

# Analyse et modélisation des textes

Lou Burnard Consulting

octobre 2012

## 1 Numériser et encoder...

La numérisation nous propose une *image*, une représentation d'un objet déjà existant

L'encodage nous permet de représenter l'image des *idées* résultant de ces représentations

Difficile de limiter les idées, mais on peut distinguer plusieurs axes...

- infos 'intrinsèques' à l'objet : les formes, couleurs, etc. d'une image; les sons, rythmes, etc. d'une musique ; les structures linguistiques (mot, phrase, paragraphe) ou formels (chapitre, titre, nom de lieu) d'un texte...
- infos 'extrinsèques' ou 'meta' sur l'objet : son type, ses origines, ses buts, ses usages ...
- infos 'interprétatives': la portée d'un texte ou d'un dessin, le programme d'une musique ou d'un rite...

## 2 Modéliser et structurer

Il existe (quelques) méthodes classiques d'analyses de données. L'important c'est de bien comprendre que toute méthode ne serait qu'une modélisation. Cette modélisation devrait faire ressortir la structuration essentielle d'un objet complexe, priorisée selon les besoins prévus pour ce modèle.

Après une modélisation, on peut donner une implémentation informatisée ; sans modèle, une implémentation risque d'être incompréhensible et aléatoire, inutile...

L'analyse des données classique nous requiert de identifier ...

- les 'objets d'intérêt'
- leurs attributs/propriétés
- les relations entre objets
- les procédures/traitements essentiels envisagés

Dans cet exercice, on vous propose de faire pareille pour les documents en répondant aux questions suivantes :

- quelles sont les unités que vous souhaitez traiter dans votre version numérisée ?
- comment sont-elles structurées (quels composants, quels attributs) ?
- quelles relations éventuelles peut on envisager entre ces composants) ?
- est-ce que les occurrences de ces objets sont clairement identifiables dans un flux textuel ?

Les trois documents que vous expérimentez sont disponibles dans votre dossier Travaux: cartePostale.jpg, manuscrit.jpg et tapuscrit.jpg

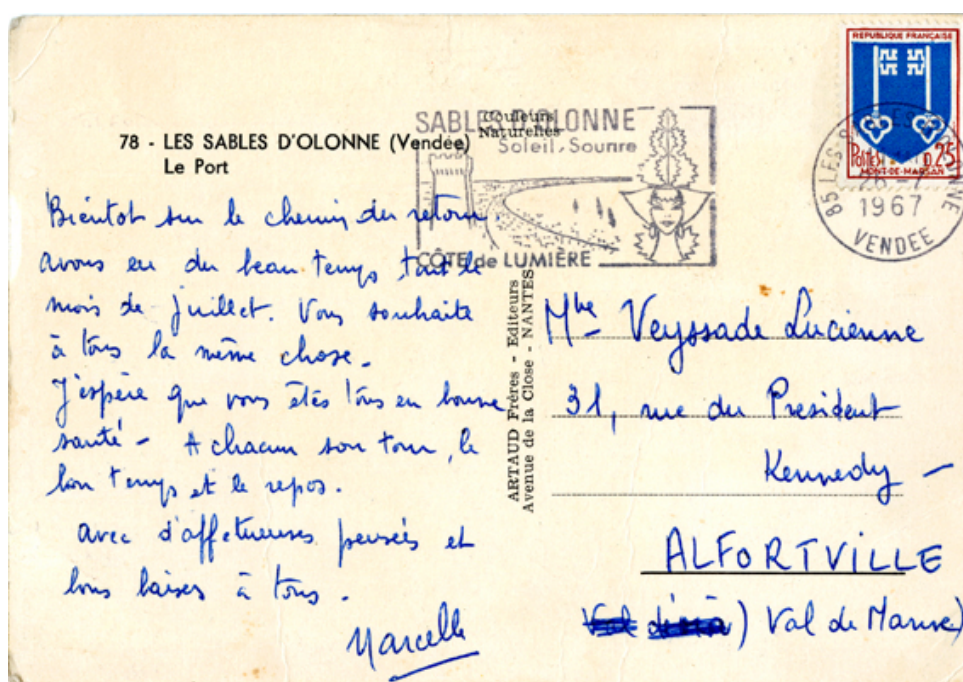
## 3 Démarche de l'exercice

- Dans votre groupe, sélectionner d'abord un rapporteur
- Regarder bien votre document. Supposer que vous en avez quelques centaines d'autres de même nature.

## 6 DOCUMENT C : UN TAPUSCRIT

- Comment est organisé votre document ?
- Quels sont ses composants essentiels, commun à tous ces documents ?
- Quels sont ses composants intéressants, possibles dans tous ces documents ?
- Sauriez-vous faire en sorte qu'une autre personne reconnaîtrait les mêmes objets ?
- Faire une liste de tous les objets et propriétés essentiels de votre document
- Supposez que le budget de votre projet soit coupé en deux: de quelles distinctions pourriez-vous vous passer ?
- Présentez cette liste... et justifiez-la !

## 4 Document A : une carte postale



## 5 Document B : un manuscrit

Nota bene : ce document, daté de 1763, est conservé aux Archives nationales sous la cote O1 284, n° 525. Il s'agit de l'enregistrement, par le gouverneur du château de Marly, d'ordres du roi relatifs au confort des courtisanes. Son édition en TEI, accompagnée d'un commentaire diplomatique et historique, est publiée dans les dossiers documentaires de l'application Theleme (Techniques pour l'Historien en Ligne : Études, Manuels, Exercices, Bibliographies) de l'Ecole nationale des chartes. Vous pouvez la consulter à la page <http://theleme.enc.sorbonne.fr/dossiers/notice105.php>.

## 6 Document C : un tapuscrit

Mai 1763  
ordres du Roy

Le pavillon du Roy comme  
à l'ordinaire, mais l'on ne  
donnera de bois d'arrivée à  
personne.



Le voyage étant comme  
celui de Versailles, l'on n'est  
obligé à rien, mais on peut  
leur donner quelque secours en  
bois et lumière, Mr. Boucheman  
et à moins ou autrement.

oui sans doute

oui en outre avec le  
plus exact

ou non, une denie

1782 n° 595

non mais la  
valeur en argent

Le Roy chauffe et chauffe les conseillers  
d'état à leur pied à terre qui est au 5 au commun

Le S. Luthier demande si l'on éclairera  
et chauffera le Château de Marly le  
voyage prochain, comme dans les voyages  
ordinaires.

Si le Concierge, les Garçons du Gardemeuble,  
Le S. Boucheman, et le Moine, Les  
Garçons du Château, l'Orloger, le Garçon  
des Châmes, et les Porteurs seront chauffés  
et éclairés.

Si on éclairera les Salles des Gardes  
de Votre Majesté, des Gardes de  
la Porte, et des Com-Suisses.

Si les femmes de chambre de Madame  
qui logent hors du château, seront chauffées  
et éclairées.

Si on donnera une bougie blanche par  
jour au Supérieur, des Cordeliers de  
Noisy.

Et si on donnera une denie bougie par  
jour au Vicaire de la Paroisse qui dessert  
la Chapelle du Commun.

Verailles



F. LE LIONNAIS  
23, route de la Reine  
92 - BOULOGNE SUR SEINE  
Tél. 605 90 13

001

le 3 Mars 1969

Chers Brigadiers,

Le sort en reste jeté. Raymond QUENEAU et René THOM viennent de me confirmer - avec des accents qui ne laissent aucun doute sur leur sincérité - que notre prochaine réunion aura bien lieu :

Lundi 17 Mars à 17 Heures, chez Georges PEREC (4<sup>e</sup> Etage)  
92, rue du Bac PARIS 7<sup>e</sup> - Téléphone : 222 95 68

On est prié de ne pas arriver en retard, car certains participants m'ont exprimé leur désir de se libérer avant 18 H 45.

Il avait récemment couru le bruit que le signataire de cette lettre était coutumier du parjure et qu'on ne pouvait lui faire confiance pour la livraison du "Deuxième Manifeste" destiné aux Editions Cape. C'était là pure calomnie. Voici le texte ou du moins un projet que vous êtes invités à analyser, critiquer et compléter. Voyez-vous d'autres aspects de l'idéal OuLiPien qui mériteraient d'être mentionnés ? Ou croyez-vous que l'aspiration sémantique devrait être développée ?

Je vous laisse sur ces méditations et vous adresse mes pensées amicales.

P.S. Claude BERGE : Et bien entendu, je t'attends chez moi Jeudi 13 Mars à 11 Heures.



Georges PEREC : THOM aimerait beaucoup avoir un tableau noir. Pourrez-vous vous le procurer ?

Jacques BENS : Merci pour votre lettre du 25 Février. L'idée d'une distinction entre "Lipo pure" et "Lipo appliquée" me semble d'une valeur certaine. Il vaudrait certainement la peine de s'y intéresser et peut-être cela donnerait-il un alinéa de plus dans le Manifeste ci-joint.

Venez-vite (et restez longtemps), on sera content de vous revoir.