

# **La base de données IPFC: travaux en cours**

Isabelle Racine & Sylvain Detey

ELCF, U. de Genève & U. Waseda,

Journée IPFC2016

Paris, Maison de Norvège, Cité Internationale,  
21-22 novembre 2015



WASEDA University



UNIVERSITÉ DE GENÈVE



FONDS NATIONAL SUISSE  
DE LA RECHERCHE SCIENTIFIQUE

# Plan

---

- ▶ Introduction: pourquoi une base de données IPFC?
- ▶ La base de données IPFC – phase pilote
- ▶ Perspectives



# Pourquoi une base de données IPFC?

---

- ▶ Pour que le projet IPFC acquiert le statut de corpus phonologique.

“... a collection of recordings which are available in a **computer-readable form** (e.g. wav format) and which are **accompanied by transcriptions and annotations aligned with the signal**. The transcriptions and annotations should be in standardized formats [...] or in formats easily convertible to them [...]. They should contain **essential metadata**: information about how and when the recordings were made, how the speakers were selected and who the speakers are (age, sex, social status, etc.). The transcriptions and annotations should be accompanied by a documentation explaining how they were devised. **All these requirements should be met if a corpus is to be searchable** so that analyses made by other users of the corpus can be verified and (in)validated. Finally, the collection of the data and its availability for users must follow **agreed ethical guidelines** which can vary from country to country.”

Detey, Durand, Laks & Lyche (2016)

- ▶ Voir également Gut (2014).
- 



# La base de données IPFC – phase pilote

---

- ▶ Pour atteindre cet objectif, projet pilote lancé en 2015 (financement : Université de Genève).
- ▶ Deux objectifs:
  - ▶ Pérennisation des données IPFC
  - ▶ Recherche dans les données
- ▶ Concordancier permettant des recherches par mot-clé en fonction de la LI, du locuteur ou du type de tâche (+ critères liés au locuteur: p. ex. durée des études).
- ▶ Illustration...



# La base de données IPFC – phase pilote

ipfc.unige.ch/index.php?page=concordancier

Rechercher

HOME **CONCORDANCIER** SIGNALER UNE ERREUR

Merci d'utiliser un navigateur compatible HTML5 tel que Firefox et Chrome afin que la lecture des sons soit possible.

▼ **Contexte lexical** Réinitialiser les critères

**Recherche principale** français Mot entier ?

Ajouter un critère

<b>Contexte antérieur</b>	
<b>Recherche principale</b>	Mot entier français
<b>Contexte postérieur</b>	

► **Locuteur** Réinitialiser les critères

► **Enregistrement** Réinitialiser les critères

Rechercher Réinitialiser tous les critères

© Copyright 2004-2016 - CMS Made Simple  
This site is powered by CMS Made Simple version 1.12

# La base de données IPFC – phase pilote

ipfc.unige.ch/index.php?page=concordancier

HOME **CONCORDANCIER** SIGNALER UNE ERREUR

Merci d'utiliser un navigateur compatible HTML5 tel que Firefox et Chrome afin que la lecture des sons soit possible.

▸ **Contexte lexical** Réinitialiser les critères

▸ **Locuteur** Réinitialiser les critères

▸ **Enregistrement** Réinitialiser les critères

Rechercher Réinitialiser tous les critères

80 Entrées [Premier](#) [Précédant](#) [1](#) [2](#) [3](#) [4](#) [Suivant](#) [Dernier](#) [Aller](#) Page 1 de 4

Locuteur	Texte	Tout/Rien
esgelm1	je parle le <b>français</b> et l'anglais	<input type="checkbox"/>
esgelm1	< <b>français</b> > j'ai commencé deux a- deux années avant de venir de venir en Suisse	<input type="checkbox"/>
esgelm1	j'ai étudié à l'i- à l'institut <b>français</b> à Madrid	<input type="checkbox"/>
esgelm1	euh c'est là que (j'ai commencé/je commençais) à étudier le <b>français</b>	<input type="checkbox"/>
esgelm1	<euh> on parlait que le <b>français</b> dans (le/les) cours et c'est là que j'ai vraiment commencé à apprendre le <b>français</b>	<input type="checkbox"/>
esgelm1	parce que je (je/j'ai) (je/j'ai) aimé (j'aimais/j'ai aimé) toujours le <b>français</b> sais pas pourquoi	<input type="checkbox"/>
esgelm1	même j'ai si j'ai été obligée d'apprendre l'anglais je (j'ai préféré/je préférais) de de étudier le <b>français</b> c'est parce que j'étudiais peut-être	<input type="checkbox"/>
esgelm1	et je veux mh ah après l'anglais j'ai décidé de choisir le <b>français</b>	<input type="checkbox"/>
esgelm1	<parce que> mes études simplement parce que j'aime le < <b>français</b> sais> pas j'aime comment (c-/sont) (les/le) (sons/son) et euh	<input type="checkbox"/>
	et on savait pas vraiment (c'est/ce) comment est-ce que c- (le/les) ( <b>français</b> /Français) (sonne/son/sont)	<input type="checkbox"/>

# La base de données IPFC – phase pilote

et après c'était en option mais moi j'ai choisi de faire le le mode bilingue italien-**français**

30 [79.91] - [85.4]

Télécharger le fichier son

Télécharger le fichier TextGrid

Interval: 30 [79.91] - [85.4], 19 mots

**Information sur le locuteur**

Code	it_ge_cr1		
Sexe	Femme		
Âge au moment de l'enregistrement	25		
Spécialité	Maitrise en anglais		
Le français est une langue étrangère	Oui		
Langues premières	italien		
Études de français			
	Nombre de mois	Pays - Institution	Volume horaire hebdomadaire
	106	école primaire, Pollegio, Biasca (Tessin), Suisse	2
	46	école secondaire, Biasca (Tessin), Suisse	3
		école de commerce (bilingue),	

les deux derniers alors j'ai dit je change je vais en **français** langue étrangère je me sens plus à l'aise

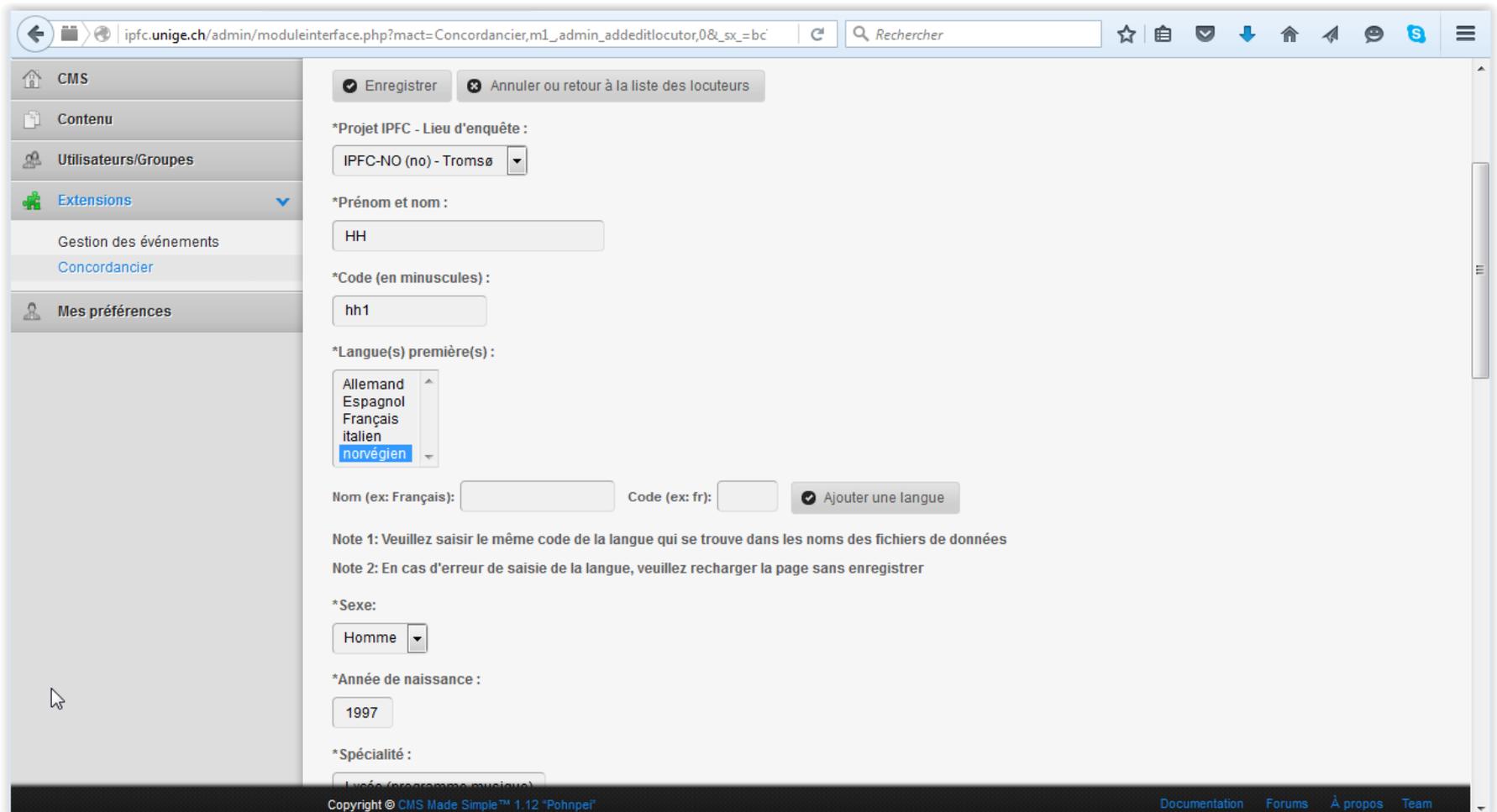
# La base de données IPFC – phase pilote

The screenshot displays the administration interface for the IPFC database. The browser address bar shows the URL: `ipfc.unige.ch/admin/moduleinterface.php?_sx_=bc7aa45468fc6b21&module=Concordancier`. The page title is "Administration de CMS Made Simple™ - IPFC - Concordancier - Concordancier". The user is logged in as "ipfc\_admin". The main content area is titled "Concordancier" and includes a navigation menu with "Projets IPFC", "Locuteurs", "Enregistrements" (selected), and "Données". A button "Ajouter un enregistrement" is visible above the table.

ID	Nom	Projet IPFC (langue) - Lieu d'enquête	Lieu de l'enregistrement	Date	Enregistré par	Transcrit par	Vérifié par	Responsable(s)	Note
5	it_ge_cr1	IPFC-IT (it) - Genève	Genève, ELCF	2014-10-15	Isabelle Racine	Roberto Paternostro	Marion Didelot	Isabelle Racine	
3	no_tr_hh1	IPFC-NO (no) - Tromsø	Tromsø, UIT Université arctique de Norvège	2014-09-17	Vincent Carretier	Danielle Stephan	Helene N. Andreassen	Helene N. Andreassen	Pour plus d'informations concernant l'enquête de Tromsø, voir <a href="http://cbille.tufs.ac.jp/ipfc/index.php?id=3">http://cbille.tufs.ac.jp/ipfc/index.php?id=3</a>
4	no_os_ib1	IPFC-NO (no) - Oslo	Oslo, Université	2014-10-15	Chantal Lyche	Danielle Stephan	Chantal Lyche et Helene N. Andreassen	Chantal Lyche, Helene N. Andreassen	Pour plus d'informations concernant l'enquête d'Oslo, voir <a href="http://cbille.tufs.ac.jp/ipfc/index.php?id=3">http://cbille.tufs.ac.jp/ipfc/index.php?id=3</a>

Copyright © CMS Made Simple™ 1.12 "Pohnpei" | Documentation | Forums | À propos | Team

# La base de données IPFC – phase pilote



The screenshot shows a web browser window with the URL `ipfc.unige.ch/admin/moduleinterface.php?mact=Concordancier,m1_admin_ajoutlocutor,0&_sx_=bc`. The browser's address bar also contains a search field with the text "Rechercher".

On the left side, there is a navigation menu with the following items:

- CMS
- Contenu
- Utilisateurs/Groupes
- Extensions (with a dropdown arrow)
- Gestion des événements
- Concordancier
- Mes préférences

The main content area displays a form for adding a new participant. At the top, there are two buttons: "Enregistrer" (checked) and "Annuler ou retour à la liste des locuteurs".

The form fields are as follows:

- \*Projet IPFC - Lieu d'enquête : A dropdown menu with "IPFC-NO (no) - Tromsø" selected.
- \*Prénom et nom : A text input field containing "HH".
- \*Code (en minuscules) : A text input field containing "hh1".
- \*Langue(s) première(s) : A dropdown menu with "norvégien" selected. The list of languages includes Allemand, Espagnol, Français, italien, and norvégien.
- Nom (ex: Français): [ ] Code (ex: fr): [ ]  Ajouter une langue

Below the form, there are two notes:

Note 1: Veuillez saisir le même code de la langue qui se trouve dans les noms des fichiers de données

Note 2: En cas d'erreur de saisie de la langue, veuillez recharger la page sans enregistrer

Additional form fields include:

- \*Sexe: A dropdown menu with "Homme" selected.
- \*Année de naissance : A text input field containing "1997".
- \*Spécialité : A text input field containing "Lyce (programme musique)".

At the bottom of the page, there is a footer with the text "Copyright © CMS Made Simple™ 1.12 'Pohnpei'" and navigation links for "Documentation", "Forums", "À propos", and "Team".

# Perspectives

---

## ▶ A faire:

- ▶ Tester le concordancier et corriger les premiers problèmes
- ▶ Introduire davantage de données (corpus complet alémanique et tessinois + données norvégiennes) et tester le concordancier
- ▶ Ajuster la partie accessible des métadonnées

## ▶ Procédure d'intégration des données:

- ▶ Fait à Genève
- ▶ Nécessaire: métadonnées (locuteur + enquête) + 2 fichiers par tâche et par locuteur (son + grid avec transcription orthographique selon conventions IPFC)
- ▶ Seules des données vérifiées et anonymisées (en fonction des exigences de chaque pays) seront entrées dans la base
- ▶ Un numéro DOI par enquête (avec renvoi aux références à citer pour l'utilisation de chaque point d'enquête)

## ▶ Les accès:

- ▶ Public: 1-2 locuteurs par enquête
  - ▶ Corpus IPFC-recherche (à terme): consultation après accord
  - ▶ Corpus IPFC-interne: un accès par équipe pour recherche dans ses propres données
- 

