

# 21963 Informàtica 2013

Sessió 8 (9a setmana)

Tema: traducció automàtica vs. traducció  
assistida

Lluís de Yzaguirre

Iria da Cunha

Alba Coll

Amor Montané

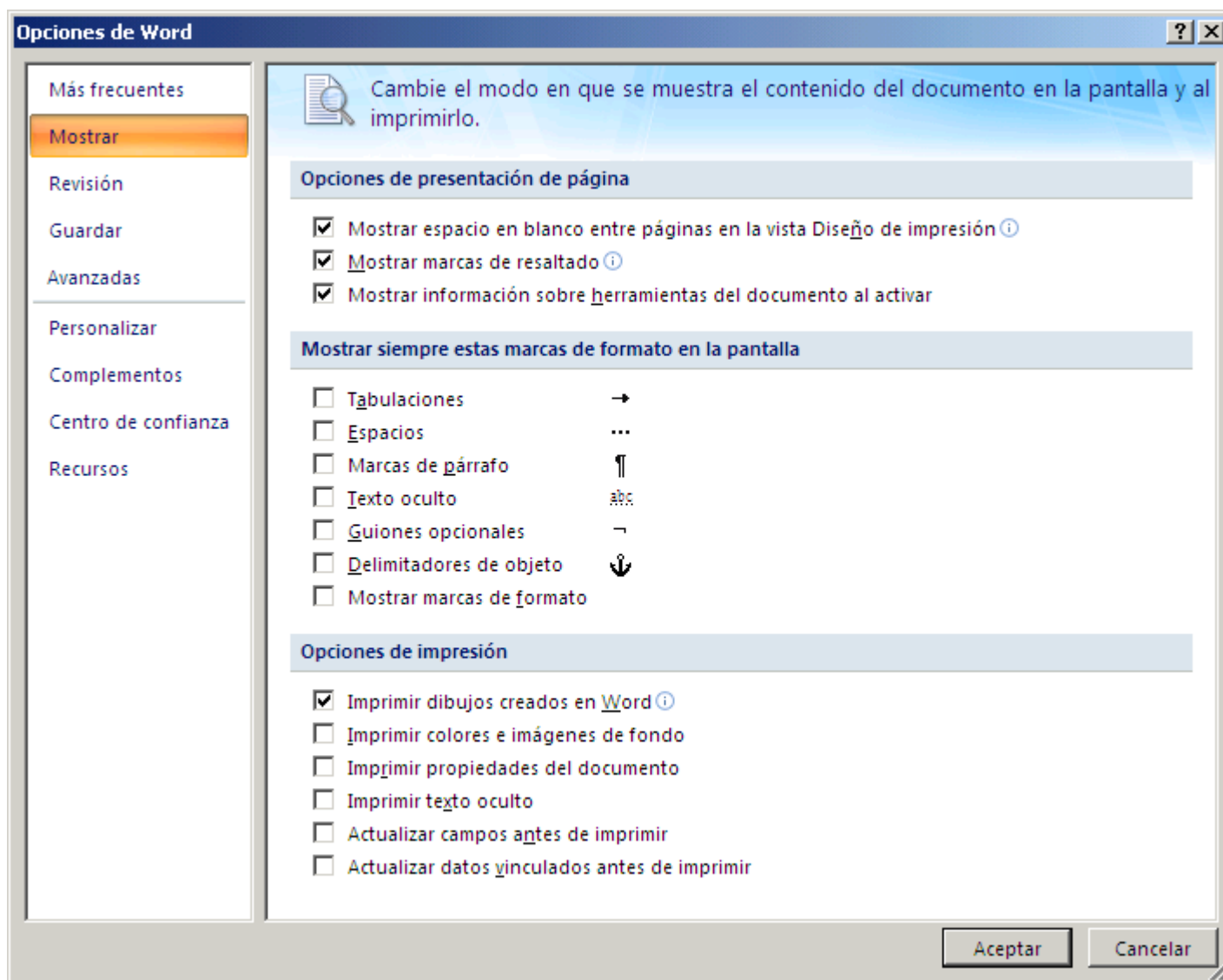
## Seminari 7:

- Valoració macro+excel

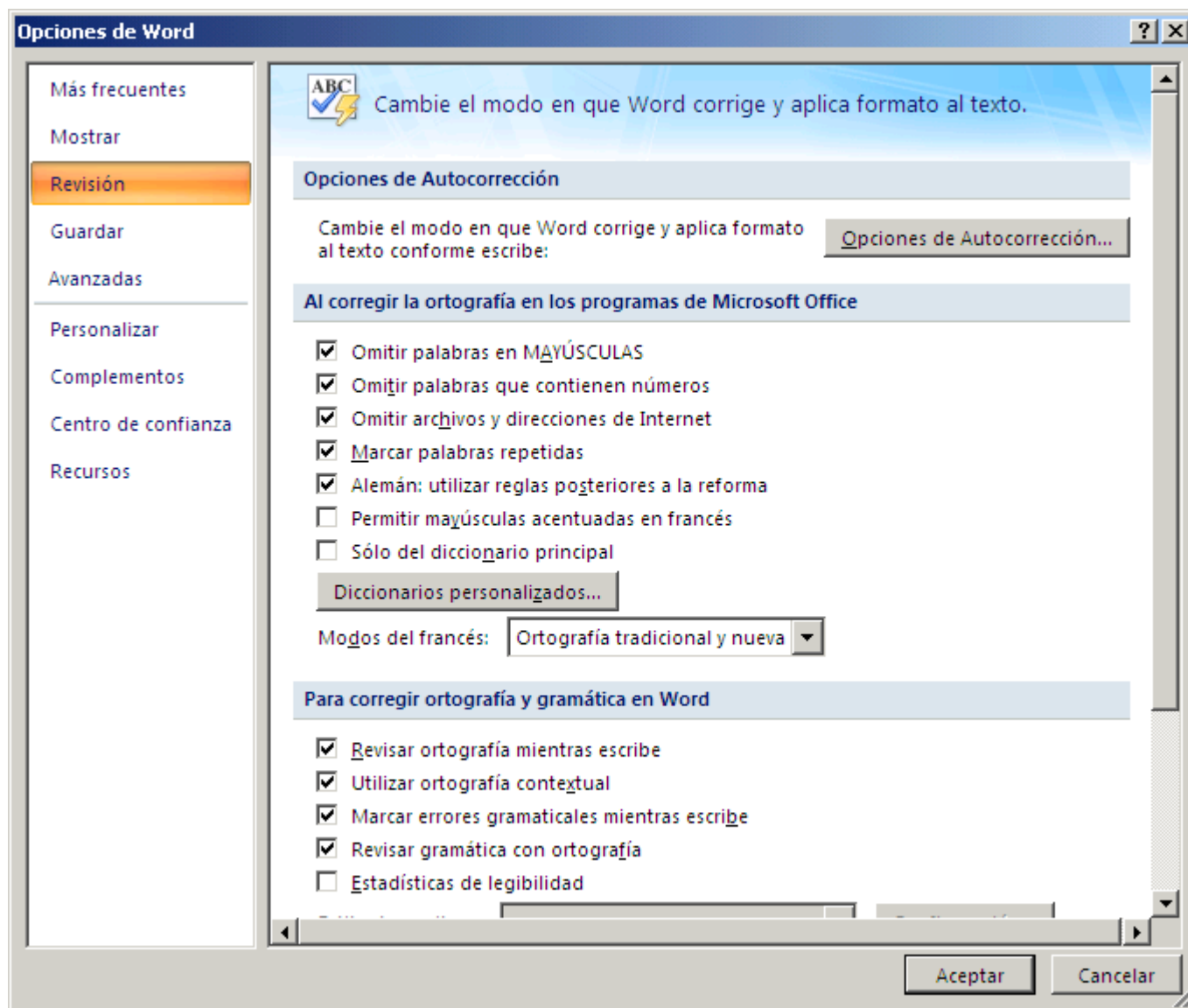
## Avís

- Gala dels Curts: al final

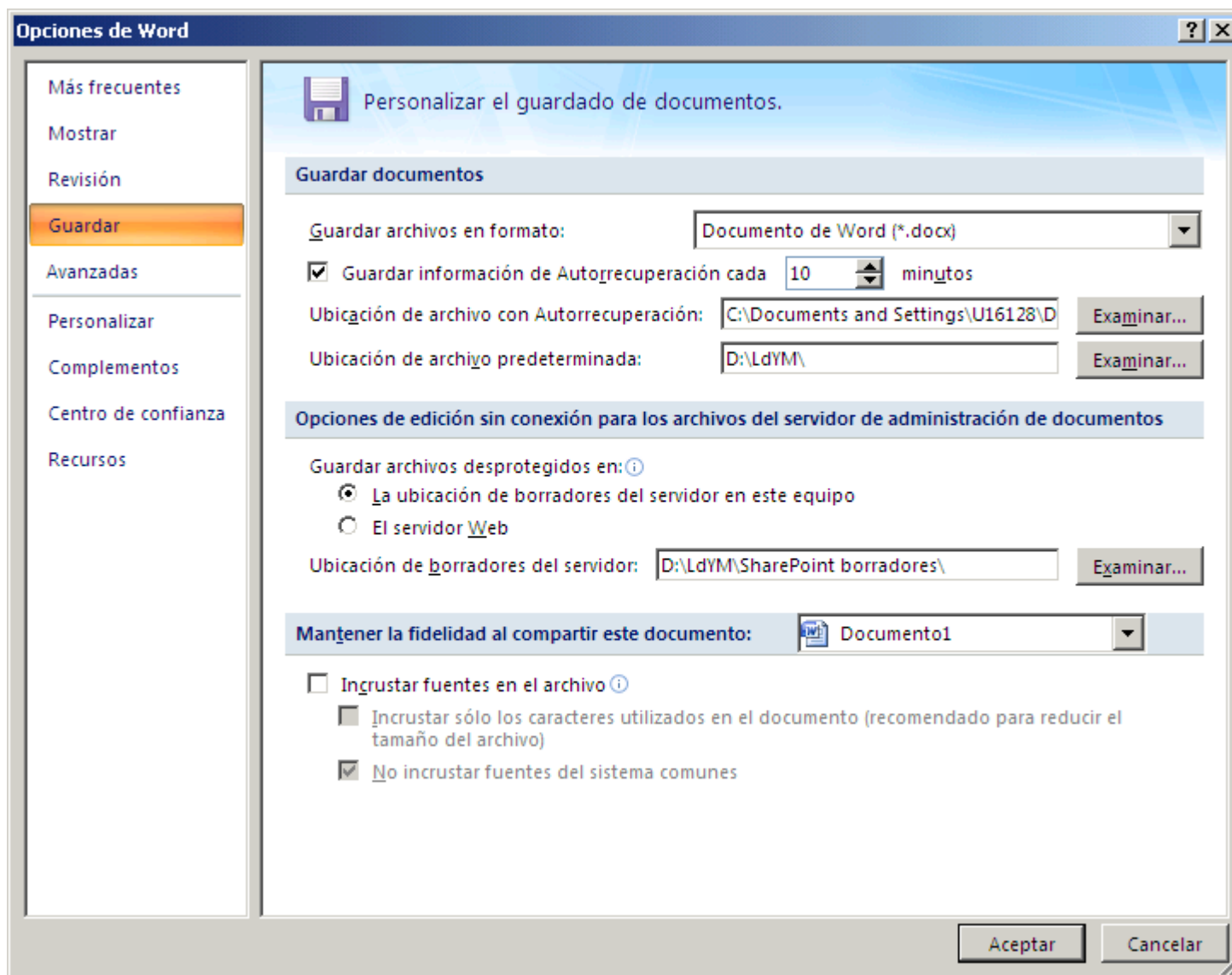
# Valoració macro



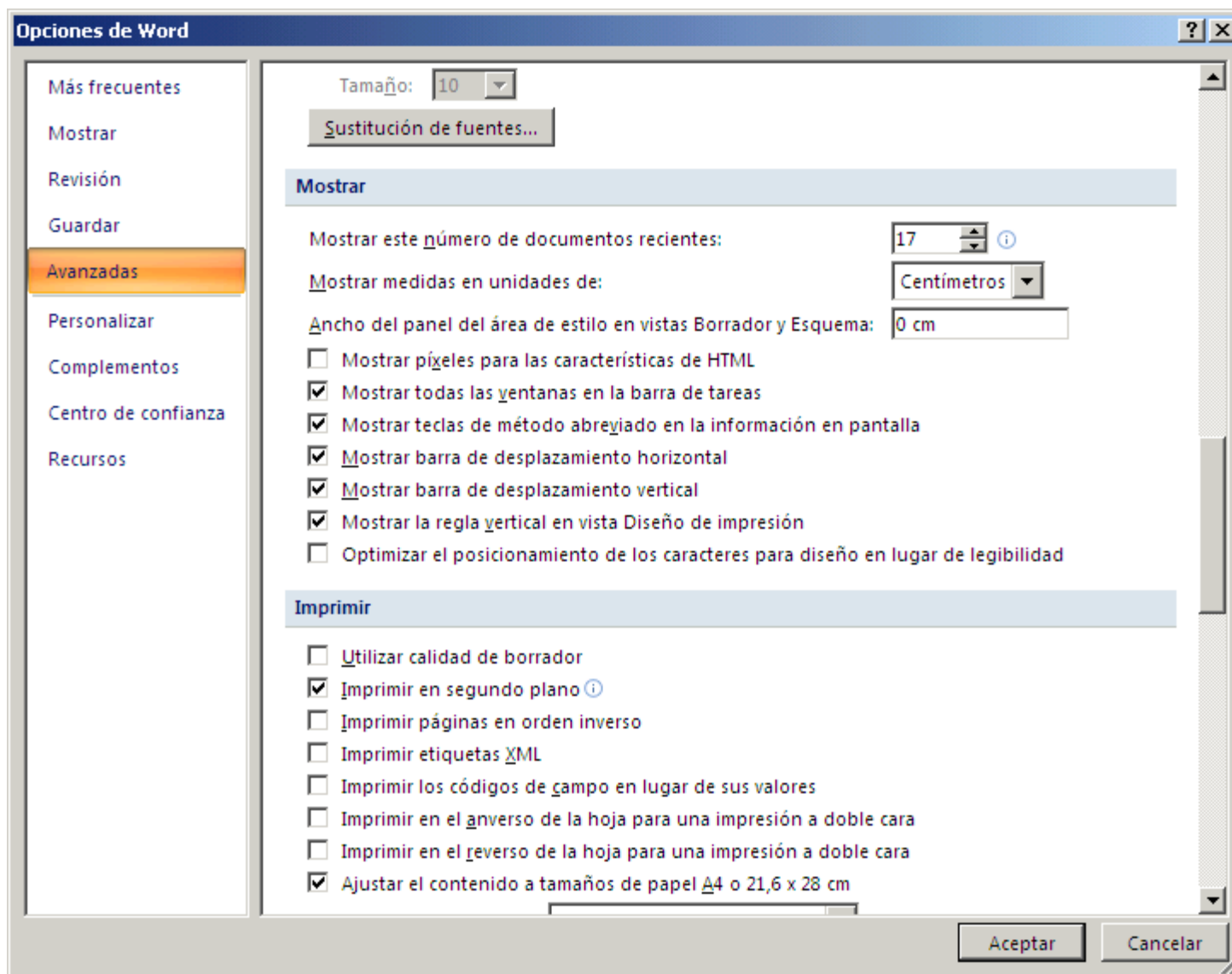
# Valoració macro



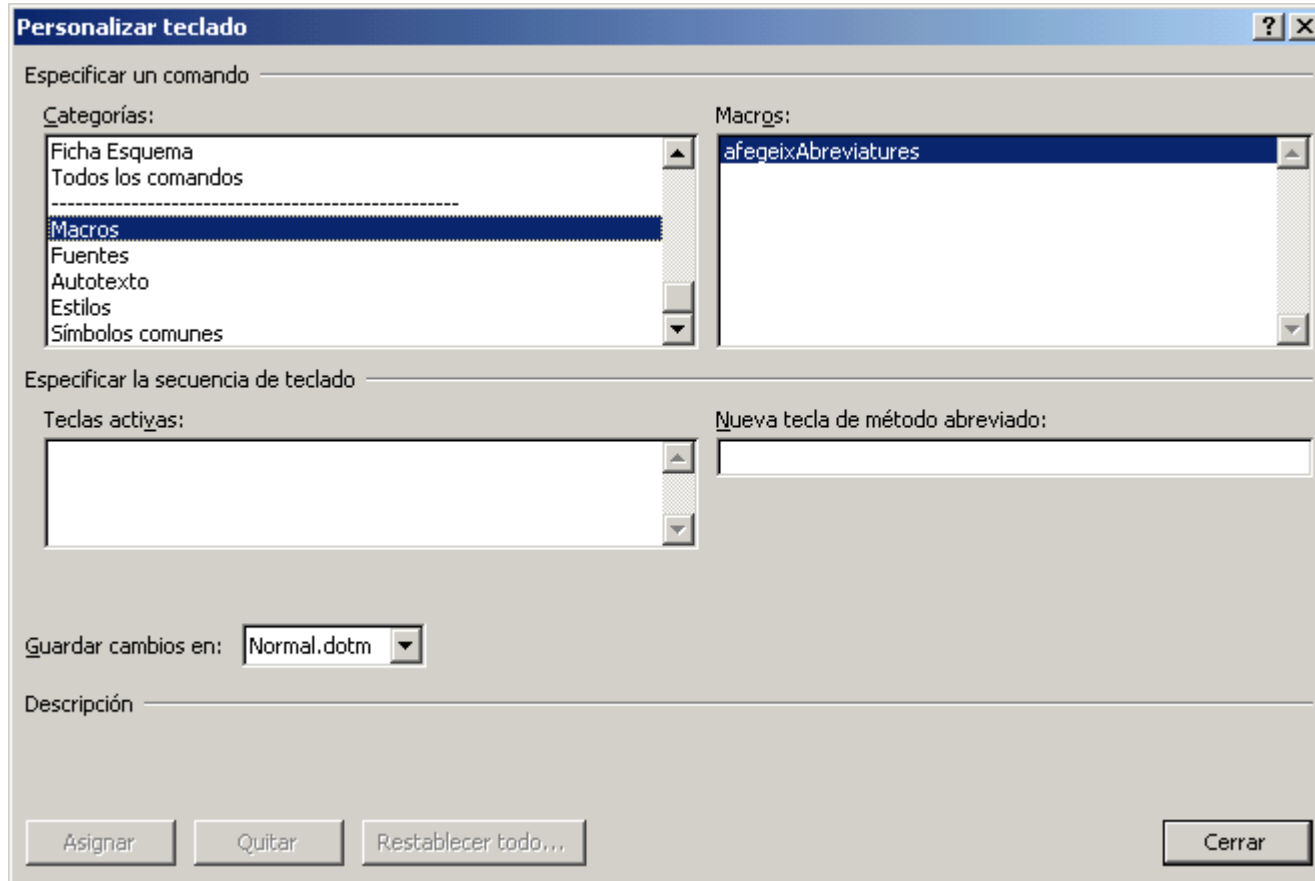
# Valoració macro



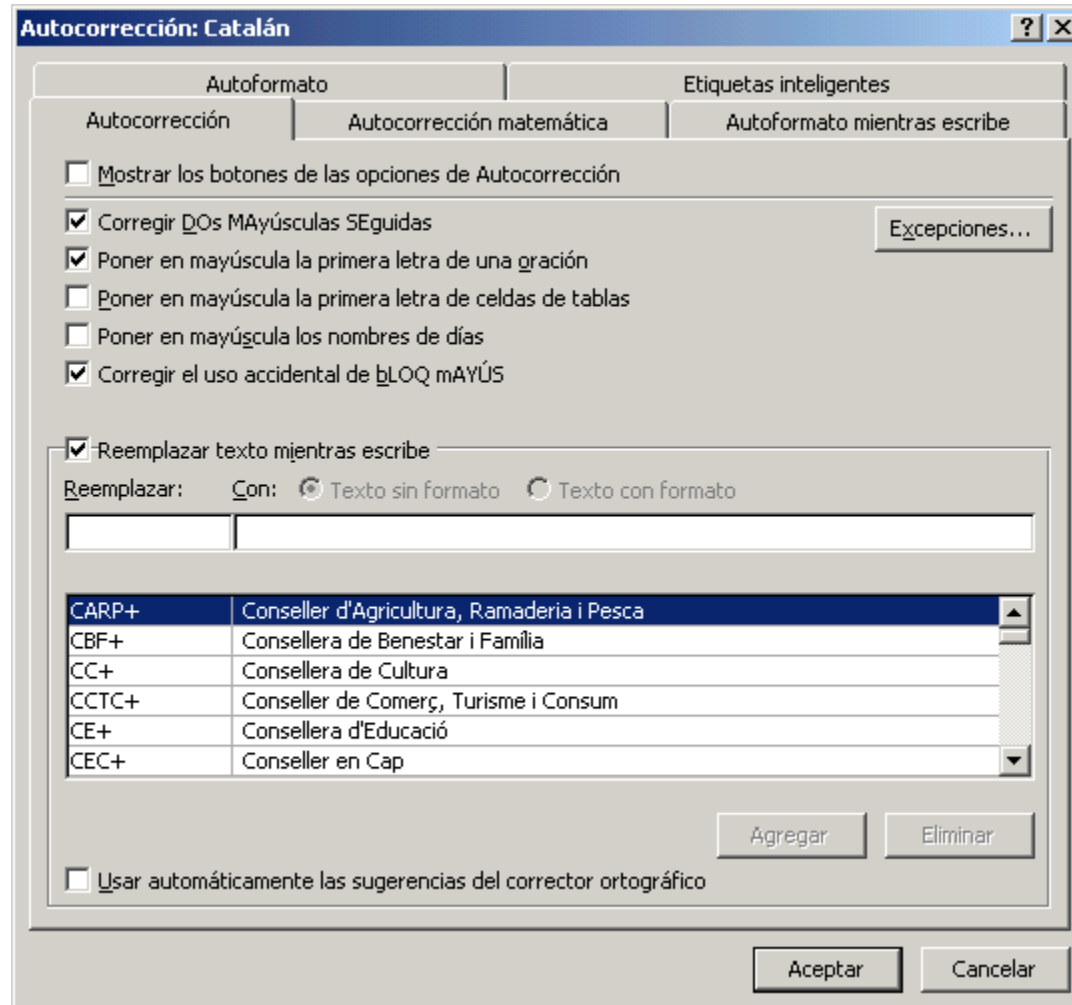
# Valoració macro



# Valoració macro



# Valoració macro





- Traducció automàtica
- Traducció assistida
- Mostra de corpus etiquetat

## Traducció automàtica: introducció

La **Traducció Automàtica** (TA) és un dels camps més investigats dins l'àrea de Processament del Llenguatge Natural (PLN).

Factors de **demanda**: societat de la informació, globalització, multilingüisme (UE), mercat internacional, etc.

No impliquen la substitució de traductors professionals: són eines útils com a **ajuda** a la traducció.

Tipus de textos en què la TA **funciona bé**: butlletins meteorològics, disposicions legals o administratives, textos jurídics, manuals tècnics, butlletins informatius, patents, etc.

Tipus de textos en què la TA **no funciona bé**: textos creatius (poesia, teatre, ressenyes, crítiques, etc.) i expressius (llenguatge periodístic molt directe, llenguatge col·loquial, jocs de paraules, etc.).

## 6 fases de la TA

### Fase 1:

Fins 1955 aprox. Tècnica principal: substitució de paraules (amb diccionaris).

### Fase 2:

Fins 1965 aprox. Molts projectes a tot el món, sobre tot en USA, però les traduccions no aconsegueixen una gran qualitat.

### Fase 3:

Fins 1975 aprox. Es redueixen les inversions en TA, degut als resultats negatius obtinguts i s'analitzen les dificultats.

## 6 fases de la TA

### Fase 4:

Fins 1990 aprox. La Comunitat Europea comença a utilitzar de nou la TA, degut a la gran quantitat de textos a la què s'ha d'enfrontar. També Canadà i Japó reprenen les seves investigacions.

### Fase 5:

Fins l'any 2005. Desànim generalitzat degut als alts costos de desenvolupament de la TA i escassa amortització.

### Fase 6:

Fins ara. Nova etapa en la què els sistemes de TA són empleats pel públic general.

## Mètodes simbòlics vs. mètodes empírics

**Mètodes simbòlics:** coneixement lingüístic codificat explícitament en forma de regles.

Systran (van Slype, 1979; Hutchins, 1986)

Eurotra (Allegranza et al., 1991)

Metal (Schneider, 1989)

Lucy LT KWIK Translator

**Mètodes empírics:** coneixement inferit a partir de les dades per mitjà de tècniques estadístiques.

Moses, Guiza ++ (Knight et al., 1999; Och i Ney, 2003; Koehn, 2004)

Google Translator

### 2 sistemes de TA:

- *Lucy LT* (basat en regles lingüístiques):  
<http://www.lucysoftware.com/english/home/>

- *Google Translate* (basat en tècniques estadístiques):  
<http://translate.google.es/?hl=es>

<http://www.vilaweb.cat/noticia/4091206/20130605/fot-li-castanya-catanyol.html>



## TRANSLATION SOLUTIONS MADE TO MEASURE

Contact | Sitemap | Legal Notice

english

- Home
- Company
- SAP Translation
- Machine Translation**
  - Lucy LT Technology Solutions Research
  - Lucy LT KWIK Translator
- Customers & References
- News

### Lucy LT KWIK Translator



Got a short text or a website that is not in your language and you need a quick translation? Then Lucy KWIK Translator is your tool. Paste in your text or enter the URL of the web page you want translating, choose your languages and off you go!

#### Short texts

Catalan-English

Translate

General Vocabulary

#### Web pages

Catalan-English

Translate

2,155,420,213 Words

#### ... translated in 2008!

If you found the Lucy KWIK Translator useful, imagine what you could do with Lucy Software's professional solutions for enterprises and public institutions!

[More information](#)

#### What Customers Say

**SEGRE**

"The Diari Segre newspaper is issued in both Spanish and Catalan. The machine translation engine provides us with a high-performance solution that is integrated into our overall processes."

Diari Segre Newspaper, Spain

**paymetric**

One Company. One Solution.

"Lucy Software's knowledge and experience saved us when we were stuck. We now know who to turn to in future for any questions related to SAP translation."

# 21963 Informàtica 2013

Imágenes Maps YouTube Noticias Gmail Docs Calendar Más -

Del: español ▾



Al: catalán ▾

Traducir

francés

español

inglés

catalán

cción de un sitio web, o bien, [traduce un documento](#).

Traductor de Google para empresas:

[Google Translator Toolkit](#)

[Traductor de sitios web](#)

[Global Market Finder](#)

<http://translate.google.es/>



## Exemples d'errors

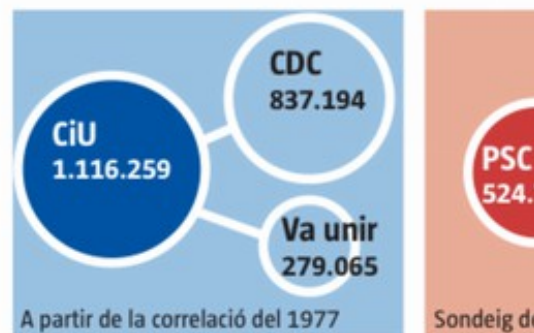
### POLÍTICA

#### No a la noia

José Bono incendia la lluita pel control de la cúpula socialista en declarar: "Crec que hem d'apostar 'a la grande'... i no 'a la chica'", en una possible referència a Carme Chacón. **PÀGINA 18**

*Una hipòtesi de ruptura en e*

### AUTONÒMIQUES



### GENERALS

# Exemples d'errors

*Mundo*. En esta información se indicaba que el pasado 14 de noviembre el corrector automático del DOG cambió al catalán más de un centenar de los 294 apellidos de los opositores. Por ejemplo, Yolanda Hidalgo Cumplido pasaba a ser Yolanda de Gentilhome Complert.

“El traductor automático tradujo” también los apellidos, y “no se puso la atención suficiente de que los nombres propios no se tenían que traducir”, sostuvo la titular de Justicia, que convino que era un

## El corrector automático del DOG catalanizó apellidos de opositores

EL PAÍS, Barcelona

La consejera de Justicia, Montserrat Tura, atribuyó ayer a una “mala jugada” de la tecnología la traducción al catalán de los apellidos de los opositores, cuando el concurso de traslados de la administración judicial fue publicado en el *Diari Oficial de la Generalitat de Catalunya* (DOG), según publicó también ayer el diario *El Mundo*. En esta información se indicaba que el pasado 14 de noviembre el corrector automático del DOG cambió al catalán más de un centenar de los 294 apellidos de los opositores. Por ejemplo, Yolanda Hidalgo Cumplido pasaba a ser Yolanda de Gentilhome Complert.

“El traductor automático tradujo” también los apellidos, y “no se puso la atención suficiente de que los nombres propios no se tenían que traducir”, sostuvo la titular de Justicia, que convino que era un error humano. “Pedimos disculpas a las personas que hayan visto su nombre reconvertido”, dijo Tura, quien recordó que ya se publicó un anexo para subsanar este error.

## Exemples d'errors

### Renuncia por edad el cardenal Ortega, mediador ante el régimen cubano

---

**ELISABET SABARTÉS**  
Mèxic. Corresponsal

---

El cardenal Jaime Ortega, máxima autoridad de la Iglesia católica en Cuba cumplió ayer 75 años y, acatando la legislación canónica, presentó al papa Benedicto XVI su renuncia como arzobispo de La Habana. El Pontífice debe decidir ahora si acepta la preceptiva jubilación o autoriza al prelado a seguir en el cargo, que ocupa desde 1981. La mayoría de cubanólogos considera poco probable

que el Vaticano admita la renuncia de Ortega, cabeza de la organización no gubernamental más poderosa de la isla, cuyo régimen se declaró ateo hasta 1991, y figura reconocida dentro y fuera del país por su papel de mediador ante el presidente Raúl Castro en pro de los presos políticos.

El cardenal abrió el marco de diálogo que permitió la excarcelación de 115 prisioneros de conciencia entre el 2010 y el 2011, en un proceso en el que estuvo apoyado por el Gobierno español. La

### El cardenal Ortega, mediador davant el règim cubà, renuncia per edat

---

**ELISABET SABARTÉS**  
Mèxic. Corresponsal

---

El cardenal Jaime Ortega, máxima autoridad de l'Església catòlica a Cuba va fer ahir 75 anys i, com que acabava la legislació canònica, va presentar al papa Benet XVI la renúncia com a arquebisbe de l'Havana. El Pontífex ha de decidir ara si accepta la preceptiva jubilació o autoritza el prelat a seguir al càrrec, que ocupa des del 1981. La majoria de cubanòlegs considera poc

probable que el Vaticà admeti la renúncia d'Ortega, cap de l'organització no governamental més poderosa de l'illa, el règim de la qual es va declarar ateu fins al 1991, i figura reconeguda dins i fora del país pel paper de mitjançar davant el president Raúl Castro a favor dels presos polítics.

El cardenal va obrir el marc de diàleg que va permetre l'excarceració de 115 presoners de consciència entre el 2010 i el 2011, un procés en què va tenir el suport del Govern espanyol.

# El Periódico vs La Vanguardia

<http://ec.europa.eu/translation/bulletins/puntoycoma/51/periodico.htm>

<http://ec.europa.eu/translation/bulletins/puntoycoma/numeros.html>

[http://www.incyta.com/archivos/Noticia1\\_cat.pdf](http://www.incyta.com/archivos/Noticia1_cat.pdf)

<http://www.fti.uab.es/tradumatica/revista/num4/articles/09/09art.htm>

<http://www.aptic.cat/noticia/charla-como-se-hacen-los-periodicos-bilinguees>

- Traducció automàtica
- Traducció assistida
- Mostra de corpus etiquetat

Traducció assistida: es farà a altres assignatures més endavant

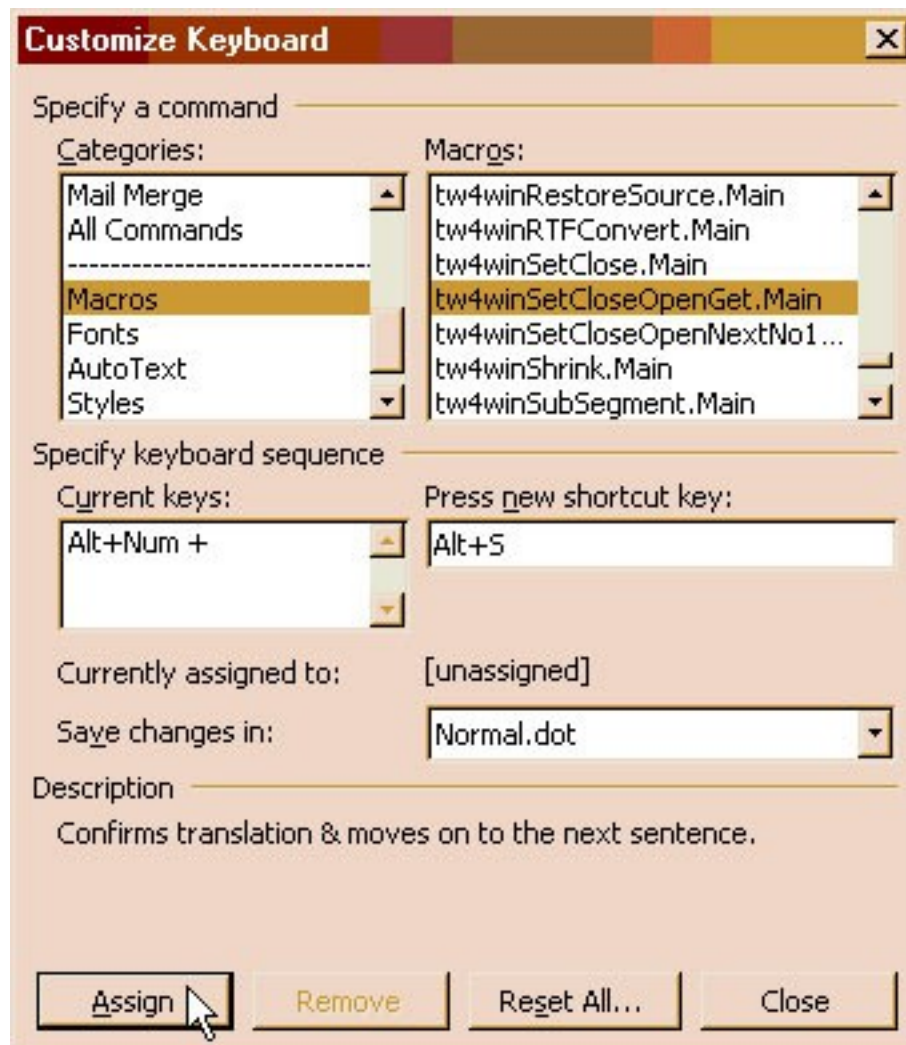
Alguns avançaments:

Memòries de traducció + Gestors de terminologia

- ApSIC Xbench: [http://www.apsic.com/en/products\\_xbench.html](http://www.apsic.com/en/products_xbench.html)
- DéjàVu + base de dades terminològiques: <http://www.atril.com/>
- Transit XV (memòria de traducció) + TermStar (gestor de terminologia): <http://www.star-spain.com/es/tecnologia/transit.php>
- Translator's Workbench (memòria de traducció) + Multiterm (gestor de terminologia): <http://www.trados.com/en/>



# Retrobem les macros \*



# Retrobem les macros \*

## Trados Translators Workbench too slow in Word 2007?

 Posted by **Dallas**


 November 9, 2009

Trados Translators Workbench is actually an add-on to Word. Microsoft has released a patch that solves that speed problem of macro codes and it also solves the speed problem of Trados. If you use other macro programs like Logoport and Translation Workspace and have the same issue, this patch may also help you. You can download the patch locally here at my site:

### Word

[office.microsoft.com/office\\_365](http://office.microsoft.com/office_365)

Utilice Office cuando lo necesite. ¡Infórmese y Pruébalo Gratis!

AdChoices 

You may also download the patch from Microsoft site:

<http://support.microsoft.com/kb/952701>

Open the page and click “View and request hotfix downloads”.

Please also check out [Google Translate for Translators](#), a tool that seamlessly integrates Google Translate to Trados, Word and all text applications.



- Traducció automàtica
- Traducció assistida
- **Mostra de corpus etiquetat**

Al Corpus Tècnic de l'IULA (Bwananet),  
fem la cerca següent:

**a) Definició d'un subcorpus**

a1) Per àmbit temàtic

Subdominis	Documents
Informàtica	
Medi ambient	
Dret	
<b>Medicina</b>	
Genoma	
Economia	
General	

**Medicina**  
mor Organismes  
mgs Planificació i gestió sanitària  
mte Tècniques i equipaments analítics, diagnòstics i terapèutics  
mfs Antropologia, educació, sociologia i fenòmens socials  
mps Psiquiatria i psicologia

Document original/traduït:  Tipus de document:

a2) Per quantitat de paraules:  +/-  %  Màxima quantitat de documents  Selecció aleatòria  
 Mínima quantitat de documents

a3) Recuperar una selecció guardada

seleccionem dos gerundis seguits:

Selecció feta:

Llengua dels documents: Català

Àmbits temàtics seleccionats: Medicina

Nombre de paraules: 2625392

quantitat de documents: 236

**a) Informació específica sobre la concordança** [Afegir més columnes](#)

Unitats	<>	Unitat #1	Unitat #2	Unitat #3	Unitat #4	Unitat #5	</>
- Formes							
- Lemes							
- Categories	<input type="checkbox"/>	v. gerundi	v. gerundi	-	-	-	<input type="checkbox"/>
Repetició		-	-	-	-	-	
Negació		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ordenat per	<input checked="" type="radio"/> no	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
		Ordre alfabètic per: <input type="radio"/> Formes <input checked="" type="radio"/> Lemes					

## centrem els resultats...

Llengua dels documents: Català  
Àmbits temàtics seleccionats: Medicina  
Nombre de paraules: 2625392  
quantitat de documents: 236

Atenció! el nombre de concordances que es mostren està limitat a 50

a: [ pos="VG----" ] [ pos="VG----" ] :: ( ( a.doc\_area="m" ) ) within text

Nombre de concordances: 12 --[Centrar resultats-->](#)

---

- 1 < m00009 >
- 2 < m00025 >
- 3 < m00056 >
- 4 < m00101 >
- 5 < m00178 >
- 6 < m00178 >
- 7 < m00318 >
- 8 < m00352 >
- 9 < m00526 > <s>Així , si realment aquesta isoforma és específica de cèl·lules germinals , s ' hauria d ' observar una ex
- 10 < m00610 >
- 11 < m00610 >
- 12 < m00628 >

Resultats en format: [Text](#)

[Enviar comentaris en relació a aquest resultat](#)

## i els analitzem

varia segons els casos i de vegades ha d' anar **canviant seguint** les diverses etapes de la malaltia en una mateixa persona .</s>  
<s>**Essent important** el grau termomètric desitjat , preveure no sobreescalfar el vapor d '  
substituir - la pe ls cumarínics ( Sintrom ) , bo i **continuant fent** el control hematològic , de bon principi diàriament .</s>  
à de seguir cap mena de dieta , si bé continuarà **sent important** un règim alimentari adequat .</s>  
la és conscient i té ganes que la relació continuï **sent plaent** estalviant - se a l màxim els problemes , descobrirà que les possibilit  
conscient i té ganes que la relació continuï sent **plaent estalviant** - se a l màxim els problemes , descobrirà que les possibilitats d ' obt  
<s>La construcció d ' aquest mapa s ' està **realitzant emprant** STS , YACs , híbrids de radiació , hibridació in situ ( FISH ) i informa  
gens s ' han anat formant i diversificant , i han **anant donant** les opcions que han permès a la selecció natural conduir l ' evolució  
s en que les cèl·lules germinals van proliferant i **madurant donant** lloc a ls espermatozous .</s>  
i MODY 5 constitueixen formes poc freqüents , **havent sent** descrits només en algunes famílies .</s>  
<s>Decantar el **sobrenedant deixant** la capa cel·lular .</s>  
<s>Continuament s ' **estant seqüenciant** nous gens i la informació augmenta de forma extraordinària .</s>

## Seminari 8:

- Treballarem amb els dos sistemes que hem mostrat, els posarem a prova i compararem els resultats.
- Temps d'enquestes