

English

The TEI Internationalization Proposal: Issues and Criteria of Cultural Localization of the TEI Guidelines

Italian

La TEI Internationalization Proposal: problemi e criteri di adattamento culturale delle TEI Guidelines

Valentina Notarberardino (valentina.not@tiscali.it)
Università La Sapienza Roma (Rome, Italy)

PAPER

[English]

The *TEI Internationalization Proposal* (<http://www.tei-c.org/I18N/>) is a multicultural and multilinguistic initiative undertaken from the TEI Consortium. It represents a “cultural challenge”: to develop an infrastructure through which it will be possible to translate the *TEI Guidelines* (<http://www.tei-c.org/P4X/index.html>) from English towards other languages and to store all the translations together.

The *Text Encoding Initiative* (<http://www.tei-c.org/>) is a milestone in the Humanities Computing field. It is the most important worldwide initiative for establishing standard guidelines to codify texts through the use of XML tags. However non-native English speakers are disadvantaged in understanding the *Guidelines*.

There are already a few translated versions of the *TEI Lite* (<http://www.tei-c.org/Lite/>) (in Italian, French, Spanish, Korean, Japanese, Chinese and Russian). Although they are useful documents, we must emphasise that the *TEI Lite* is just a short surrogate of the real *Guidelines*. Plus those versions are not yet synchronized and some of them (the Korean, Chinese, Japanese and the Russian ones) have kept the original English examples.

The *TEI Guidelines* are written in a modular way: each chapter deals with one main topic. Each paragraph contains descriptive prose and examples. There are “theoretical examples” (showing just the tags’ syntax) and “applicative examples” (the tags mark-up real texts). Our analysis will be focused on the latter.

This submission is focused on the issue of translating these “applicative examples” into Italian.

Although it is “relatively” easy to translate the *Guidelines*’ prose, the element and attribute names, and technical descriptions for tag usage, it is more problematic for the examples. We have three options: keeping the English originals, translating them, or looking for corresponding Italian texts. This submission will address the problems and the main methods involved in the last possible solution.

The main aim of cultural localization is choosing an Italian text that preserves as much information as possible from the original one. The decisions by which this is done needs to follow various analytic and selective criteria: *tag used*, *structure*, *genre*, *content* and *length* of the text and, if possible, its *date*.

In order to illuminate these problems we will analyse four examples in particular (three taken from the seventh chapter of the guidelines, one from the sixth), indicating that for each one of them there are different dominant criteria; and that sometimes these criteria overlap in importance. We aim to demonstrate that it is not possible to establish a hierarchy of importance for these criteria. Differences between texts oblige us to privilege different criteria in different cases.

The first text [see Appendix] (chap. 7.1.3.) we are going to analyse is a passage taken from an English novel.

It shows numbered divisions for texts (<div0>, <div1>, etc.) and their responding attributes (n, id, type). Therefore our Italian example should be taken from a novel (genre) made up of two books containing chapters and paragraphs (structure). Concerns about tags overlap with those about genre and textual structure. In this particular case the content is not so important as it is irrelevant for the argument presented in the guidelines.

Sometimes there are complications due to some “additional tags”, that are extraneous to the aim of the example. Here we find the <trailer> tag marking-up the ending part of the first book; it is not necessary to preserve it. We must maintain the <p> and <head> tags as the descriptive prose links to paragraphs dealing with these elements. They are core elements and we will find them in most of the examples.

Our second example [see Appendix] (chap. 7.1.4.) is an extract from a newspaper. Here the main feature is the org=“composite” attribute. It is very easy to find Italian texts that allow us to use the same tags and attributes related to the journalistic textual genre. Neither content is important, neither chronology. But how do we deal with the <soCalled> tag? In this case we can adapt the translation via rejection. The example aims to show the tag is placed inside the *Quotation* section. If the descriptive prose had indicated any link about it we would have preserved that element. If not, we would have to modify the prose as well, with all the complications that would involve.

Our third example [see Appendix] (chap. 6.3.3.) deals with the <soCalled> tag. We run into four different usages of this element. There are three sentences (we could translate and/or re-invent them) and a passage taken from an essay. Here, different usages are related to different communicative purposes of the text. In this case the discriminating criterion is not about structure but it is about content, that is the linguistic context of the sentence.

Looking for general conclusions, we could state that every time specific tags of this kind occur, it is necessary to privilege the content. The length of the text is also very important when there are many instances of the same element.

The fourth example [see Appendix] (chap. 7.2.1.) considers where chronological issues could be significant. The tag used here is the <trailer> of an ecclesiastical biography of the bishop of Tours. The text is very old, as we can see from the theme, the rhetorical language, and from the headings' length. The criterion of textual extent means that we need to find an Italian text that is more or less the same age, where it is more likely to have headings and trailers of a similar length.

This research aims to clarify the criteria informing translation of examples. The cultural localization of examples cannot be undertaken by a normal translator. It needs a person with a genuine experience in text encoding and with a good literary background to understand and apply different criteria where necessary. Consequently, an interdisciplinary and intercultural collaboration is necessary in order to ensure the project is successful.

[Italian]

La *TEI Internationalization Proposal* (<http://www.tei-c.org/I18N/>) è un'iniziativa multilinguistica e multiculturale intrapresa dal TEI Consortium. Il suo obiettivo rappresenta una “sfida culturale”: creare un'infrastruttura attraverso la quale sia possibile tradurre le *TEI Guidelines* (<http://www.tei-c.org/P4X/index.html>) dall'inglese verso altre lingue, ed archiviare insieme tutte le traduzioni.

La *Text Encoding Initiative* (<http://www.tei-c.org/>) è una pietra miliare nel campo dell'*Humanities Computing*. Si tratta della più importante iniziativa mondiale volta alla diffusione di strutture standard per la codifica testuale in XML. I codificatori non inglesi sono svantaggiati nella comprensione del testo delle *Guidelines*.

Esistono già alcune traduzioni della documentazione *TEI Lite* (<http://www.tei-c.org/Lite/>) (in italiano, francese, spagnolo, coreano, giapponese, cinese e russo). Sebbene si tratti di documentazioni molto utili, bisogna ricordare che la *TEI Lite* è solo un breve surrogato delle *Guidelines*. Tali versioni non sono ancora sincronizzate, ed alcune (quella coreana, cinese, giapponese e russa) hanno mantenuto gli esempi originali inglesi.

Le *TEI Guidelines* sono costruite seguendo una struttura modulare: ogni capitolo argomenta un tema principale. Ciascun paragrafo contiene prosa descrittiva ed esempi. Bisogna distinguere tra “esempi teorici” (mostranti solo la sintassi dei *tag*) ed “esempi applicativi” (i *tag* marcano testi reali).

Tale analisi focalizza il problema di tradurre questi “esempi applicativi” in italiano.

Sebbene sia “relativamente” semplice tradurre la prosa delle *Guidelines*, i nomi degli elementi e degli attributi, nonché la scheda di descrizione tecnica dei *tag*, il testo degli esempi presenta più questioni. Abbiamo tre opzioni: mantenere l’originale, tradurlo, oppure cercare un corrispondente italiano. Esporremo i maggiori problemi ed i principali metodi impiegati nell’intraprendere l’ultima possibilità.

Lo scopo finale dell’adattamento culturale sarà scegliere un testo italiano che preservi il maggior numero possibile di informazioni rispetto all’originale inglese. Per ottenere tale obiettivo bisogna tener conto di alcuni criteri analitici e selettivi: *tag usati*, *struttura*, *genere*, *contenuto* e *lunghezza* del testo e, per quanto possibile, *datazione* dello stesso.

Per esporre questi problemi, analizzeremo quattro esempi in particolare (tre tratti dal settimo capitolo ed uno dal sesto), indicando come in ognuno di essi risulti vitale un criterio diverso; talvolta i principi si sovrappongono. Vogliamo dimostrare come sia impossibile stabilire una gerarchia d’importanza valida una volta per tutte, in quanto la varietà dei testi ci obbliga a privilegiare ora l’uno ora l’altro criterio.

Il primo testo [vedi Appendice] (cap. 7.1.3.) che analizzeremo mostra alcuni passaggi di un romanzo inglese volto ad esemplificare i *tag* specifici per le divisioni testuali numerate (<div0>, <div1>, ecc.) e gli attributi correlati (n, id, type) ad essi. Quindi il nostro esempio italiano dovrebbe essere tratto da un romanzo (genere) composto di due libri, contenenti capitoli e paragrafi (struttura). La priorità data alla scelta dei *tag* si interseca con il discorso sul genere e sulla struttura testuale. In questo caso il contenuto non è così importante poiché è irrilevante per l’argomento presentato nelle *Guidelines*.

Talvolta c’è una complicazione dovuta alla presenza dei “*tag* addizionali”, cioè estranei allo scopo dell’esempio. Qui troviamo il *tag* <trailer> a marcare la parte finale del primo libro; non è indispensabile riproporlo. È necessario preservare i vari <p> ed <head>, in quanto la prosa descrittiva contiene *link* che rimandano ai paragrafi dove si argomenta l’uso di tali marcatori. Si tratta di elementi base che saranno presenti nella maggior parte degli esempi.

Il nostro secondo esempio [vedi Appendice] (cap. 7.1.4.) è un testo tratto da un quotidiano. La sua caratteristica principale è l’attributo org=“composite”. È molto semplice trovare corrispondenti italiani che permettano di usare gli stessi *tag* e gli stessi attributi. Il contenuto non è importante, nemmeno la datazione. Ma come trattare il <soCalled> *tag*? In questo caso adattamento è sinonimo di rinuncia. L’esempio che mostra l’uso di questo elemento è presente nella sezione *Quotation*. Se nella didascalia introduttiva ci fosse stato un riferimento in merito, con il relativo *link*, avremmo dovuto ripresentarlo. Diversamente, saremmo costretti a modificare la stessa prosa descrittiva con tutte le complicazioni del caso.

Il nostro terzo esempio [vedi Appendice] (cap. 6.3.3.) analizza il <soCalled> *tag*. Ci imbattiamo in quattro diversi usi di tale elemento. Si tratta di tre frasi (che potremmo re-inventare e/o tradurre) e di un passo tratto da un saggio. Qui i diversi usi corrispondono ai diversi scopi comunicativi del testo. In questo caso il criterio discriminante non è rappresentato dalla struttura ma dal contenuto, cioè dal contesto linguistico della frase.

Cercando di trarre conclusioni generali, si potrebbe asserire che ogni qual volta ci si imbatte in marcatori specifici di questo tipo, bisogna privilegiare il contenuto. La lunghezza del testo è importante nella misura in cui esso presenti varie istanze dello stesso elemento.

Il quarto caso [vedi Appendice] (cap. 7.2.1.) è un esempio per il quale potrebbero essere significativi i criteri di datazione. Il *tag* usato è il <trailer> di una biografia ecclesiastica del vescovo di Tours. Il testo è molto antico, come possiamo vedere dal tema, dal linguaggio retorico e dalla lunghezza degli *headings*. Il criterio dell’estensione testuale si intreccia con la necessità di risalire ad un testo altrettanto datato, nel quale è maggiormente probabile trovare titolazioni e *trailers* della stessa lunghezza.

La presente ricerca mira a chiarificare i criteri che ci guidano nella traduzione degli esempi. La localizzazione culturale degli esempi non può essere intrapresa da un semplice traduttore. È necessario possedere esperienza nel campo della codifica testuale ed un buon *background* letterario per capire ed applicare criteri

differenti dove necessario. Conseguentemente, la collaborazione interdisciplinare ed interculturale risulta indispensabile ai fini del successo del progetto.

Appendix/Appendice

First Example / Primo Esempio (cap. 7.1.3.)

<http://www.tei-c.org/P4X/DS.html >

```
<div0 type="book" n="I" id="JA0100">
  <head>Book I.</head>
  <div1 type="chapter" n="1" id="JA0101">
    <head>Of writing lives in general, and particularly of Pamela, with a word by the bye of Colley
    Cibber and others.</head>
    <p>It is a trite but true observation, that examples work more forcibly on the mind than
    precepts: ... </p>
    <!-- ... remainder of chapter 1 here ... -->
  </div1>
  <div1 type="chapter" n="2" id="JA0102">
    <head>Of Mr. Joseph Andrews, his birth, parentage, education, and great endowments; with a word
    or two concerning ancestors.</head>
    <p>Mr. Joseph Andrews, the hero of our ensuing history, was esteemed to be the only son of
    Gaffar and Gammar Andrews, and brother to the illustrious Pamela, whose virtue is at present so
    famous ... </p>
    <!-- ... remainder of chapter 2 here ... -->
  </div1>
  <!-- ... remaining chapters of Book 1 here ... -->
  <trailer>The end of the first Book</trailer>
</div0>
<div0 type="book" n="II" id="JA0200">
  <head>Book II</head>
  <div1 type="chapter" n="1" id="JA0201">
    <head>Of divisions in authors</head>
    <p>There are certain mysteries or secrets in all trades, from the highest to the lowest, from
    that of <term>prime-ministering</term>, to this of <term>authoring</term>, which are seldom
    discovered unless to members of the same calling ... </p>
    <p>I will dismiss this chapter with the following observation: that it becomes an author
    generally to divide a book, as it does a butcher to joint his meat, for such assistance is of great
    help to both the reader and the carver. And now having indulged myself a little I will endeavour to
    indulge the curiosity of my reader, who is no doubt impatient to know what he will find in the
    subsequent chapters of this book.</p>
  </div1>
  <div1 type="chapter" n="2" id="JA0202">
    <head>A surprising instance of Mr. Adams's short memory, with the unfortunate consequences which
    it brought on Joseph.</head>
    <p>Mr. Adams and Joseph were now ready to depart different ways ... </p>
  </div1>
</div0>
```

Second Example / Secondo Esempio (cap. 7.1.4.)

<http://www.tei-c.org/P4X/DS.html >

```
<div1 type="storylist" org="composite">
  <head>News in brief</head>
  <div2 type="story">
    <head>Police deny <soCalled>losing</soCalled> bomb</head>
    <p>Scotland Yard yesterday denied claims in the Sunday Express that anti-terrorist officers
    trailing an IRA van loaded with explosives in north London had lost track of it 10 days ago.</p>
  </div2>
  <div2 type="story">
    <head>Hotel blaze</head>
    <p>Nearly 200 guests were evacuated before dawn yesterday after fire broke out at the Scandic
    Crown hotel in the Royal Mile, Edinburgh.</p>
  </div2>
  <div2 type="story">
    <head>Test match split</head>
    <p>Test Match Special next summer will be split between Radio 5 and Radio 3, after protests
    this year that it disrupted Radio 3's music schedule.</p>
  </div2>
  <!-- other stories here -->
</div1>
```

Third Example / Terzo Esempio (cap. 6.3.3.)

<http://www.tei-c.org/P4X/CO.html >

- a) <head>PM dodges <soCalled>election threat</soCalled> in interview</head>
- b) He hated <soCalled>good</soCalled> books.
- c) <soCalled>Croissants</soCalled> indeed! toast not good enough for you?
- d) Although Chomsky's decision that all NL sentences are finite objects was never justified by arguments from the attested properties of NLs, it did have a certain <soCalled>social</soCalled> justification. It was commonly assumed in works on logic until fairly recently that the notion <mentioned>language</mentioned> is necessarily restricted to finite strings.

Fourth Example / Quarto Esempio (cap. 7.2.1.)

<http://www.tei-c.org/P4X/DS.html >

```
<div1 type="book" n="I"><head>In the name of Christ here begins the first book of the ecclesiastical
history of Georgius Florentinus, known as Gregory, Bishop of Tours.</head>
  <div2>
    <head>Chapter-Headings</head>
    <list>
      <!-- list of chapter heads omitted -->
    </list>
  </div2>
  <div2>
    <head>In the name of Christ here begins Book I of the history.</head>
    <p>Proposing as I do ...</p>
    <!-- ... -->
    <p>From the Passion of our Lord until the death of Saint Martin four hundred and twelve years
passed.</p>
    <trailer>Here ends the first Book, which covers five thousand, five hundred and ninety-six years
from the beginning of the world down to the death of Saint Martin.</trailer>
  </div2>
</div1>
```