

Compte rendu de la réunion du 17 novembre

Gwenaëlle Fabre, Badreddine Hamma, Noëlle Serpollet, Yvan Stroppa, Gabriel Bergounioux

1. Retour sur les critères de sélection des unités à étudier

Rappel des critères retenus pour la construction d'une maquette du dictionnaire à partir de quelques termes considérés comme à même d'éclairer la méthodologie à mettre en place.

- 1°) Mots de deux syllabes ou plus (ce qui le plus souvent équivaut à deux morphèmes),
- 2°) Une fréquence élevée en soustrayant ce qui résulte d'un effet corpus soit dans les thèmes (« Orléans » n'est pas le quatre-vingt-deuxième mot le plus courant en français), soit dans l'interaction (« évidemment » n'est pas le deux cent treizième mot le plus fréquent).
- 3°) Une dispersion répartie entre différents locuteurs.

Outil Il y a possibilité de construire un programme qui discrimine et identifie automatiquement combien de locuteurs ont employé telle forme, en éliminant du calcul les enquêteurs.

- 4°) Une seule définition correspondante, autrement dit exclusion des homonymes.

Question quel traitement pour les homonymes discriminables par la prosodie comme « maintenant » entre l'emploi de ponctuant et sa fonction adverbiale équivalant à un complément circonstanciel de temps ?

2. Un exemple : « ouvrier »

Les caractéristiques morpho-phonologiques du mot : attaque vide, présence d'un groupe complexe requérant l'épenthèse d'un /i/ soit /u-vri-jé/, fragilité du vocalisme (cf. *infra*), opacité partielle de la construction morphologique.

Question L'épenthèse a reçu une explication phonétique qui en justifiait la présence par la difficulté de réaliser à la suite une occlusive, une constrictive et une semi-voyelle avant la voyelle finale. Il semble au contraire que ce soit une extension d'une réalisation où le /i/ est présent dans le morphème comme dans « crier » ou « briller » alors que, si le syntagme est réparti sur deux mots successifs, la réalisation se fait sans épenthèse : cf. « i(l) trouv(e) rien », « i(l) sauv(e) rien ».

L'analyse a une visée structurale :

- en phonologie : construire une compétence d'auditeur (et non de locuteur) : quelles sont les variations intégrées dans le système du locuteur à partir de quoi il est à même d'effectuer les neutralisations nécessaires à l'identification des mots, éventuellement de leur assigner une interprétation sociale (de lieu, d'âge, de niveau d'études...) à la différence des formes ?
- en morphologie : au-delà d'une étude mot à mot, observer la multiplicité de réalisation de ses morphèmes.

Verbe	N. d'agent	N. d'action	Adj.	Hors champ ?		
opérer	opérateur	opération	opératoire	opéra		
œuvrer	(manœuvre)	œuvre	œuvré	hors	d'œuvre,	chef
d'œuvre						

(ouvrier) ouvrier ouvrage ouvrage ouvrierable

3. Sélection de termes

Pour une première étude, s'impose la nécessité de sélectionner des mots correspondant aux critères définis *supra* et qui soient attestés en nombre limité (30 à 60 occurrences). En restreignant l'inventaire aux quatre-vingts entretiens exploitables d'ESLO2, il s'avère que les 2.700 occurrences d'*Orléans* (environ 10.000 dans l'ensemble de tous les enregistrements d'ESLO 1 + ESLO 2) ou les 520 de *problème(s)* (1.800 dans tout le corpus) exigeraient trop de temps et ne pourront être traités qu'à partir du moment où des procédures automatiques auront permis d'en faire l'analyse.

Question On note une extrême disparité entre le nombre total de termes et ceux attestés dans l'échantillon retenu qui, le plus souvent, en conservent entre le tiers et le quart mais *enseignement(s)*, qui a environ 1.500 occurrences, n'en présente pourtant que 5 dans l'échantillon. C'est une information qui appelle réflexion.

Pour une première étude, cinq termes seront examinés lors de la prochaine séance de travail :

- *création*
- *difficulté*
- *événement*
- *heureusement*
- *ingénieur*

Prochaine réunion en décembre

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

4. 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.