité : ☐haute ☐moyenne ☐basse |
| Date : | Signature : |

| | |
|---|---|
| Décision de la COINF : | Priorité : |
| Date : | Signature : |

UNIVERSITÉ DE GENÈVE
FACULTÉ DES LETTRES
Ecole de langue et de civilisation françaises

**Prof. Isabelle Racine**
Ligne directe: 022 379 74 46
Isabelle.Racine@unige.ch

# Constitution d'une base de données informatique pour le projet « InterPhonologie du Français Contemporain » (IPFC)

**Requérante :** Prof. Isabelle RACINE

**Département/Faculté :** ELCF, Lettres

**Exposé des motifs :**

Au cours des dernières années, l'accroissement des capacités de stockage et de traitement des données numériques a ouvert la voie à la normalisation de la constitution de larges bases de données orales transcrites, alignées et annotées, lesquelles permettent de revisiter, sur une base empirique étendue, certains modèles et analyses linguistiques hérités de la tradition. En didactique des langues étrangères, les résultats de ces analyses de corpus, permettent d'affiner nos connaissances en matière de parole non-native et d'aboutir *in fine* à un renouvellement du matériel pédagogique.

Dans le domaine de la phonétique et de la phonologie du français L2, bien qu'il existe, depuis longtemps, différents travaux décrivant les principales caractéristiques de prononciation du français d'apprenants de diverses langues premières (L1), aucun corpus de ce type n'existait. C'est ce manque qui a motivé, en 2008, le lancement du projet « InterPhonologie du Français Contemporain » (IPFC, Detey & Kawaguchi, 2008 ; Racine *et al.*, 2012), qui constitue le volet non-natif d'un corpus devenu une référence en matière de français oral natif, le projet « Phonologie du Français Contemporain » (PFC, Durand *et al.*, 2014, site du projet www.projet-pfc.net). Le projet IPFC, co-dirigé par S. Detey et Y. Kawaguchi, à Tokyo, ainsi que par I. Racine, à Genève, vise à constituer et mettre à disposition des chercheurs et des enseignants une large base de données, permettant des analyses de la parole non-native en français oral, notamment au niveau phonétique et phonologique. La collecte des données a débuté en 2008 avec les deux premiers corpus, IPFC-japonais, à Tokyo, et IPFC-espagnol, à l'Université de Genève, soutenus par divers financements (au Japon : Société Japonaise pour la Promotion de la Science et Université Waseda ; en Suisse : Société Académique de Genève, Faculté des lettres de l'UniGE et FNS no 132144/1). A l'heure actuelle, le projet regroupe quinze équipes de chercheurs de différents pays, qui recueillent des données d'apprenants de différentes origines[1], par le biais d'un protocole identique, ce qui permet la comparabilité des données. Depuis 2010, des « Journées IPFC », qui permettent aux membres du projet, y compris des doctorants et étudiants, de se rencontrer et de présenter leurs travaux, sont organisées chaque année en décembre, à Paris, et un volume sur la prononciation du français dans le monde (plus de 40 contributeurs) sera publié cette année (Detey *et al.*, à paraître)[2].

Depuis sept ans, le projet IPFC n'a donc cessé de croître et il occupe, à l'heure actuelle, une place non négligeable dans le domaine de l'étude de la prononciation du français langue étrangère. Toutefois, comme pour tout corpus, sa pérennisation passe indiscutablement par la constitution et la mise à disposition d'une base de données informatique, permettant le stockage et la consultation des données collectées par chaque équipe. Le projet dispose actuellement d'un site web (http://cblle.tufs.ac.jp/ipfc/), qui a pu être créé grâce à un financement unique de TUFS (Tokyo University of Foreign Studies) et qui présente le projet IPFC, les équipes, les publications et les colloques. L'élaboration d'une plateforme informatique permettant le stockage et la consultation des données via un concordancier constitue toutefois un travail conséquent, qui requiert à la fois un budget et des compétences spécifiques en matière d'informatique. C'est pour la réalisation de cette plateforme informatique que nous sollicitons la COINF. Mentionnons encore

[1] Les L1 représentées dans le projet sont les suivantes: allemand, anglais, arabe, coréen, danois, espagnol, grec, italien, japonais, néerlandais, norvégien, portugais, suédois, russe et turc. Pour plus de détails, voir: http://cblle.tufs.ac.jp/ipfc/.
[2] La liste des publications liées au projet peut être consultée sur le site IPFC (http://cblle.tufs.ac.jp/ipfc/).

que le choix de l'Université de Genève pour héberger cette plate-forme est lié au rôle important que joue l'ELCF dans la co-direction du projet IPFC depuis ses débuts ; en outre, le fait qu'une telle base de données soit *in fine* consultée par tout un chacun, utilisée – et par conséquent citée dans les publications – de chercheurs de tous les continents, participerait au rayonnement et à la visibilité internationale de notre institution[3].

Plus concrètement, deux étapes sont prévues :

a) Création d'un site, avec implémentation du « back office » pour la saisie des données du projet (1h de données orales en fichiers .wav par locuteur, produites dans différentes tâches : listes de mots lues et répétées, texte lu et conversations), avec les métadonnées correspondantes (informations relatives au locuteur et à l'enregistrement) et la possibilité de télécharger les fichiers sons et les fichiers TextGrids contenant la transcription orthographique alignée sur le signal sonore (sous Praat, www.praat.org). Implémentation également d'un concordancier permettant d'effectuer des requêtes dans les données. L'hébergement de cette plateforme (base de données IPFC et concordancier) est prévu sur le site de l'UniGE, via un serveur dédié virtuel (cf. discussion avec Patrick Grespan). Les outils prévus pour la conception de cette base de données sont gratuits et nous nous sommes assurés qu'ils soient pérennes.

b) Une fois la base de données implémentée, saisie des données du projet, étape qui implique un important travail préalable d'homogénéisation, de vérification et d'anonymisation des données, ainsi que l'élaboration d'un protocole de préparation des données destinés aux différentes équipes de chercheurs, ce qui leur permettra de fournir du matériel préparé pour alimenter la base à l'avenir.

**Description des besoins :**

Pour la phase a), un développeur en informatique expert dans la conception de base de données est nécessaire. Nous proposons de confier ce mandat à M. Pierre Ménétrey (CV en annexe), qui a élaboré la base de données et le concordancier OFROM (Corpus Oral de Français Parlé), hébergés à l'Université de Neuchâtel (http://www11.unine.ch/). M. Ménétrey a également conçu la plateforme Labguistic (www.labguistic.com), que les chercheurs de l'ELCF et du Département de linguistique de l'UniGE utilisent pour des enquêtes expérimentales.

Pour la phase b), des connaissances spécifiques au niveau du projet IPFC sont requises, notamment en matière de données et de traitement des données. Nous proposons de confier ce travail à M. Roberto Paternostro, qui occupe actuellement un poste de post-doc (sur fonds overhead de la Faculté, contrat en annexe). M. Paternostro sera engagé deux mois pour la réalisation exclusive de cette phase b) (juillet et août 2015).

**Budget :**

[3] A titre d'exemple, pour le site du projet PFC (www.projet-pfc.net), qui inclut la base de données du projet et le concordancier, on compte plus de 10'000 consultations des pages chaque mois.

**Planification prévue :**

| | |
|---|---|
| 1er avril 2015 – 30 juin 2015 : | Phase a) élaboration de la base de données et du concordancier + phase de tests (accès aux données, requêtes, etc.). |
| 1er juillet 2015 – 31 août 2015 : | Phase b) préparation (vérification, homogénéisation et anonymisation) et saisie des premières données IPFC (projets pilotes : IPFC-espagnol, IPFC-italien, IPFC-grec et IPFC-norvégien) + préparation d'un guide de préparation des données pour l'intégration dans la base et d'utilisation de la plateforme. |

**Suivi et pérennisation :**

Une fois les données des projets pilotes saisies, le fonctionnement de la plateforme et du concordancier sera testé. La plateforme sera présentée lors des Journées IPFC2015, qui auront lieu à Paris en décembre, et les différentes équipes seront encouragées à préparer leurs données de manière à les intégrer dans la plateforme. Celle-ci sera conçue de manière à pouvoir être nourrie avec des données sans nécessiter de suivi informatique particulier.

**Coordination :** Prof. I. Racine, ELCF, Faculté des Lettres.

**Bibliographie sommaire :**


Detey, S. & Kawaguchi, Y. (2008). Interphonologie du Français Contemporain (IPFC): récolte automatisée des données et apprenants japonais. *Journées PFC: Phonologie du français contemporain: variation, interfaces, cognition,* Paris, 11-13 décembre 2008.

Detey, S., Racine, I., Kawaguchi, Y. and Eychenne, J. (éds) (à paraître). *La prononciation du français dans le monde : du natif à l'apprenant*. Paris: CLE international.

Durand, J., Laks, B. & Lyche, C. (2014). French phonology from a corpus perspective: the PFC programme, in: J. Durand, U. Gut & G. Kristoffersen (eds), *The Oxford Handbook of Corpus Phonology,* Oxford: Oxford University Press, 486-497.

Racine, I., Detey, S., Zay, F. & Y. Kawaguchi (2012). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet «Interphonologie du français contemporain» (IPFC), in: A. Kamber & C. Skupiens (éds), *Recherches récentes en FLE*, Berne: Peter Lang, 1-19.


**Annexes :**

- CV de M. Pierre Ménetrey
- Devis de M. Pierre Ménetrey
- Copie du contrat de M. Roberto Paternostro