

**TD
M2 MIAGE
PRM2
Géo-Marketing**

Fichier : TD_01_PRM2_Oct-2016.docx

Date : **Octobre 2016**

Rédacteur : DCN

Équipe d'étudiant(e)s

-
-
-
-
-
-

Diffusion

- Étudiants M2 Miage

Géo-Marketing

TD n°1 :

Visualisation de données cachées réparties sur une aire géographique

1- Problème à traiter

Dante Alighieri (dit « Dante », poète, écrivain et homme politique florentin, né en 1265 et mort en 1321) a rendu célèbre la distinction des langues romanes en les classant en fonction de la façon avec laquelle on disait **oui** :

- langues d'**oïl** au nord de la France,
- langues d'**oc** au sud de la France,
- langues de **si** pour l'Italie.

Au-delà de ces différentes façons de dire **oui**, les langues d'oc et d'oïl présentaient des différences syntaxiques qui provenaient des influences antérieures des langues parlées par les envahisseurs successifs qui avaient déferlé, siècle après siècle, sur les régions qui constituent maintenant la France.

Ainsi, **dans la partie nord de la France**, sous l'influence du Francique, **l'adjectif précède le nom**. On retrouve cette syntaxe en anglais ainsi que dans des tournures du Nord de la France et de la Belgique. Par exemple :

- en anglais : my new hat
- en picard : *min neu capiau*
- dans les toponymes et les noms de villes : *Neuville* (ville nouvelle), *Achicourt*, *Azincourt*, *Méricourt* (domaine de Achi, Azin, Méri), *Francheville* (Ville affranchie, ville libre), *Rougemont* (montagne rouge), *Brunehaut* (colline brune), *Blanc-Mont*, (montagne blanche), *Vauclair* (vallée claire), etc.

Au contraire, **dans la partie sud de la France**, sous l'influence du Latin, **l'adjectif suit le nom**. On retrouve cette syntaxe en espagnol, en italien ainsi que dans des tournures du sud de la France. Par exemple :

- en espagnol : *ciudad nueva*
- en italien : *citta nuova*
- en catalan : *ciutat nova*
- dans les toponymes et les noms de villes : *Villeneuve* (ville nouvelle), *Villefranche* (Ville affranchie, ville libre), *Montrouge* (montagne rouge), *Montauban*, (montagne blanche), *Clairvaux* (vallée claire), etc.

Le problème à traiter est le suivant. Nous allons établir une cartographie de la France hexagonale en analysant les noms des 36 000 communes et nous indiquerons, par une couleur adéquate, les zones qui relèvent probablement de la langue d'Oc et celles qui relèvent probablement de la langue d'Oïl.

D'après l'INSEE (Institut national de la statistique et des études économiques), au 1er janvier 2015, la France comptait 36 658 communes dont 36 529 en France métropolitaine (Corse comprise).

Ces communes sont identifiées de manière non-ambigüe à l'aide d'un identificateur INSEE (COG : Code Officiel Géographique). Mais on peut aussi les identifier (moins finement) à l'aide de leur code postal à 5 chiffres.

Les deux premiers chiffres (12) du code postal (12345) correspondent au numéro du département. Les trois derniers chiffres (345) correspondent au bureau distributeur du courrier de la commune dans le département.

Il existe quelques cas rares de communes dont le courrier est distribué par un bureau qui ne se trouve pas dans le même département (cas de villages limitrophes). Toutefois, ces cas rares ne perturbent pas les calculs statistiques et, en première approche, on peut les négliger.

Bibliographie - webographie :

Voir, notamment,

http://www.lexilogos.com/etymologie_oil_oc.htm

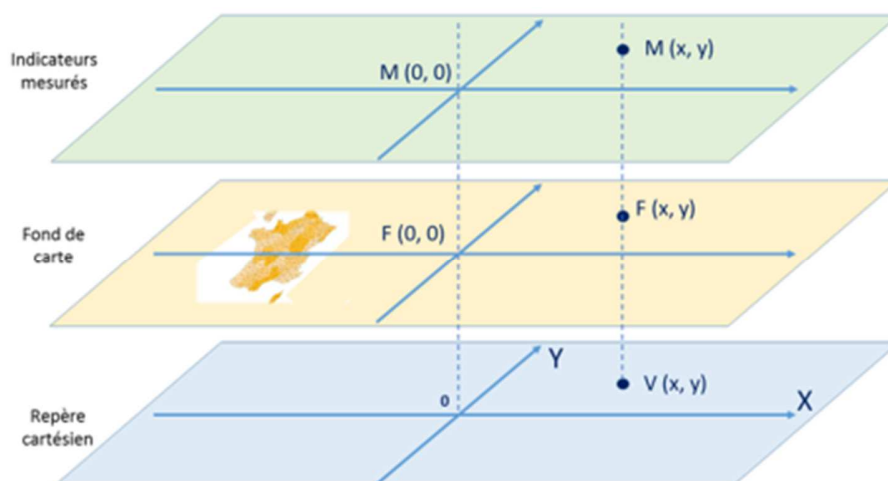
<http://bbouillon.free.fr/univ/hl/Fichiers/Cours/oil-oc.htm>

<http://www.insee.fr/fr/methodes/nomenclatures/cog/>

<https://www.data.gouv.fr/en/datasets/base-officielle-des-codes-postaux/>

2- Analyse du problème

On s'appuie sur le modèle « multi-couche » suivant pour analyser le problème à traiter.



2.1 – Couche basse (repère cartésien)

En raison de la surface à prendre en compte (surface géodésique égale à 551 695 km² pour la France métropolitaine), on adoptera les coordonnées géographiques usuelles (latitude et longitude). On ne prendra pas en compte l'altitude (ou élévation) par rapport au niveau de la mer parce que cette coordonnée n'est pas pertinente dans le problème posé.

Il faut évidemment choisir une unité de mesure adéquate pour la latitude et la longitude.

2.2 – Couche médiane (fond de carte)

Puisqu'on cherche à modéliser la **France métropolitaine**, il nous faut trouver un fond de carte représentant la métropole.

Et puisqu'on utilise les codes postaux qui sont corrélés aux départements, il nous faut trouver un fond de carte représentant les **départements** en France métropolitaine.

Pour superposer correctement le fond de carte et le repère cartésien, il est indispensable de connaître les coordonnées d'au moins trois points (voire quatre points) sur le fond de carte. Par exemple, on peut utiliser les coordonnées géographiques de Lille, Strasbourg, Marseille et Brest.

Lille : Coordonnées	50° 38' 14" Nord	3° 03' 48" Est
Strasbourg : Coordonnées	48° 34' 24" Nord	7° 45' 08" Est
Marseille : Coordonnées	43° 17' 47" Nord	5° 22' 12" Est
Brest : Coordonnées	48° 23' 27" Nord	4° 29' 08" Ouest

2.3 – Couche haute (indicateurs mesurés)

Les indicateurs que nous allons utiliser (et superposer aux couches précédentes) sont destinés à représenter le caractère « OC » ou « OïL » des communes de la France métropolitaine.

Soit L une variable linguistique définie sur l'intervalle $[-1 \dots +1]$ pour chaque commune.

Si $L = -1$, on dira que la commune est sûrement « OïL ».

Si $L = +1$, on dira que la commune est sûrement « OC ».

Dans tous les autres cas, on est dans un continuum entre -1 et $+1$ tel que :

si $L = 0$, on dira que la commune est indéterminée (sur le plan linguistique).

si $L < 0$, on dira que la commune est probablement « OïL »

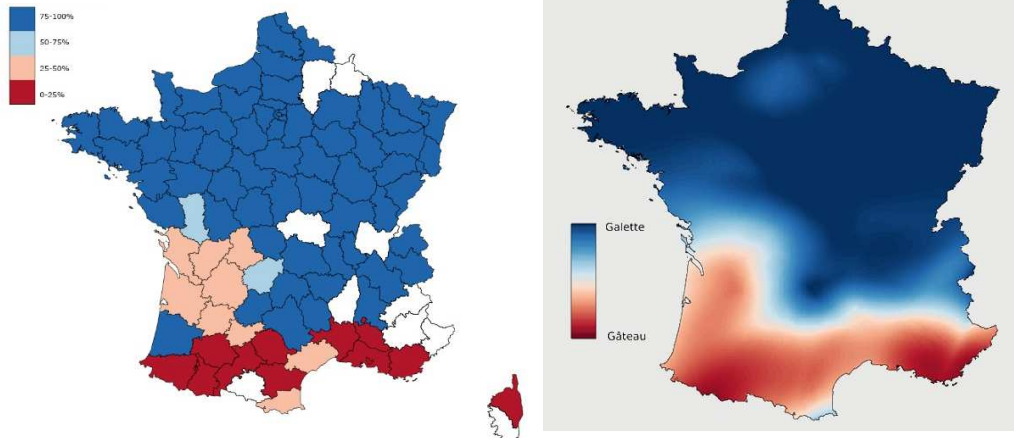
si $L > 0$, on dira que la commune est probablement « OC »

Selon la granularité souhaitée, on peut travailler commune par commune. Ou bien département par département, voire région par région.

Sur le plan graphique, on choisira une couleur associée à « OïL » et une autre associée à « OC ». Ces deux couleurs permettront de cartographier la France hexagonale en indiquant, par une couleur adéquate, les zones qui relèvent probablement de la langue d'Oc et celles qui relèvent probablement de la langue d'Oïl.

3- Résultat attendu

On livrera des cartes ressemblant aux exemples suivants (à ne surtout pas prendre au pied de la lettre). Bien entendu, on précisera la légende des cartes.



(Source : <http://www.slate.fr/story/96613/galette-ou-gateau-carte>)

Le choix des couleurs, des algorithmes ainsi que des bases de données et des outils et langages de développement est entièrement libre.

À noter : La complexité des données ne relève pas du « *big data* ». Le volume des données est de l'ordre de grandeur du méga-octet. Toutefois, il est hors de question de travailler « à la main ». Il faut entièrement automatiser les processus afin de limiter les temps de production (et donc les coûts de production) de ces cartes.