

Rapport sur les travaux de mon mémoire

Theophile Miaillhe

1/04/22

1 Le travail accompli :

Grâce à la disponibilité en format CSV de la base de données des morts en opération extérieure, je n'ai pas eu besoin de faire un scrapping du site du SHD. J'ai téléchargé les données directement de ce lien: https://www.memoiredeshommes.sga.defense.gouv.fr/fr/arkotheque/navigation_facette/index.php?f=opendata. J'ai procédé à un traitement de la base « Militaires décédés sur les théâtres d'opérations extérieurs (1905-1962) » avec la librairie **Pandas** de **Python**. La base comportait 20.059 entrées qui correspondaient chacune à un militaire mort.

Out[6]:

	id_confli_t_intitule	id_sous_confli_t_intitule	id_famille_cote_intitule	sous_serie	serie	article	nom	prenom	nom_autre	pseudonyme	...	disparition_da
0	Théâtres d'opérations extérieurs	Afrique du Nord		NaN	NaN	NaN	AB EL OUAHAB BEN SAID BEN BELGACEM	NaN	NaN	NaN	...	
1	Théâtres d'opérations extérieurs	Levant		AC	40.0	R 4063.0	ABABOU	Abdelkader		NaN	NaN	...
2	Théâtres d'opérations extérieurs	Levant		AC	40.0	R 4063.0	ABABOU	Abdelkader		NaN	NaN	...
3	Théâtres d'opérations extérieurs	Afrique du Nord		AC	40.0	R 4034.0	ABABSA	Belgacem		NaN	NaN	...
4	Théâtres d'opérations extérieurs	Afrique du Nord		AC	40.0	R 4034.0	ABABSA OTHMANE BEN AHMED		NaN	NaN	NaN	...
...
20054	Théâtres d'opérations extérieurs	Afrique du Nord		NaN	NaN	NaN	LUZET		NaN	NaN	NaN	...
20055	Théâtres d'opérations extérieurs	Levant		AC	40.0	R 4070.0	LVOFF	André		NaN	NaN	...
20056	Théâtres d'opérations extérieurs	Afrique du Nord		AC	40.0	R 4046.0	LYARID BEN FARRADJ		NaN	NaN	NaN	...
20057	Théâtres d'opérations extérieurs	Afrique du Nord		AC	40.0	R 4046.0	LYAZID BEL FKJ AOMAR		NaN	NaN	NaN	...
20058	Théâtres d'opérations extérieurs	Levant		AC	40.0	R 4070.0	LYSRI	Youcef		NaN	NaN	...

20059 rows x 74 columns

Figure 1: Base de données originelle

Ensuite j'ai filtré la base par année et par lieu de décès. La période qui m'intéresse est de janvier 1925 à décembre 1926 où la majorité des combats ont eu lieu. Dans la base, il y a une colonne pour le pays de décès et là j'ai filtré pour le Maroc. Le script que j'ai utilisé pour filtrer la base se trouve

sur la **branch v.02** de mon projet sur [Github](#). Ce travail m'a permis de créer une base de données personnelle des sujets qui m'intéressent :

Out [7]:

	id_conflit_intitule	id_sous_conflit_intitule	id_famille_cote_intitule	sous_serie	serie	article	nom	prenom	nom_autre	naissance_jour_mois_annee	...
0	Théâtres d'opérations extérieures	Afrique du Nord		NaN	NaN	NaN	AB EL OUAHAB BEN SAID BEN BELGACEM	NaN	NaN	0000-00-00	...
1	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4034.0	ABASSIA OTHMANE BEN AHMED	NaN	NaN	NaN	...
2	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4034.0	ABAD	Tayeb	NaN	NaN	...
3	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4034.0	ABADIE	Jean Aubin	NaN	0000-00-00	...
4	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4034.0	ABBADIE	Jules Jean Marie	NaN	NaN	...
...
4717	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4046.0	LOUZABI	Ahmed	NaN	NaN	...
4718	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4046.0	LUCAS	Marcel Charles	NaN	NaN	...
4719	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4046.0	LUDWIG	Otto	NaN	NaN	...
4720	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4046.0	LUJEMANN	Clément	NaN	NaN	...
4721	Théâtres d'opérations extérieures	Afrique du Nord		AC	40.0	R 4046.0	LUXEMBURGER	Jean	NaN	NaN	...

4722 rows x 39 columns

Figure 2: Base de données finale

Avec cette base constituée, j'ai commencé les premières manipulation des caractéristiques des soldats telles que leur régiment ou leur lieu de décès. Mon but était de faire une pré-analyse avant de commencer à traiter les soldats par nom et faire l'OCR des fiches matricules. Le script que j'ai utilisé pour faire mes premières figures se trouve la **branch v.03** de mon projet sur [Github](#). J'ai chercher les 20 régiments avec le plus de pertes et les 20 lieux avec le plus de morts dans ma base :

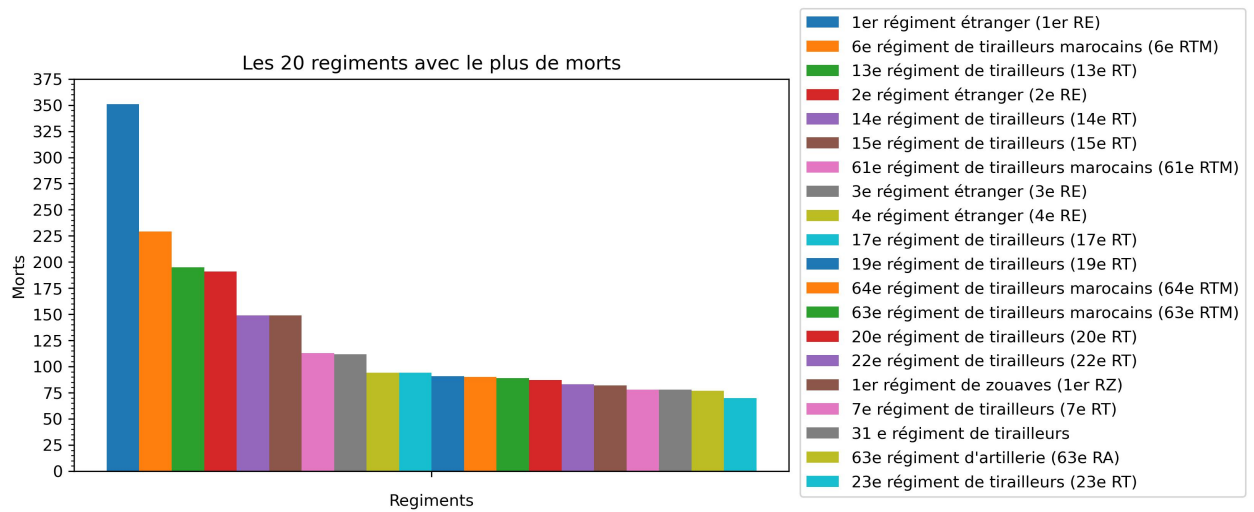


Figure 3: Les 20 premiers régiments avec le plus de morts

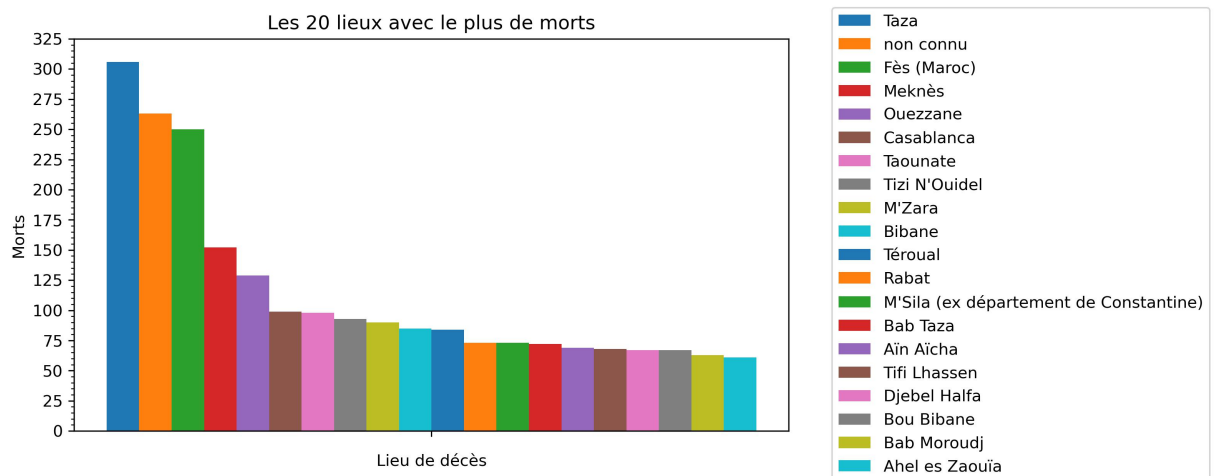
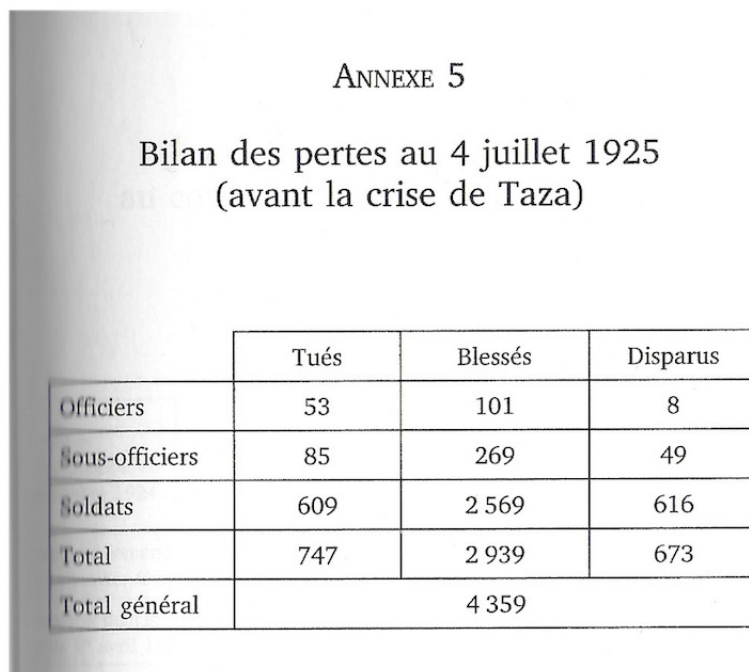


Figure 4: Les 20 premiers lieux avec le plus de morts

2 Les problèmes :

Le premier problème qui m'a interpellé est la vérification de mes données avec celles Max Schiavon dans son ouvrage *La guerre du Rif*. Je partage ici un scan de l'ouvrage de Schiavon qui décrit les pertes de l'armée avant l'offensif rifaine de 1925.^[1] Dans le support de son ouvrage, il inclut beaucoup



The image shows a scan of a page from a book. At the top, it says 'ANNEXE 5'. Below that, the title of the table is 'Bilan des pertes au 4 juillet 1925 (avant la crise de Taza)'. The table itself has four columns: the first column lists the categories of personnel, and the next three columns list the number of 'Tués' (killed), 'Blessés' (wounded), and 'Disparus' (missing). The rows are: 'Officiers', 'Sous-officiers', 'Soldats', 'Total', and 'Total général'. The numbers are: Officers (53 killed, 101 wounded, 8 missing), Sub-officers (85 killed, 269 wounded, 49 missing), Soldiers (609 killed, 2 569 wounded, 616 missing), Total (747 killed, 2 939 wounded, 673 missing), and Total général (4 359 total losses).

	Tués	Blessés	Disparus
Officiers	53	101	8
Sous-officiers	85	269	49
Soldats	609	2 569	616
Total	747	2 939	673
Total général	4 359		

Figure 5: Bilan de pertes 1925

de données quantitatives sur la guerre mais pas le bilan des pertes totales. Donc il faudrait que j'essaie de trouver dans quel référence au SHD il a trouvé ces chiffres.

Je me suis aussi rendu compte en créant mes premières figures que la base de données était pas forcément très propre. J'ai trouvé déjà un cas d'un double de la même personne et dans la Figure 4 vous pouvez voir qu'il y a « M'Sila » qui est en Algérie et pas au Maroc. Il faudrait que je trouve un moyen de filtrer les doubles et procéder à une vérification des noms de lieu.

3 Le travail à faire :

Maintenant que j'ai ma base constituée, il faut:

1. La nettoyer
2. Contacter le SHD pour vérifier l'origine des données chiffrés
3. Trouver une méthode pour diviser les soldats par origines
4. Constituer mon échantillon des fiches matricules numérisées
5. Commencer l'OCR des fiches matricules et vérifier sa faisabilité

List of Figures

1	Base de données originelle	1
2	Base de données finale	2
3	Les 20 premiers régiments avec le plus de morts	3
4	Les 20 premiers lieux avec le plus de morts	3
5	Bilan de pertes 1925	4

References

- [1] Max Schiavon, (1986) *La guerre du Rif - Un conflit colonial oublié (1925-1926)*, Éditions Pierre de Taillac, 2016, 352 pages.