



Développement de la plateforme Dolmen pour IPFC

Julien EYCHENNE (*Hankuk University of Foreign Studies*),
Sylvain DETEY (*U. Waseda & U. de Rouen*), ***Isabelle RACINE***
(*ELCF, U. de Genève*), ***Laurie BUSCAIL*** (*U. de Toulouse*)

Journées IPFC2012: corpus, liaison, interprétation
Maison de Norvège, CIU, Paris, 10-11 décembre 2012

Plan

I. Il était une fois Dolmen...

II. Puis vint IPFC

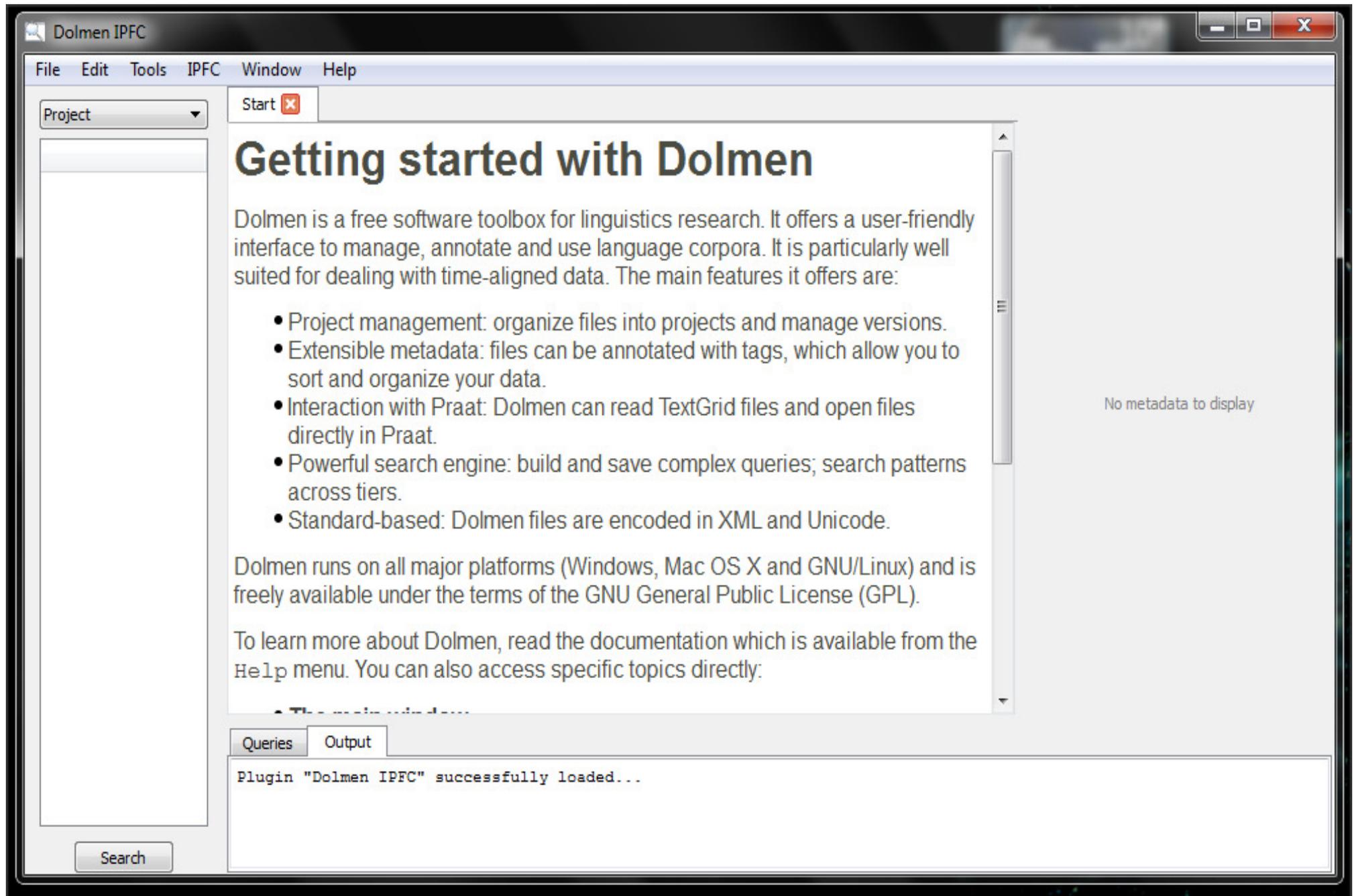
III. Et ainsi Dolmen-IPFC

IV. Perspectives

I. Il était une fois Dolmen...

- Le projet PFC: comptage, comparaison
 - classeur schwa, classeur liaison, comparateur...
- Plusieurs outils écrits spécifiquement pour PFC
 - plateforme PFC
- Besoin identiques ou similaires dans d'autres projets travaillant avec des transcriptions et codages alignés sur le signal (ex : PAC)
 - Dolmen

DOLMEN



Dolmen

- Dolmen est un concordancier pour les transcriptions et codages alignés sur le signal
- Open source (licence GPL) / Multi-plateforme (Windows, Mac OS X, Linux)
- Compatible avec fichiers Praat et WaveSurfer / Écrit en C/C++
- Moteur de recherche avancé : expressions rationnelles, recherche multi-niveaux, sauvegarde de requêtes
- **Interaction avec Praat**
 - Lecture/écriture de fichiers TextGrid
 - Ouverture de fichiers dans Praat (via sendpraat).
- Téléchargement :
www.julieneychenne.info/dolmen

II. Puis vint IPFC

Objectifs de IPFC:

- similaire: description à large échelle
- différent: évaluation des productions

Besoins de IPFC:

- automatisation
- codage (vs transcription phonétique)
- plateforme comparable à celle de PFC

III. Et ainsi Dolmen-IPFC

Décembre 2011:

- Présentation du code voyelles nasales
- Amorce de Dolmen-IPFC
- Comparaison Tests perceptifs / Codage & Dolmen

En 2012:

Julien Eychenne : IPFC-néerlandais, IPFC-coréen
Développement dans le cadre d'un projet KAKEN
(étude longitudinale en production/perception – Japon)

Trois interfaces quasi-achevées:

- voyelles nasales
- voyelles orales
- liaison

Fonctionnalités soutenues par le Kaken

- Export des résultats au format CSV (import dans Excel, R, SPSS, etc.)
- Extensions utilisateurs :
 - définition de grammaires de recherche pour faciliter la recherche
 - compatible avec tout système de codage pouvant être décrit à l'aide d'expressions rationnelles (automate à états finis)
- Extension IPFC pour les voyelles nasales et orales, et la liaison
- Comparaison inter-annotateurs

Code IPFC – Grammaire de recherche de Dolmen

La grammaire autorise :

- Un nombre arbitraire de **champs**
- Un nombre arbitraire de **valeurs** pour chaque champ
- Diverses options (tire de recherche, sensible à la casse, etc...)

- **Exemple: les voyelles nasales (6 champs)**
(Detey, Racine et Kawaguchi à paraître dans Durand, Kristoffersen & Laks, *La phonologie du français : normes, périphéries et modélisation*, Hommage à Chantal Lyche)

- 1= Voyelle cible (2 valeurs, 2 digits)
- 2= Contexte gauche (3 valeurs, 2 digits)
- 3= Contexte droit (5 valeurs, 3 digits)
- 4= Evaluation nasalité (4 valeurs, 1 digit)
- 5= Evaluation timbre (ou catégorie si V orale) (2 valeurs, 2 digits)
- 6= Evaluation appendices consonantiques (3 valeurs, 2 digits)

Illustrations

Les voyelles nasales

Les voyelles orales (en phase de finalisation)

La liaison (voir présentation Racine & Detey demain)

Illustrations

Le code des voyelles orales :

1= Cible (7 valeurs – 1 digit)

2= Contexte phonologique gauche (3 valeurs – 3 digits)

3= Contexte phonologique droit (2 valeurs – 2 digits)

4= Qualité globale de réalisation (4 valeurs – 3 digits)

Ex : le vi1_20F_2L_100llage de Beaulieu est en grand émoi



Résultats en cours de collecte:

Données codées :

Total : 1170 [i] + 464 [y] + 412 [u] + 1459 [a] = 3505 V

Hispanophones:

Mots (Répétition & Lecture) : 80 (40 L + 40 R) pour Y/U;

897 texte (4 sujets)

Japonophones :

Mots (Répétition & Lecture) : 80 (40 L + 40 R) pour Y/U;

695 texte (3 sujets) ; 1760 guidée (3 sujets)

Résultats illustratifs préliminaires (nbr occ.) :

[y] mieux réalisée en R qu'en L (hispanophones et japonophones)

[u] mieux réalisée en R pour les hispanophones mais en L pour les japonophones

V les mieux réalisées (Texte & Conversation) : [a] (50% guidée japonophone ; 40% texte japonophone ; 41% texte hispanophone)

V les moins bien réalisées (Texte & Conversation) : [u] (76% guidée japonophone ; 47% texte japonophone) [y] (83% texte hispanophone)

Perspectives

Jusqu'à présent : requêtes manuelles

A présent : catégories

Chantiers :

- Automatisation du calcul des pourcentages
- Intégration de fonctions de comparaison supplémentaires (p. ex. inter-tâche/groupes/sujets)
- Tests du système de comparaison inter-annotateurs
- Poursuivre l'automatisation des fonctions statistiques afin d'obtenir facilement le profil phonologique des locuteurs (30 sujets X 4 séances X 6 tâches dans Kaken)
- Diffusion de Dolmen-IPFC dans la communauté IPFC

Remerciements

Les travaux présentés ici ont bénéficié du soutien:

Au Japon (S. Detey):

de la Société Japonaise pour la Promotion de la Science (JSPS)
(Grant-in-Aid for Scientific Research B n° 23320121)

de l'Université Waseda (Special Research Grant, 2011B-297)

En Suisse (I. Racine):

du Fonds national suisse de la recherche scientifique (132144/1)

de la Faculté des Lettres de l'Université de Genève

de la Société Académique de Genève (Fonds Ch. Bally)

Remerciements à Y. Kawaguchi, J. Durand, B. Laks, C. Lyche, M. Kondo, J. Eychenne, L. Buscail et tous les collaborateurs de IPFC.

VOTRE ATTENTION !



Développement de la plateforme Dolmen pour IPFC

*Julien EYCHENNE (Hankuk University of Foreign Studies),
Sylvain DETEY (U. Waseda & U. de Rouen), Isabelle RACINE
(ELCF, U. de Genève), Laurie BUSCAIL (U. de Toulouse)*

Journées IPFC2012: corpus, liaison, interprétation
Maison de Norvège, CIU, Paris, 10-11 décembre 2012