

# **Traducció i tecnologies: eines, processos i recursos**

**Antoni Oliver**

# Introducció

Les tecnologies informàtiques s'han introduït plenament en el món de la traducció i els professionals de la traducció comptem amb un gran ventall d'aplicacions dissenyades per facilitar-nos la feina, millorant el rendiment i assegurant uns nivells de qualitat òptims. Aquestes tecnologies van des de la traducció assistida a la traducció automàtica, passant pels sistemes d'extracció i gestió de la terminologia. El traductor, a més, té a la seva disposició una gran quantitat de recursos lingüístics: glossaris terminològics, diccionaris i corpus paral·lels que, emprats de la manera adequada poden suposar un gran estalvi de temps.

Aquest llibre presenta d'una manera clara i amb detall tots els aspectes de les tecnologies aplicades a la traducció. També ensenya a trobar, crear i gestionar recursos lingüístics per a la traducció, especialment bases de dades terminològiques i memòries de traducció. El llibre pot ser un manual de consulta per al traductor en actiu i pot també servir com a material de referència principal en un curs universitari de grau, postgrau o màster sobre tecnologies aplicades a la traducció.

El llibre es complementa amb una sèrie d'enunciats de pràctiques que es poden descarregar lliurement de <http://lpg.uoc.edu/TIT>.

L'autor del llibre està obert a rebre comentaris, crítiques i suggeriments a través del seu correu electrònic.

Antoni Oliver

[aoliverg@uoc.edu](mailto:aoliverg@uoc.edu)

Novembre de 2014

# Índex

## 1. La traducció assistida per ordinador

- 1.1. Introducció
- 1.2. Components bàsics d'un sistema de traducció assistida
  - 1.2.a Entorn de treball
  - 1.2.b. Consulta a memòries de traducció
  - 1.2.c. Consulta a glossaris i diccionaris
  - 1.2.d. Combinació de traducció assistida i traducció automàtica
  - 1.2.e. Tractament de formats
- 1.3. El procés de traducció amb un sistema de traducció assistida
  - 1.3.a. Tractament del format
  - 1.3.b. Segmentació: el format SRX
  - 1.3.c. Formats de projectes de traducció: el format XLIFF
  - 1.3.d. Assignació de recursos a un projecte de traducció
  - 1.3.e. Traducció
  - 1.3.f. Revisió dins de l'eina de traducció assistida
  - 1.3.g. Creació dels documents traduïts
  - 1.3.h. Revisió en format final
  - 1.3.i. Gestió dels recursos generats
- 1.4. Els sistemes de traducció assistida
- 1.5. Conclusions
- 1.6. Per ampliar coneixements
  - 1.6.1. Eines de traducció assistida on-line
  - 1.6.2. SRX i les expressions regulars
  - 1.6.7. El corrector gramatical LanguageTool

## 2. Les memòries de traducció

### 2.1. Introducció

### 2.2. Indexació i recuperació de segments

#### 2.2.1. Indexació de memòries de traducció

#### 2.2.2. Càlcul de la similitud de segments

### 2.3. Coincidència exacta i parcial

### 2.4. Combinació d'unitats subsegmentals

### 2.5. Format d'intercanvi de memòries de traducció: TMX

### 2.6. Creació de memòries de traducció

#### 2.6.a. Traducció amb sistemes de traducció assistida

#### 2.6.b. Memòries de traducció i corpus paral·lels

#### 2.6.c. Alineació manual de documents

#### 2.6.d. Alineació automàtica de documents

### 2.7. Memòries de traducció remotes compartides i públiques

### 2.8. Treball amb memòries de traducció

### 2.9. Anàlisi de projectes i tarifació

### 2.8. Noves funcionalitats

#### 2.8.1. Sistema de traducció automàtica integrat

#### 2.8.2. Autocompletat intel·ligent

#### 2.8.3. Mesures de confiança

### 2.9. Conclusions

### 2.10. Per ampliar coneixements

#### 2.10.1. Corpus paral·lels i memòries de traducció disponibles públicament

#### 2.10.2. Etiquetadors morfosintàctics

#### 2.9.3. Propietat de les memòries de traducció

### **3. Les bases de dades terminològiques**

- 3.1. Introducció
- 3.2. Terminologia i traducció
- 3.3. Bases de dades terminològiques, diccionaris generals i coneixement enciclopèdic
- 3.4. Cerca automàtica a bases de dades terminològiques
- 3.5. Format d'intercanvi de bases de dades terminològiques: TBX
- 3.6. Creació de bases de dades terminològiques
  - 3.6.a. A mesura que es tradueix
  - 3.6.b. A partir de recursos terminològics a Internet
  - 3.6.c. A partir de la Vikipèdia
  - 3.6.d. Extracció automàtica de terminologia
- 3.7. Extracció automàtica de terminologia
  - 3.7.1. Definició
  - 3.7.2. Classificació de mètodes per a l'extracció automàtica de terminologia
  - 3.7.3. Mètodes estadístics per a l'extracció automàtica de terminologia
  - 3.7.4. Mètodes lingüístics per a l'extracció automàtica de terminologia
  - 3.7.5. Mesures estadístiques
- 3.8. Conclusions

### **4. La traducció automàtica**

- 4.1. Introducció
- 4.2. Història de la traducció automàtica
- 4.3. Tipus de sistemes de traducció automàtica
  - 4.3.1. Traducció directa
  - 4.3.2. Sistemes de transferència
  - 4.3.3. Sistemes d'interlingua
  - 4.3.4. Sistemes estadístics
  - 4.3.5. Sistemes de traducció automàtica basada en exemples

#### 4.3.6. Sistemes híbrids

#### 4.4. Usos de la traducció automàtica

#### 4.5. La traducció automàtica a Internet

#### 4.6. Combinació de traducció automàtica i traducció assistida

#### 4.7. Postedició de traducció automàtica

#### 4.8. Conclusions

#### 4.9. Per ampliar coneixements

##### 4.9.1. Història de la traducció automàtica

##### 4.9.2. Moses: un sistema de traducció automàtica estadística

##### 4.9.3. Altres sistemes de traducció automàtica

### 5. Tractament de formats

#### 5.1. Introducció

#### 5.2. Representació de la informació textual: codi de caràcters

#### 5.3. El llenguatge intern de l'ordinador i les unitats de mesura en informàtica

##### 5.3.a. Els múltiples del byte

#### 5.4. Representació d'informació no numèrica

##### 5.4.a. Representació de text

#### 5.5. Representació de la informació textual: codi de caràcters

##### 5.5.a. Conceptes bàsics

##### 5.5.b. Algunes definicions importants

##### 5.5.c. Els codis de caràcters més habituals

##### 5.5.d. Unicode

##### 5.5.e. Detecció de la codificació de caràcters

##### 5.5.f. Canvi de la codificació de caràcters

#### 5.6. La representació de la informació no textual

##### 5.6.a. Noms d'arxiu i extensions. Relació amb el format i l'aplicació

##### 5.6.b. El format HTML

##### 5.6.c. L'XHTML

#### 5.6.d. Open Document

#### 5.6.e. Els formats de documents de Microsoft Word

Microsoft Word DOC (.doc)

Microsoft Word 2003 XML (.xml)

Microsoft Word 2007/2010 XML (.docx)

#### 5.6.f. El format LaTeX

#### 5.6.g. El format DocBook

#### 5.6.h. El format PDF

### 5.7. XML

#### 5.7.a. Introducció

#### 5.7.b. Exemples senzills de documents XML

#### 5.7.c. Estructura dels documents XML

#### 5.7.d. Els documents XML ben formats

#### 5.7.e. Tecnologies associades: XSLT i XPATH

#### 5.7.f. Editors d'XML

#### 5.7.g. Traducció de documents XML

### 5.8. Els formats XML emprats en el món de la traducció

#### 5.8.a. Intercanvi de memòries de traducció: TMX

#### 5.8.b. Intercanvi de bases de dades terminològiques: TBX

#### 5.8.c. Intercanvi de projectes de traducció: XLIFF

#### 5.8.d. Intercanvi de regles de segmentació: SRX

#### 5.8.e. Mètriques GILT: GMX

## 6. Conclusions

Per ampliar coneixements

Tutorials d'W3Schools

Detecció automàtica de llengua

Taules d'Unicode

Problemes de visualització de documents relacionats amb les fonts

SC Unipad: Editor d'Unicode