
1 – Introduction

Les *Recommandations* de la TEI – *Text Encoding Initiative (TEI) Guidelines* – s’adressent à tous ceux qui souhaitent échanger des informations stockées sous forme électronique. Elles mettent l’accent sur l’échange des données textuelles mais d’autres types de données comme les images et les sons sont également pris en compte. Les *Recommandations* peuvent être appliquées aussi bien pour créer de nouvelles informations que pour échanger des informations existantes.

Les *Recommandations* fournissent le moyen de rendre explicites certaines caractéristiques d’un texte, de façon à faciliter le traitement de ce texte par des programmes informatiques pouvant s’exécuter sur des plate-formes différentes. Cette tâche d’explicitation est appelée «balisage» ou «codage». La représentation d’un texte sur un ordinateur met toujours en œuvre une forme de balisage ou une autre. La TEI tire son origine d’une part de l’anarchie qui règne dans la communauté scientifique en matière de format, et d’autre part du nombre croissant de traitements que les chercheurs opèrent sur les textes sous forme électronique.

Les *Recommandations* de la TEI s’appuient sur SGML (*Standard Generalized Markup Language*) pour définir leurs règles de codage¹. SGML est une norme internationale de plus en plus utilisée par les entreprises spécialisées dans le traitement de l’information. Cette norme permet de définir les règles de codage en termes d’éléments, d’attributs et de règles régissant l’utilisation de ces éléments et attributs. L’utilisation que la TEI fait de SGML est ambitieuse, tant par sa complexité que par sa recherche d’universalité; elle n’est cependant pas fondamentalement différente de tous les autres schémas de balisage fondés sur SGML. La preuve en est que tous les outils SGML généralistes sont capables de traiter des textes conformes à la TEI.

La TEI est soutenue par l’*Association for Computers and the Humanities*, l’*Association for Computational Linguistics* et l’*Association for Literary and Linguistic Computing*. Le projet a été en partie financé par le *National Endowment for the Humanities* américain, la DG XIII de la CEE, la fondation Andrew W. Mellon et le *Social Science and Humanities Research Council* du Canada. Les *Recommandations* ont été publiées en mai 1994, après six ans de travaux auxquels ont participé des chercheurs de toute nationalité et de toute discipline.

1. Voir la bibliographie sommaire donnée en section 22 «Références bibliographiques»

Au début de cette entreprise, les objectifs généraux de la TEI ont été définis par la résolution finale de la conférence préparatoire tenue au Vassar College de New York en novembre 1987. Cette résolution connue sous le nom de «Principes de Poughkeepsie» fut peu à peu précisée à travers une série de documents de travail. D'après ces documents les *Recommandations* devaient :

- être suffisamment précises pour représenter les propriétés des textes intéressants pour les chercheurs ;
- être simples, claires et concrètes ;
- être utilisables facilement par les chercheurs et ne pas nécessiter l'utilisation de logiciels spécifiques ;
- permettre une définition rigoureuse des textes en vue de traitements efficaces ;
- être modifiables par l'utilisateur ;
- respecter les normes en vigueur ou sur le point d'être adoptées.

Le monde de la recherche est large et divers. Pour que ces recommandations aient une large audience, il était important de s'assurer que :

1. les descriptions des caractéristiques fondamentales d'un texte puissent être facilement échangées ;
2. les descriptions spécialisées puissent être facilement ajoutées (ou supprimées) d'un texte ;
3. la même caractéristique puissent être encodée en parallèle de plusieurs façons ;
4. la richesse du balisage puisse être déterminée par l'utilisateur de la façon la plus simple possible ;
5. une documentation relative au texte et à la façon dont il a été codé soit fournie.

Ce document décrit un sous-ensemble utilisable des éléments définis par la TEI et fournit des recommandations en conformité avec les principes directeurs qui viennent d'être rappelés plus haut.

En effectuant un choix parmi les centaines d'éléments définis par la TEI, nous nous sommes efforcés d'identifier un ensemble initial contenant les éléments susceptibles de servir à n'importe quel utilisateur. L'expérience acquise en ne

travaillant qu’avec la «TEI Lite» sera d’une grande utilité pour comprendre la DTD TEI complète et pour identifier les parties de cette DTD qui sont nécessaires pour travailler sur un type de texte donné.

Nous pouvons résumer comme suit les objectifs que nous nous sommes assignés en définissant ce sous-ensemble :

- il doit comprendre la plupart des éléments qui appartiennent au *core tag set* de la TEI ; ce dernier contient en effet des éléments généraux susceptibles d’être utilisés pour tout type de texte et tout type de traitement ;
- il doit permettre de traiter correctement une large gamme de texte, avec un niveau de précision correspondant aux pratiques courantes des chercheurs ;
- il doit pouvoir être utilisé aussi bien pour produire de nouveaux documents que pour coder des documents existants ;
- il doit pouvoir fonctionner avec la plupart des logiciels SGML ;
- il doit être dérivé de la DTD TEI complète en utilisant les mécanismes d’extension décrits dans les *Recommandations* de la TEI ;
- il doit enfin respecter les contraintes énoncées ci-dessus de la façon la plus simple et la plus concise possible.