

# **Rapport de stage L3**

## **Laboratoire ERTIM**

Ousseynou Gueye

June 1, 2018

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Définition des objectifs et description du corpus</b> | <b>2</b> |
| 1.1      | Présentation et Constitution du corpus . . . . .         | 2        |
| 1.1.1    | Pourquoi ce corpus ? . . . . .                           | 2        |
| 1.1.2    | Description du corpus . . . . .                          | 2        |
| 1.1.3    | Choix de normalisation . . . . .                         | 3        |
| 1.2      | Hypothèse principale . . . . .                           | 3        |
| 1.3      | Choix des outils . . . . .                               | 3        |
| 1.3.1    | Iramuteq . . . . .                                       | 3        |
| 1.3.2    | TXM . . . . .  | 3        |
| 1.3.3    | Arbitrage . . . . .                                      | 3        |
| <b>2</b> | <b>Résultats</b>   | <b>4</b> |
| 2.1      | Spécificités et hypothèses de départ . . . . .           | 4        |
| 2.1.1    | Discussions et interpretation . . . . .                  | 4        |
| 2.1.2    | Ouverture . . . . .                                      | 5        |
| 2.2      | Champ lexicaux et spécificités . . . . .                 | 5        |
| 2.2.1    | Vocabulaire mecquois . . . . .                           | 5        |
| 2.2.2    | Vocabulaire médinois . . . . .                           | 6        |
| 2.2.3    | Analyses . . . . .                                       | 7        |
| <b>3</b> | <b>Points positifs, limites et ouvertures</b>            | <b>9</b> |

# CHAPTER 1

## DÉFINITION DES OBJECTIFS ET DESCRIPTION DU CORPUS

*Étude de l'évolution de certaines formes interpellatives dans un texte arabe (le Coran) selon un découpage chronologique (les deux grandes périodes de sa révélation (mecquoise et médinoise)).*

### 1.1 Présentation et Constitution du corpus

Notre corpus est constitué du texte coranique.

**NB :** le choix du corpus, ainsi que le travail qui va suivre n'ont aucun but religieux. Notre texte nous intéresse avant tout pour sa valeur littéraire.

#### 1.1.1 Pourquoi ce corpus ?

Pour au moins deux raisons :

- c'est un corpus large mais assez restreint pour être étudiés.
- c'est un corpus connu avec des études depuis des siècles.

#### 1.1.2 Description du corpus

Le corpus est composé de :

- 114 chapitres de longueur très variables (de 3 à 286 versets).
- Ces chapitres sont réparties en deux parties, correspondant aux 22 ans de sa révélation : mecquoise (610-622), médinoise (622-632).
- Environ 85 000 mots graphiques.
- 349 formules interpellatives commençant par **﴿** qui est l'outil du vocatif en arabe.

### 1.1.3 Choix de normalisation

- Nous avons enlevé tous les signes de vocalisation. Pour une étude plus poussée, peut-être seraient-ils importants, mais dans notre cas, ils nous ajoutent des contraintes non-justifiées.
- Les formules interpellatives sont toujours formées d'au moins deux mots graphiques. Or les calculs de spécificités ne se font que sur un mot graphique, donc nous avons remplacé des formules par des lettres selon le schéma suivant :

| Formule arabe                 | remplacé par : | Trad formule             | Nombre |
|-------------------------------|----------------|--------------------------|--------|
| يا أَيُّهَا النَّاسُ          | A              | Ô vous les humains       | 8      |
| يا أَيُّهَا الَّذِينَ آمَنُوا | B              | Ô vous qui croyez        | 89     |
| يا أَيُّهَا النَّبِيُّ        | C              | Ô Toi le Prophète        | 13     |
| يا أَيُّهَا + x               | D              | Ô                        | 17     |
| يا أَهْلَ الْكِتَابِ          | E              | Ô vous les gens du livre | 12     |
| يا                            | non-remplacée  | Ô                        | 189    |

## 1.2 Hypothèse principale

Nous pensons que la formule (*Ô vous les humains* - **A**) devrait être caractéristique de la 1ère période, car étant au début de la révélation, il n'y pas de communauté, là où l'on devrait voir plus de (*Ô vous qui croyez* - **B**) à partir de la seconde période (médinoise).

## 1.3 Choix des outils

Nous avons hésité et testé deux logiciels (Iramuteq et TXM).

### 1.3.1 Iramuteq

- Avantage : prise en charge utf8, capacité d'importer de txm.
- Inconvénient : format d'entrée non pratique. Le problème peut être résolu grâce à un script, cependant il aurait fallu une longue procédure à chaque changement

### 1.3.2 TXM

- Avantage : prise en charge utf8, mais surtout la capacité à utiliser un fichier csv pour les informations sur le corpus.
- Inconvénient : impossibilité de changer la police, et présentation des résultats de gauche à droite, ce qui n'est pas commode pour de l'arabe.

### 1.3.3 Arbitrage

Notre choix s'est porté sur TXM, pour l'avantage que nous avons énoncé.

## 2.1 Spécificités et hypothèses de départ

Voici le premier tableau des spécificités, classées selon l'ordre alphabétique, ce qui nous permet de nous occuper de nos lettres (A - E)

| Units ^ | Frequency T 75445 | mecque t=47261 | score | medina t=28184 | score |
|---------|-------------------|----------------|-------|----------------|-------|
| ...     | 6236              | 4621           | 88.6  | 1615           | -88.6 |
| A       | 19                | 9              | -0.9  | 10             | 0.9   |
| B       | 89                | 0              | -38.1 | 89             | 38.1  |
| C       | 13                | 0              | -5.6  | 13             | 5.6   |
| D       | 20                | 15             | 0.7   | 5              | -0.7  |
| E       | 12                | 0              | -5.1  | 12             | 5.1   |
| ...     | 196               | 160            | 8.3   | 36             | -8.3  |

Figure 2.1: Spécifités par ordre alphabétique.

### 2.1.1 Discussions et interpretation

#### 1. Cas de A (Ô vous les humains)

Notre hypothèse est invalidee, avec quasiment le même nombre d'occurence dans les deux parties (9 vs 10)

Lecture possible : même ds la deuxième partie de sa révélation, le texte coranique, qui a une vocation universelle continue de s'intéresser aux humains en général, croyants ou pas.

#### 2. Cas de B (Ô vous les croyants)

Notre hypothèse est validée. Sur les 89 occurrences dans le corpus, les 89 (100%) se trouve dans la deuxième partie.

Lecture possible : dans cette seconde partie, nous avons une communauté naissante, ce qui justifie la mention des croyants.

## 2.1.2 Ouverture

Après nous être penchés sur nos deux hypothèses de base, nous avons voulu aller plus loin dans l'exploration du vocabulaire de ces deux parties.

## 2.2 Champ lexicaux et spécificités

Pour cela, nous allons nous intéresser à la spécificités selon son étude naturelle qui est celle du score.

Nous regarderons les 35 termes les spécifiques à chaque partie, tout en proposant notre explication sur les phénomènes qui nous semblent les plus intéressants.

### 2.2.1 Vocabulaire mecquois

| Units   | Frequency T 75445 | mecque t=47261 | score | medine t=28184 | score |
|---------|-------------------|----------------|-------|----------------|-------|
| ...     | 6236              | 4621           | 88.6  | 1615           | -88.6 |
| قال     | 416               | 339            | 16.4  | 77             | -16.4 |
| رَبِّكَ | 220               | 192            | 15.6  | 28             | -15.6 |
| رَبِّي  | 94                | 89             | 12.4  | 5              | -12.4 |
| هذا     | 190               | 163            | 11.9  | 27             | -11.9 |
| إِنَّهُ | 147               | 129            | 11.1  | 18             | -11.1 |
| رَبِّ   | 130               | 114            | 9.8   | 16             | -9.8  |
| كُنَّا  | 63                | 60             | 8.8   | 3              | -8.8  |
| يا      | 196               | 160            | 8.3   | 36             | -8.3  |
| إِنِّي  | 131               | 112            | 8.2   | 19             | -8.2  |
| وقال    | 85                | 76             | 7.6   | 9              | -7.6  |
| مبين    | 84                | 75             | 7.4   | 9              | -7.4  |

Figure 2.2: Spécificités de la période mecquoise 1

|             |     |     |     |    |      |
|-------------|-----|-----|-----|----|------|
| ولقد        | 129 | 109 | 7.4 | 20 | -7.4 |
| يومئذ       | 65  | 60  | 7.4 | 5  | -7.4 |
| قومه        | 36  | 36  | 7.3 | 0  | -7.3 |
| إِنَّا      | 156 | 128 | 7.0 | 28 | -7.0 |
| الرَّحْمَنُ | 157 | 128 | 6.7 | 29 | -6.7 |
| كانوا       | 229 | 179 | 6.5 | 50 | -6.5 |
| كَلَّا      | 38  | 37  | 6.3 | 1  | -6.3 |
| موسى        | 129 | 106 | 6.0 | 23 | -6.0 |
| أم          | 124 | 102 | 5.8 | 22 | -5.8 |
| يوم         | 217 | 168 | 5.7 | 49 | -5.7 |
| فرعون       | 67  | 59  | 5.5 | 8  | -5.5 |
| أنا         | 49  | 45  | 5.5 | 4  | -5.5 |

Figure 2.3: Spécificités de la période mecquoise 2

| ويوم    | 44  | 40  | 4.6 | 4   | -4.6 |
|---------|-----|-----|-----|-----|------|
| إلا     | 664 | 466 | 4.6 | 198 | -4.6 |
| نحن     | 65  | 56  | 4.6 | 9   | -4.6 |
| أولم    | 34  | 32  | 4.6 | 2   | -4.6 |
| يوسف    | 22  | 22  | 4.5 | 0   | -4.5 |
| أكثرهم  | 38  | 35  | 4.4 | 3   | -4.4 |
| وكذلك   | 38  | 35  | 4.4 | 3   | -4.4 |
| أرسلنا  | 51  | 45  | 4.4 | 6   | -4.4 |
| الإنسان | 59  | 51  | 4.3 | 8   | -4.3 |
| الفلك   | 21  | 21  | 4.3 | 0   | -4.3 |
| يروا    | 27  | 26  | 4.3 | 1   | -4.3 |
| الحمد   | 26  | 25  | 4.1 | 1   | -4.1 |

Figure 2.4: Spécificités de la période mecquoise 3

## 2.2.2 Vocabulaire médinois

| Units  | Frequency T 75445 | mecque t=47261 | score  | medine t=28184 | score |
|--------|-------------------|----------------|--------|----------------|-------|
| الله   | 2265              | 894            | -113.1 | 1371           | 113.1 |
| والله  | 240               | 33             | -54.7  | 207            | 54.7  |
| B      | 89                | 0              | -38.1  | 89             | 38.1  |
| ورسوله | 66                | 2              | -24.5  | 64             | 24.5  |
| بالله  | 139               | 35             | -18.6  | 104            | 18.6  |
| فإن    | 98                | 24             | -13.8  | 74             | 13.8  |
| منكم   | 107               | 28             | -13.7  | 79             | 13.7  |
| تكذبان | 31                | 0              | -13.3  | 31             | 13.3  |
| ريكما  | 33                | 2              | -10.9  | 31             | 10.9  |
| وائقوا | 42                | 5              | -10.9  | 37             | 10.9  |
| سبيل   | 87                | 24             | -10.5  | 63             | 10.5  |
| آلاء   | 34                | 3              | -10.1  | 31             | 10.1  |

Figure 2.5: Spécificités de la période médinoise 1

| Units  | Frequency T 75445 | mecque t=47261 | score | medine t=28184 | score |
|--------|-------------------|----------------|-------|----------------|-------|
| قلوبهم | 65                | 15             | -10.0 | 50             | 10.0  |
| جناح   | 26                | 1              | -9.5  | 25             | 9.5   |
| فبأئ   | 35                | 4              | -9.3  | 31             | 9.3   |
| عليم   | 106               | 35             | -9.2  | 71             | 9.2   |
| الرسول | 45                | 8              | -9.1  | 37             | 9.1   |
| كفروا  | 189               | 78             | -8.7  | 111            | 8.7   |
| الحرام | 23                | 1              | -8.2  | 22             | 8.2   |
| واليوم | 23                | 1              | -8.2  | 22             | 8.2   |
| غفور   | 52                | 12             | -8.2  | 40             | 8.2   |
| الآخر  | 29                | 3              | -8.1  | 26             | 8.1   |
| الذين  | 722               | 378            | -8.0  | 344            | 8.0   |
| عطیما  | 22                | 1              | -7.8  | 21             | 7.8   |
| عليما  | 22                | 1              | -7.8  | 21             | 7.8   |

Figure 2.6: Spécificités de la période médinoise 2

| Units    | Frequency T 75445 | mecque t=47261 | score | medine t=28184 | score |
|----------|-------------------|----------------|-------|----------------|-------|
| واعلموا  | 18                | 0              | -7.7  | 18             | 7.7   |
| وانتم    | 45                | 10             | -7.4  | 35             | 7.4   |
| المؤمنين | 78                | 25             | -7.4  | 53             | 7.4   |
| النساء   | 24                | 2              | -7.4  | 22             | 7.4   |
| أو       | 281               | 132            | -7.2  | 149            | 7.2   |
| مريم     | 34                | 6              | -7.0  | 28             | 7.0   |
| حكيمًا   | 16                | 0              | -6.8  | 16             | 6.8   |
| بين      | 16                | 0              | -6.8  | 16             | 6.8   |
| يحب      | 39                | 9              | -6.3  | 30             | 6.3   |
| رحيم     | 61                | 19             | -6.2  | 42             | 6.2   |
| آمنوا    | 174               | 77             | -6.2  | 97             | 6.2   |
| عليكم    | 164               | 72             | -6.1  | 92             | 6.1   |

Figure 2.7: Spécificités de la période médinoise 3

### 2.2.3 Analyses

On remarque entre autre :

- Un vocabulaire du récit privilégié dans la partie mecquoise.

Nous disons cela à cause de la présence de (قال - Il dit) présent 424 fois dans cette période soit 84.6% des occurrences totales (501).

- Un discours eschatologique accentué dans la partie mecquoise.

On voit que les termes (يومئذ et اليوم) faisant référence à la fin des temps et le jugement dernier apparaissent à 81% (228/282) dans la partie mecquoise.

- Une variation de personne

La période mecquoise est caractérisé par plusieurs termes renvoyant à la première personne (singulier ou pluriel) comme (كنا - nous sommes), (أنا, إني, نحن - pronoms personnels).

Par contre, dans la période médinoise, les termes renvoient à la deuxième ou troisième personne. Notons juste les verbes à l'impératif (اعلموا - sachez, اتقوا - ayez crainte/foi).

Pouvons-nous y voir un déplacement de sujet. Si la révélation commence par présenter Dieu qui parle à la première personne, l'accent est ensuite mis sur la communauté, les gens donc le non-je.

- L'évolution de l'appellation de la même entité (Dieu).

Dans la partie mecquoise, nous avons la présence du mot (رب - Seigneur) accompagné de pronoms personnels (Ton, mon, votre ...) 359 fois soit 83 % des occurrences (432).

Puis le terme semble évolué. De (Seigneur - رب) on glisse vers le mot (الله). On voit que 63% (1682/2644) des occurrences ont lieu dans la partie médinoise.



Si le premier est plus intime, affectueux comme appellation, on passe dans le second cas à un terme neutre voire institutionnel.

- Un vocabulaire montrant une communauté et ses différents interlocuteurs dans la période médinoise.

On voit apparaître comme spécifique à la période médinoise plusieurs termes nous indiquant la présence d'une communauté :

D'une part le (message - رسول) et les (croyants - مؤمنون) ainsi que le verbe (croire - آمن). De l'autre, des verbes relatifs à la négation (كفر - nier) ou encore (accuser de mensonge - كذب). Mais aussi des particules telles que (منكم - de vous), (عليكم - sur vous).

## CHAPTER 3

### POINTS POSITIFS, LIMITES ET OUVERTURES

À l'issue de ce travail, le principal à garder est à notre avis que les outils dont nous disposons en TAL peuvent servir à traiter de l'arabe. Il est certes vrai que l'analyse n'est pas encore au même niveau que pour les corpus en caractères latins (surtout au niveau de l'étiquetage morphosyntaxique), cependant en ce qui concerne la textométrie de base, les outils font le travail.

Quant aux limites de notre travail :

- Nous sommes conscients qu'il puisse y avoir un biais dans la réception. Au fil des siècles, les mots changent de significations, ce qui fait que nous pouvons voir un réseau de sens qui n'était pas forcément celui du public de base. Mais cela mènerait à un débat plus philosophique que technique autour de ce que H. R. Jauss appelle l'esthétique de la réception.
- Dans le même sillage, la question des limites de l'interprétation se pose. Une autre personne, devant les mêmes données aurait probablement eu des conclusions/observations différents.
- Enfin, nous n'avons analysé qu'une infime échantillon (environ 80 mots graphiques sur 15 000 soit 0.5%). Continuer aurait renforcé/infirmé nos observations. Mais le principal n'était pas là pour ce travail ci.