

Compte rendu de la réunion du 17 septembre

Noëlle Serpollet, Yvan Stroppa, Akiko Takemura-Pellard, Gabriel Bergounioux

1. Présentation du site Passy

Yvan présente le site « Passy » (en hommage à l'un des créateurs de l'API) comprenant quatre modules :

- « Accueil / Statut » où apparaîtront les chercheurs impliqués et l'actualité des travaux : ce module permet la visualisation et la consultation des informations concernant le contexte et l'environnement du projet.
 - « Consultation » pour formuler des requêtes et extraire des données à partir d'une interrogation par lemme, par forme graphique – éventuellement par la prononciation (en lettres ou en symboles API) –, par une sélection dans les informations sociales disponibles (âge, sexe, origine géographique, niveau d'étude, profession) et par les caractéristiques de l'enregistrement (type, date de collecte).
 - « Contribution » destiné à l'incrémentation des données soit par dépôt de corpus déjà constitués (à condition de renseigner les métadonnées), soit par enregistrement direct dans une perspective de développement en sciences participatives et d'enquêtes conduites sur place.
 - « Supervision » : en accès restreint, comme le précédent, ce module assure le contrôle de qualité et le suivi des apports et des corrections effectuées.
- Le site est d'ores et déjà opérationnel sans pour autant être ouvert. Il sera adapté au fur et à mesure que le dépôt des données montrera la nécessité d'aménagements.

2. Construction d'un groupe témoin

Un premier test implique de retenir seulement un nombre limité de locuteurs et un nombre limité de mots. Les données ont été récupérées à partir de la version en ligne d'ESLO2 qui comprend 84 entretiens dont 82 sont exploitables.

À l'écoute, ont été repérés six locuteurs ayant un accent plus ou moins marqué, quatre du Midi (Nice et le Sud-Ouest), une personne d'origine antillaise et une personne d'origine maghrébine.

Pour un contraste minimum entre des locuteurs issus du Val-de-Loire, on a retenu :

- quatre hommes et quatre femmes,
 - quatre locuteurs de moins de quarante ans et quatre de plus de cinquante ans,
 - quatre témoins ayant un CSP+ et quatre un CSP-,
- soit huit témoins pour lesquels on a vérifié une bonne qualité globale de l'enregistrement (avec des problèmes techniques partiels pour deux d'entre eux), tant du point de vue matériel que de l'interaction.

Liste des témoins Val-de-Loire :

Code d'enregistrement	Code témoin	Sexe	Age	CSP
1006	OS6	F	> 50	-
1010	GK11	H	< 50	+
1023	YR23	F	> 50	-
1032	FE32	H	> 50	+
1038	AJ38	F	< 50	-
1042	PM42	F	< 50	+
1057	UZ57	H	< 20	-
1065	RK65	H	> 50	+

Liste des autres témoins

Code d'enregistrement	Code témoins	Département ou pays d'origine
1043	AN43	Aude
1044	DE44	Iran (> Gironde)
1061	AJ61	Gironde
1069	FJ944	Alpes maritimes
1083	MF363	Martinique
n° ?	KG276	Algérie

3. Établissement d'une liste de mots

La discussion a été consacrée au choix d'une liste de mots à partir d'un classement par fréquence des graphies relevées dans ESLO, de « euh » (155.484 occurrences) aux 2.300 hapax. Comme il s'agit de graphies, un même nom pourra apparaître deux fois, au singulier et au pluriel, un adjectif quatre fois, etc. Aucune lemmatisation n'a été effectuée.

L'idée est de retenir des termes présentant les trois caractéristiques suivantes :

1. ne pas être monosyllabique,
2. être susceptible de variation,
3. avoir une fréquence élevée.

Les consignes sont partiellement contradictoires. Les cent graphies les plus attestées (nombre d'occurrences > 7.500) ne comprennent que 8 dissyllabes, soit par ordre décroissant : *alors*, *enfin*, *était*, *(d')accord*, *Orléans*, *beaucoup*, *voilà*, *avait*. Aussi a-t-il fallu descendre dans la liste pour retenir les mots les plus susceptibles de retenir l'attention, les polysyllabes les plus fréquents étant des adverbes en *-ment*.

Un premier tri s'était arrêté sur dix mots dont le moins fréquent présentait 400 occurrences. Si on suppose une répartition égale entre ESLO1 et ESLO2 – ce qui n'est probablement pas exact – et, dans ESLO2, entre les 82 entretiens et les autres enregistrements, on attend 200 occurrences dans ESLO2 dont 70 figureraient dans les entretiens, soit moins d'une occurrence en moyenne par locuteur, sachant qu'il faut prendre en compte les propos de l'enquêteur voire ceux d'autres participants présents lors de la rencontre (un conjoint par exemple).

Finalement, ce sont neuf formes qui pourront être retenues pour un premier examen :

Item	Catégorie	Occurrences	Autres occurrences
Orléans	Nom	10699	17
Peut-être	Adverbe	6932	
Maintenant	Adverbe	6415	
Justement	Adverbe	2366	
Évidemment	Adverbe	2316	
Problème	Nom	1793	
Difficile	Adjectif	1583	141 (+ 80 « difficilement »)
Ouvrier	Nom	1751	1
Enseignement	Nom	1459	31 (+ 182 « renseignement(s) »)

Avec 1.500 occurrences, on arrive à trois occurrences en moyenne par locuteur, ce qui devrait néanmoins laisser des témoins non représentés.

Prochaine réunion en octobre.