**Objectif**

Constituer une base de données des textes de Guillaume Apollinaire, qu’il s’agisse de publications de son vivant, de publications posthumes ou de manuscrits.

Les textes sont stockés sous la forme de fichiers XML-TEI, soit une vingtaine de fichiers correspondant soit à des livres parus du vivant d’Apollinaire (*Le Bestiaire*, *Alcools*, *Calligrammes*, etc.), soit à des collections factices (Poèmes publiés du vivant d’Apollinaire, Poèmes posthumes, Articles parus dans le Mercure de France, etc.)

Le prototype porterait sur les textes poétiques, la prose représentant un ensemble considérable mais moins complexe à traiter (parutions généralement uniques des textes).

Les trois grandes difficultés tiennent :

1) à la visée chronologique du projet : les documents n’ont pas tous une date de création (au sens de la création par Apollinaire) ; et seuls les documents parus du vivant d’Apollinaire ont une date de publication.

2) à l’existence possible de plusieurs versions d’un même texte : un texte peut avoir un ou plusieurs états manuscrits (parfois fragmentaires), ils peut avoir été publié plusieurs fois, parfois sous des titres différents. L’enjeu est évidemment de pouvoir afficher à la recherche et sélectionner les différentes versions d’un même texte.

3) à l’existence de cas particuliers. Il n’y a pas seulement des textes (un poème, un article, etc.), mais des séries de textes (partie d’un recueil, suite de poèmes parue dans un numéro de revue), des subdivisions de textes (parties de poèmes, par exemple), enfin des textes insérés dans d’autres textes (par exemple un poème inséré dans un récit, dont il interrompt le flux de texte sans former une subdivision hiérarchique de type chapitre ou section).

**Identifiants :**

Les identifiants sont uniques :

1) Les textes parus du vivant d’Apollinaire : préfixe apo\_P, suivi de la date au format aaaa-mm-jj, suivi de la page, et si nécessaire de l’ordre de succession dans une même page.

Exemple : la série sans titre, parue dans la revue *Le Festin d’Ésope*, numéro 2, décembre 1903, p. 29. Elle comprend deux poèmes : « L’Adieu », « Le Retour ». identifiants : série sans titre = apo-P\_1903-12\_festin-esope\_p029 ; « L’Adieu » = apo-P\_1903-12\_festin-esope\_p029a, « Le Retour » = apo-P\_1903-12\_festin-esope\_p029b.

Le suffixe \_excerpt indique que le texte est contenu dans un autre texte (généralement un récit ou une lettre). Ce texte sera balisé en XML en utilisant l’élément <q> avec un xml:id

2) Les textes épistolaires : préfixe apo-L, suivi de la date, suivi du nom et du prénom du destinataire ; suffixe \_excerpt si le poème ne forme pas l’ensemble de la lettre.

Exemple : la lettre à Louis de Coligny-Châtillon du 3 février 1915 (apo-L\_1915-02-03\_coligny-chatillon-louise-de) comprend 2 poèmes : « Parce que tu m'as parlé de vice… » (apo-L\_1915-02-03\_coligny-chatillon-louise-de\_excerpt1), « Nos étoiles » (apo-L\_1915-02-03-\_coligny-chatillon-louise-de\_excerpt2)

3) Les états manuscrits ou publications posthumes : préfixe apo-M, suivi du code bibliothèque, de la cote et du folio. Par exemple : apo-M\_bnf\_naf19263\_f002v

**Champs de recherche**

L’objectif est d’interroger :

1) En plein texte

2) Par date de publication

3) Par date de création (avec la possibilité d’un empan de dates)

4) Par rubrique

5) Par série

4) Par titre

5) Par source (livre ou revue)

6) éventuellement de pouvoir discriminer en tenant compte d’autres éléments :

a) médium (livre, revue, journal, plaquette, carte postale, etc.) ;

b) état du texte (manuscrit, tapuscrit, mise au propre, copie d’imprimeur, épreuves, épreuves corrigées, etc.) ;

c) présence ou non d’illustrations ;

d) lieu de publication.

Ces champs doivent être prévus mais sont loin d’être renseignés pour l’ensemble des textes.

**Informations contenues dans le XML**

Le fichier XML donné en exemple comprend les poèmes publiés par Apollinaire de son vivant, hors les versions contenues dans les recueils poétiques. Certains textes sont parus à plusieurs reprises ; certains ont été repris dans les recueils.

Outre les informations d’une édition standard, les liens vers d’autres versions, l’appareil critique, etc., le fichier comprend deux informations essentielles :

1) Un identifiant est apposé à chaque série de texte, à chaque texte, et si nécessaire à chaque partie de texte.

2) Les notes préfixées par la mention [publication] comportent l’historique de publication du texte. Il serait évidemment assez aisé d’extraire la mention de tête pour en faire un <bibl>.

**Informations sous forme tabulaire**

N.-B. : les titres de champs, leur intitulé et leurs relations sont évidemment ouverts à discussion et à modification.

Onglet poemes-publies

ID : c’est l’identifiant unique

Type : c’est le type d’élément identifié : book (livre), series (série de textes : suite dans une revue, partie d’un recueil poétique), text (texte), part (partie d’un texte), excerpt (texte publié au sein d’un autre texte). Difficultés : a) en poésie, un livre comprendra soit des poèmes (texts), soit des suites de poèmes (series), cela ne signifie pas que tous les texts soient contenus dans des series = le recueil *Calligrammes* comprend au même niveau hiérarchique des poèmes et des parties ; b) en prose il faut prévoir un type chapter pour les chapitres de livres.

IsPartOf : repris aussi dans un onglet à part. C’est une relation : series ou text compris dans un book, text dans une series, part dans un text, excerpt dans une part ou un text.

Level : deux colonnes probablement superflues, qui ont servi à usage interne, pour déterminer le type d’élément.

PubDate : date de publication

CreationDate : date de création

Heading : rubrique, intervient rarement en poésie, mais notion importante pour les textes en prose. La rubrique dans les revues et les journaux est une section récurrente du périodique, comprenant un ou plusieurs textes. Il y a un HeadingNumber qui donne le numéro d’ordre du texte d’Apollinaire dans la série.

Title : c’est le titre de l’élément, tel qu’il apparaît dans la source. La colonne suivante donne une version normalisée, pour les cas de coquilles ou pour l’accentuation.

Clé de regroupement : permet de regrouper les différentes instances de publication d’un même texte (même si le titre diffère). Voir aussi l’onglet IsVersionOf.

Reprise en recueil : c’est une colonne à usage interne.

Source : intitulé du livre ou du périodique contenant.

Series : intitulé de la série de texte. À côté colonne SeriesNumber pour le numéro d’ordre des éléments dans la série (nécessaire ?)

Tomaison et pagination : précisions bibliographiques. Pas forcément très utile sous cette forme si on trie déjà par date normalisée et par source, d’autant que l’identifiant comprend l’indication de page et d’ordre dans la page. Je pense plutôt donner une référence bibliographique complète extraite du xml.

Les autres colonnes sont à usage interne.

Onglet IsPartOf

Déjà contenu dans le premier onglet

Onglet IsVersionOf

Il n’est pas très bien nommé. C’est une table permettant de mettre en relations les états d’un même texte (dans ce cas précis les états de publications, mais s’y ajouteront à terme les états manuscrits).

L’onglet est probablement mal fait car j’ai répété les relations croisées (A = B, A = C, B = A, C = A, C = B), là où il aurait sans doute suffi de noter A = B et B = C pour induire les autres relations.

En outre c’est redondant avec la colonne Clé de regroupement

Exemple :

**Annie**

Paru deux fois

Sous le titre Fanny dans *Les Soirées de Paris*, septembre 1912, p. 238

Sous le titre Annie dans *Alcools*, achevé d’imprimer le 20 avril 1913, p. 54

Clé de regroupement = Annie

Relation apo-P\_1912-09\_soirees-paris\_p238 = apo-P\_1913-04-20\_alcools\_p054