
2 – Premiers exemples

Ce chapitre a pour titre anglais « A Short Example » et correspond intégralement à la première section 2.1 « Un court exemple ». La section 2.2 présente un exemple en français qui n'apparaît pas dans l'original mais qu'il nous a paru intéressant d'ajouter pour montrer dès à présent quelques possibilités de la TEI en ce qui concerne les textes en français. Ce second exemple n'est donc qu'un complément au premier que le lecteur est supposé avoir étudié d'abord.

{Note du traducteur}

2.1. Un court exemple

Nous commençons par un bref exemple, reproduit ci-dessous et destiné à illustrer ce qui se passe lorsqu'un passage de prose est saisi sur ordinateur par un opérateur n'ayant qu'une faible notion de ce que représente le balisage, ou du potentiel des textes électroniques. Dans un monde parfait, une telle sortie imprimée pourrait être générée par un lecteur optique très précis. Elle tente de rester fidèle à l'aspect du texte imprimé, en retenant les mêmes coupures de ligne que l'original, en insérant des blancs afin de représenter la disposition des titres originaux et des fins de page, et ainsi de suite. Lorsque certains caractères sont nécessaires mais indisponibles au clavier (tels que la lettre « a » accentuée dans le mot « faà », ou le trait long), elle essaie de reproduire leur aspect.

CHAPTER 38

READER, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said -- 'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she

was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only -- 'Have you, miss? Well, for sure!'

A short time after she pursued, 'I seed you go out with the master, but I didn't know you were gone to church to be wed'; and she basted away. John, when I turned to him, was grinning from ear to ear. 'I telled Mary how it would be,' he said: 'I knew what Mr Edward' (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) -- 'I knew what Mr Edward would do; and I was certain he would not wait long either: and h'es done right, for aught I know. I wish you joy, miss!' and he politely pulled his forelock.

'Thank you, John. Mr Rochester told me to give you and Mary this.'

I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words --

'Sh'ell happen do better for him nor ony o' t' grand ladies.' And again, 'If she ben't one o' th' handsomest, sh'es noan faa\l, and varry good-natured; and i' his een sh'es fair beautiful, onybody may see that.'

I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and

474

JANE EYRE

475

Mary approved the step unreservedly. Diana announced that she would just give me time to get over the honeymoon, and then she would come and see me.

'She had better not wait till then, Jane,' said Mr Rochester, when I read her letter to him; 'if she does, she will be too late, for our honey moon will shine our life long: its beams will only fade over your grave or mine.'

How St John received the news I don't know: he never answered the letter in which I communicated it: yet six months after he wrote to me, without, however, mentioning Mr Rochester's name or alluding to my marriage. His letter was then calm, and though very serious, kind. He has maintained a regular, though not very frequent correspond ence ever since: he hopes I am happy, and trusts I am not of those who live without God in the world, and only mind earthly things.

Cette transcription souffre d'un certain nombre de limitations :

- les numéros de page ainsi que les titres courants sont mêlés au texte de façon si inextricable qu'un logiciel aurait du mal à les démêler ;
- aucune distinction n'est faite entre l'apostrophe et les guillemets simples, d'où la difficulté de déterminer de façon sûre quels passages sont en discours direct ;
- la conservation des césures de l'original implique que des programmes de recherche peu élaborés ne trouveront pas les mots coupés ;
- la lettre accentuée dans « faàl » et le trait long ont été saisis selon des conventions ad hoc et ne seront donc traités correctement que si le transcrip-teur prend le soin de les mentionner dans la documentation ;
- les divisions de paragraphe sont marquées uniquement par l'emploi d'espaces blancs, et des fins de ligne forcées ont été introduites au bout de chaque ligne. Par conséquent, tout changement de taille de police lors de l'impression du texte engendrera des problèmes de mise en page.

Nous présentons maintenant le même passage avec un codage conforme aux recommandations de la TEI. Comme nous le verrons, ce codage pourrait être étendu de bien des façons, mais au minimum, le codage recommandé par la TEI permet d'effectuer les distinctions suivantes :

- les divisions de paragraphe sont maintenant marquées de façon explicite ;
- les apostrophes sont distinguées des guillemets ;
- les références d'entité sont employées pour la lettre accentuée et le trait long ;
- les divisions de page ont été marquées par un seul élément `<pb>` vide ;
- afin de simplifier la recherche et le traitement, la division en lignes de l'original n'a pas été retenue et des mots coupés par accident typographique en fin de ligne se trouvent réassemblés sans commentaire ; si la division en lignes de l'original avait un intérêt particulier, comme ce pourrait être le cas pour une édition rare, elle pourrait être notée facilement, bien que ce ne soit pas le cas ici ;
- afin de faciliter la correction des épreuves, une nouvelle ligne a été insérée en début de chaque paragraphe, mais l'alinéa n'a pas été conservé.

<pb n='474'>
<div1 type=chapter n='38'>

<p>Reader, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said ‐

<p><q>Mary, I have been married to Mr Rochester this morning.</q> The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only ‐

<p><q>Have you, miss? Well, for sure!</q>

<p>A short time after she pursued, <q>I seed you go out with the master, but I didn't know you were gone to church to be wed</q>; and she basted away. John, when I turned to him, was grinning from ear to ear.
<q>I telled Mary how it would be,</q> he said: <q>I knew what Mr Edward</q> (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) ‐ <q>I knew what Mr Edward would do; and I was certain he would not wait long either: and h s done right, for aught I know. I wish you joy, miss!</q> and he politely pulled his forelock.

<p><q>Thank you, John. Mr Rochester told me to give you and Mary this.</q>

<p>I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words ‐

<p><q>She'll happen do better for him nor ony o' t' grand ladies.</q> And again, <q>If she ben't one o' th'

handsomest, sh'es noan faàl, and varry good-natured;
and i' his een sh'es fair beautiful, onybody may see
that.</q>

<p>I wrote to Moor House and to Cambridge immediately, to
say what I had done: fully explaining also why I had thus
acted. Diana and <pb n='475'> Mary approved the step
unreservedly. Diana announced that she would just give me
time to get over the honeymoon, and then she would come and
see me.

<p><q>She had better not wait till then, Jane,</q> said Mr
Rochester, when I read her letter to him; <q>if she does,
she will be too late, for our honeymoon will shine our life
long: its beams will only fade over your grave or mine.</q>

<p>How St John received the news I don't know: he never
answered the letter in which I communicated it: yet six
months after he wrote to me, without, however, mentioning Mr
Rochester's name or alluding to my marriage. His letter was
then calm, and though very serious, kind. He has maintained
a regular, though not very frequent correspondence ever
since: he hopes I am happy, and trusts I am not of those who
live without God in the world, and only mind earthly things.

La décision de se concentrer sur le texte de Brontë, plutôt que sur la façon dont il a été imprimé dans cette édition spécifique, témoigne d'un problème de codage fondamental: celui de la sélectivité. Un codage rend explicite seulement celles des caractéristiques du texte qui ont une importance pour l'encodeur. Il n'est pas difficile d'imaginer comment on pourrait étendre facilement le codage d'un passage même aussi court que celui-ci. Par exemple :

- une forme en anglais officiel pourrait être fournie pour les passages en dialecte;
- des gloses ou des commentaires sous forme de notes de pied de page pourraient être ajoutés;
- des pointeurs pourraient être ajoutés afin de lier certaines parties du texte entre elles;
- des noms propres de divers types pourraient être distingués par rapport au texte environnant;
- des données bibliographiques détaillées concernant la provenance du texte ainsi que son contexte pourraient être fournies dans un en-tête;

-
- une analyse linguistique du passage en phrases, clauses, mots, etc. pourrait être fournie, chaque unité étant associée à des codes de catégorie appropriés ;
 - le texte pourrait être segmenté dans des unités de discours ou de narration ;
 - l’interprétation ou l’analyse systématique du texte pourraient être incluses dans le codage, avec des associations ou des alignements potentiellement complexes entre le texte et l’analyse, ou entre le texte et une ou plusieurs traductions de ce dernier ;
 - certains passages dans le texte pourraient être liés à des images ou à du son stockés sur d’autres supports.

La méthode recommandée par la TEI pour faire tout ceci est décrite dans le reste de ce document. La TEI dans son ensemble fournit également maintes possibilités, dont nous ne citons que quelques exemples :

- l’analyse détaillée des composants de noms ;
- des méta-informations détaillées fournissant des données quant aux origines du texte et aux sujets qu’il aborde ;
- des informations sur l’historique des impressions, ou des variantes manuscrites telles qu’elles apparaissent dans une suite de versions particulières du texte.

Pour connaître les recommandations applicables dans ces différents cas, et pour découvrir de nombreuses autres possibilités, il convient de se reporter au texte complet des recommandations de la TEI.

2.2. La TEI et le français

Le premier exemple a mis l’accent sur la typologie du texte. Nous complétons cette introduction par un exemple en langue française montrant quelques possibilités de la TEI liées à notre langue. Cette fois partons du document original¹ de la figure 1.

Si l’on devait coder ce texte, on aurait quelques problèmes plus évidents en français qu’en anglais :

- l’emploi de lettres accentuées est très important ;

1. Il s’agit du premier chapitre de *Les belles images* de Simone de Beauvoir, Gallimard, 1960.

«C'est un mois d'octobre... exceptionnel», dit Gisèle Dufrène ; ils acquiescent, ils sourient, une chaleur d'été tombe du ciel gris-bleu — Qu'est-ce que les autres ont que je n'ai pas ? — ils caressent leurs regards à l'image parfaite qu'ont reproduite *Plaisir de France* et *Votre Maison* : la ferme achetée pour une bouchée de pain — enfin, disons, de pain brioché — et aménagée par Jean-Charles au prix d'une tonne de caviar. (“ je n'en suis pas à un million près ”, a dit Gilbert), les roses contre les murs de pierre, les chrysanthèmes, les asters, les dahlias “ les plus beaux de toute l'Ile-de-France ”, dit Dominique ; le paravent et les fauteuils bleus et violet — c'est d'une audace ! — tranchent sur le vert de la pelouse, la glace tinte dans les verres, Houdan baise la main de Dominique, très mince dans son pantalon noir et son chemisier éclatant, les cheveux pâles, mi-blonds, mi-blancs, de dos on lui donnerait trente ans. ...

FIGURE 1 – *Un exemple de texte français à coder*

- mais si l'on saisit en utilisant un codage comme ISO 8859-1 (Latin1), on sera quand-même obligé de coder certains caractères en utilisant des entités SGML, par exemple «œ» ne fait pas partie de ce codage et devra donc être codé œ ;
- alors que les textes anglais peuvent provoquer des confusions entre l'apostrophe et le guillemet simple, les textes français sont souvent composés avec toute une artillerie de guillemets divers (ce texte de Simone de Beauvoir utilise les guillemets français «et» et ceux anglais “ et ”);
- la typographie française nécessite divers types d'espaces (par exemple avant le point-virgule on met une espace fine, avant le deux-points une espace justifiante, etc.) ou de tirets (par exemple « — ») qu'une saisie simple aurait du mal à distinguer ;

Un texte TEI peut tenir compte de tous ces éléments : ce texte pourrait être codé comme suit :

```
<p> <q rend=frdquo> C'est un mois d'octobre &hellip; exceptionnel
</q>, dit Gis&egrave;le Dufr&egrave;ne;
ils acquiescent, ils sourient, une chaleur d'&eacute;t&eacute;
tombe du ciel gris-bleu&mdash;
<q>Qu'est-ce que les autres ont que je n'ai pas? </q>
&mdash; ils caressent leurs regards &aacute;l' image parfaite
```

qu’ont reproduite `<title>Plaisir de France</title>` et `<title>Votre Maison</title>`: la ferme achetée pour une bouche de pain – enfin, disons, de pain brioché – et amenée par Jean-Charles au prix d’une tonne de caviar. (`<q rend=endq>` je n’en suis pas un million `</q>`, a dit Gilbert), les roses contre les murs de pierre, les chrysanthèmes, les asters, les dahlias `<q rend=endq>` les plus beaux de toute l’Ile-de-France`</q>`, dit Dominique; le paravent et les fauteuils bleus et violet – `<q>`c’est d’une audace!`</q>` – tranchent sur le vert de la pelouse, la glace tinte dans les verres, Houdan baise la main de Dominique, `<tr>` mince dans son pantalon noir et son chemisier `<clatant>`, les cheveux p&icuml;les, mi-blonds, mi-blancs, de dos on lui donnerait trente ans.

...

Dans ce texte, on a trois types de citations: celles entre guillemets français, celles entre guillemets anglais et celles sans guillemets (par exemple `<q>`c’est d’une audace!`</q>`). Pour les deux premiers types, on a délimité la citation par le couple de balises `<q>` et `</q>` et au moyen de l’attribut `rend` on a mémorisé la nature des guillemets utilisés dans le texte imprimé (frdquo pour français double quote et english double quote) de façon à pouvoir les restituer si besoin est. De même, nous supposons que c’est lors de l’édition/impression du document que l’on doit se soucier de la présence d’espaces avant les points-virgules, deux-points, etc., pas au moment de la saisie ni du traitement. Aussi codons-nous «dit Dominique;» et non «dit Dominique_;».

Dans les documents SGML, les caractères accentués doivent être notés à l’aide d’«entités» (voir section 18 «Jeux de caractères, signes diacritiques, etc.») qui prévoient, par exemple, que «é» soit codé «`é`». Le but de ce mécanisme est d’obtenir un document composé uniquement de caractères appartenant au jeu ISO 646-IRV (ASCII) de base, donc très facile à transmettre.

Cependant, pour des raisons de lisibilité, nous avons employé les caractères ISO 8859-1 (Latin-1) pour coder les lettres des exemples utilisant le français dans la suite de cette traduction. En réalité, il est techniquement possible d’échanger des documents SGML de cette façon, mais au prix d’une modification de la «déclaration SGML» qui ne peut être effectuée que par un utilisateur expérimenté et qui rend le document moins interchangeable.

Comme il a été dit plus haut², travailler directement sur le codage TEI n’est pas chose naturelle. On écrira donc, dans les exemples qui suivent, des textes

2. Voir l’article de Jacques André sur les balises, ci-dessus.

lisibles (quitte à imaginer qu'il faille passer un programme de remplacement automatique sur ces textes), comme :

```
la ferme achetée pour une bouchée de pain &mdash;  
enfin, disons, de pain brioché &mdash; et aménagée par  
Jean-Charles au prix d'une tonne de caviar.
```

Balises françaises

Un autre aspect de l'utilisation du français pour la TEI est celui de la francisation des balises et attributs.

La TEI prévoit des mécanismes pour paramétrer le nom des balises et donc, si on le souhaite, utiliser des balises en français de son choix. Il y a là un compromis entre le confort d'écriture et de lisibilité pour le créateur du document et l'échangeabilité: les textes avec des balises « localisées » ne peuvent évidemment plus être immédiatement interprétés par d'autres. À noter que les tentatives pour franciser les langages de programmation (Basic, Pascal, etc.) n'ont jamais eu de succès, sans doute pour la raison ci-dessus... Mais enfin, avec la TEI c'est possible si on le souhaite.

Notre attitude, ici, a été de garder le nom anglais des balises dans le codage SGML des exemples, mais de proposer en cas d'ambiguïté, un équivalent français dans le texte de l'article.

Enfin, en ce qui concerne les valeurs d'attribut, nous avons parfois utilisé des chaînes « françaises » à la place des chaînes « anglaises » suggérées dans l'original, du moins lorsque la DTD le permettait (par exemple quand le type de l'attribut était CDATA).

L'ensemble de ces noms, anglais et français, se retrouve dans l'index final.