

El català en el sistema de traducció automàtica de Comprendium

I Jornades sobre el català a les noves tecnologies
7 de juny de 2003



Londres

Barcelon

a Munic

Viena

Buchs (Suïssa)

París

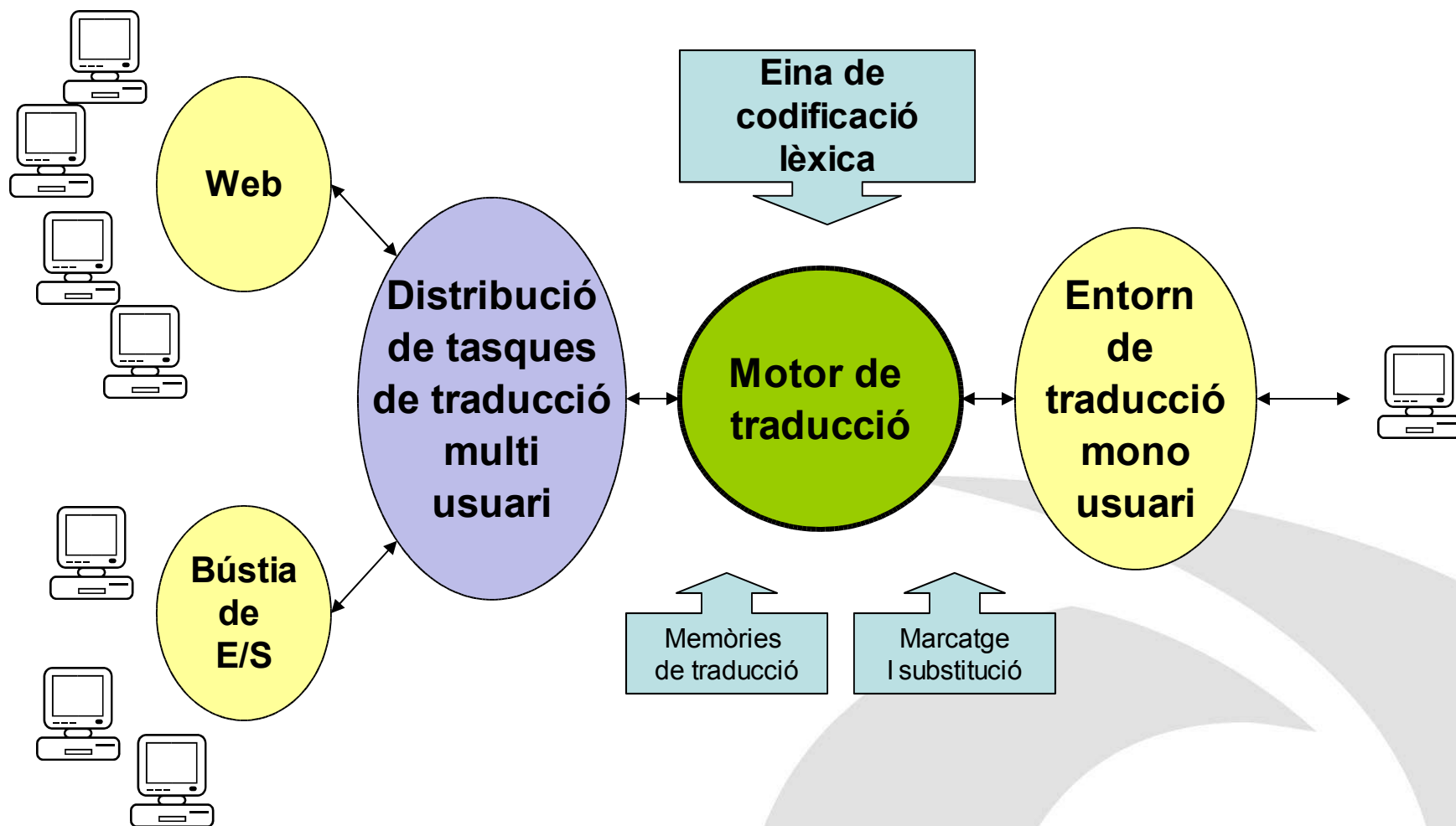
Milà

Desenvolupament de STA
Experiència TA i català (des de 1994)

incyta
MULTILANGUAGE, S.L.

distribuïdor

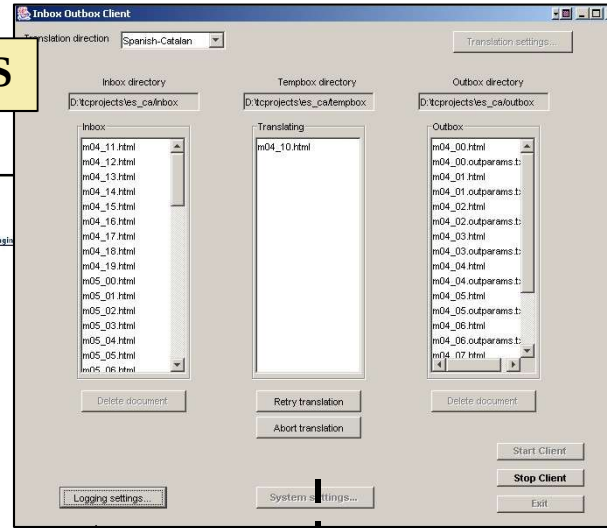
Arquitectura del sistema



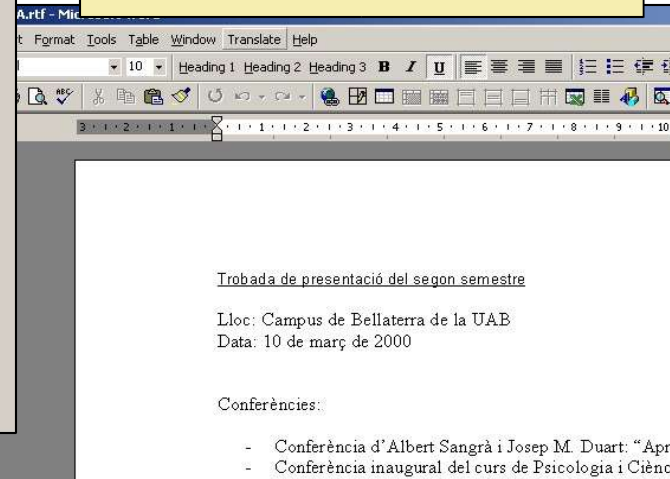
Bústia E/S



Clients basats en Web

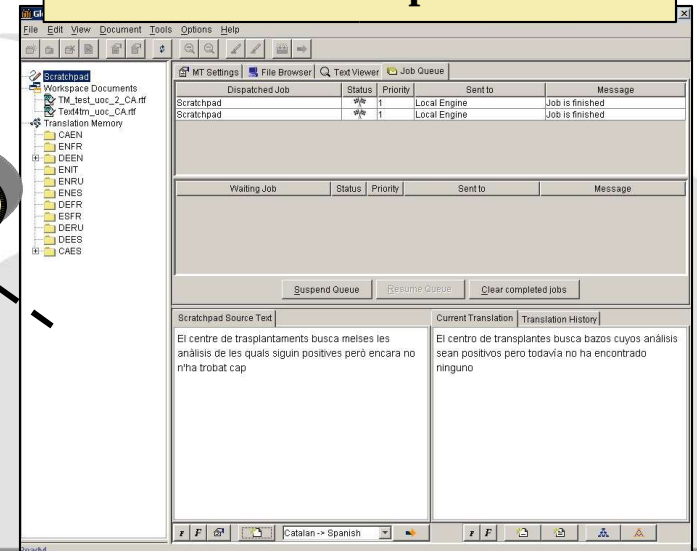


Traducció des de MS-Office



Servidor de traducció

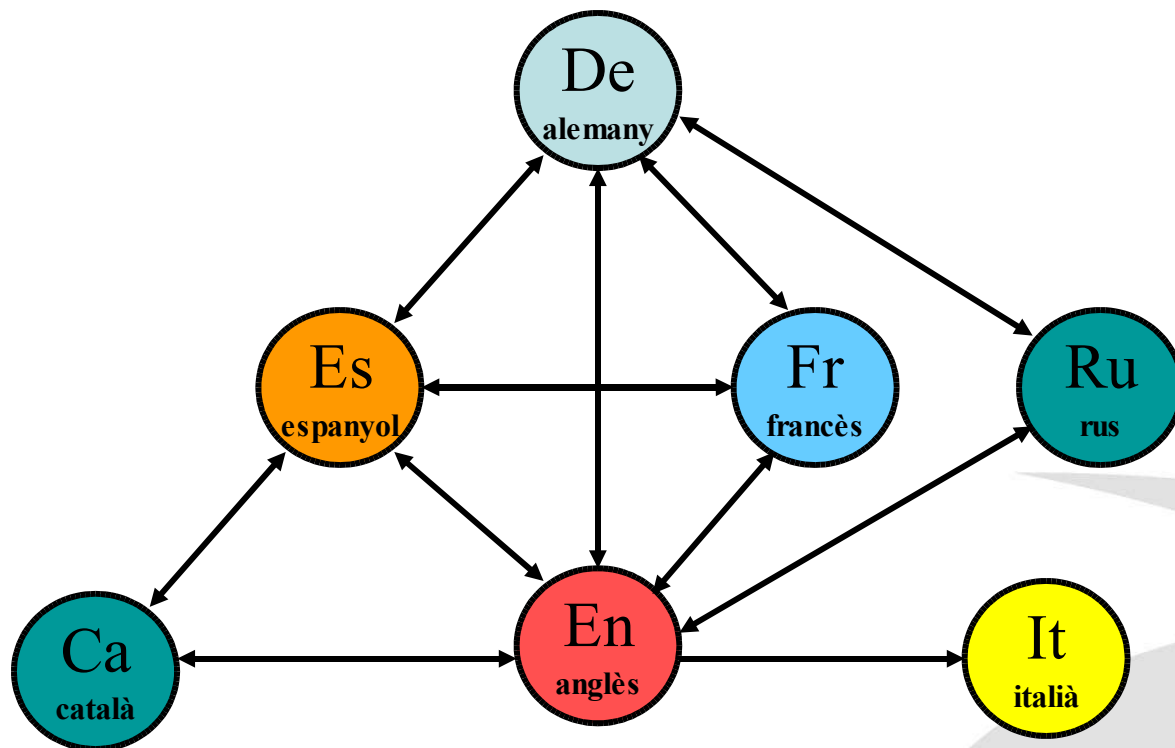
Client de traducció professional



La tecnologia Comprendium

- ✳ Ús de les últimes tecnologies informàtiques: Java, CORBA/SOAP, JSP...
- ✳ Fàcilment ampliable segons les necessitats de volum de traducció
- ✳ Sistema expandible a altres serveis (consulta BD lèxiques, sistemes de cerca d'informació, substituïdors de termes clau, etc.)
- ✳ Alt rendiment: fins a 1.200 pàgines/hora per motor de traducció (P4 a 1,8GHz amb 256 MB RAM)
- ✳ Escalabilitat lineal amb el nombre de motors de traducció i compensació automàtica de càrregues
- ✳ Autenticació d'usuaris i registre de transaccions
- ✳ Recuperació de caigudes del sistema
- ✳ Creació, edició i ús de memòries de traducció
- ✳ Possibilitat de manteniment lèxic per part de l'usuari
- ✳ Connexió d'aplicacions pròpies de l'usuari mitjançant API de programació
- ✳ Integrable dins d'entorns de gestió multilingüe de documentació

Parells de llengües disponibles



anglès <> alemany

anglès <> francès

anglès <> espanyol

anglès <> rus

anglès <> català

anglès > italià

alemany <> francès

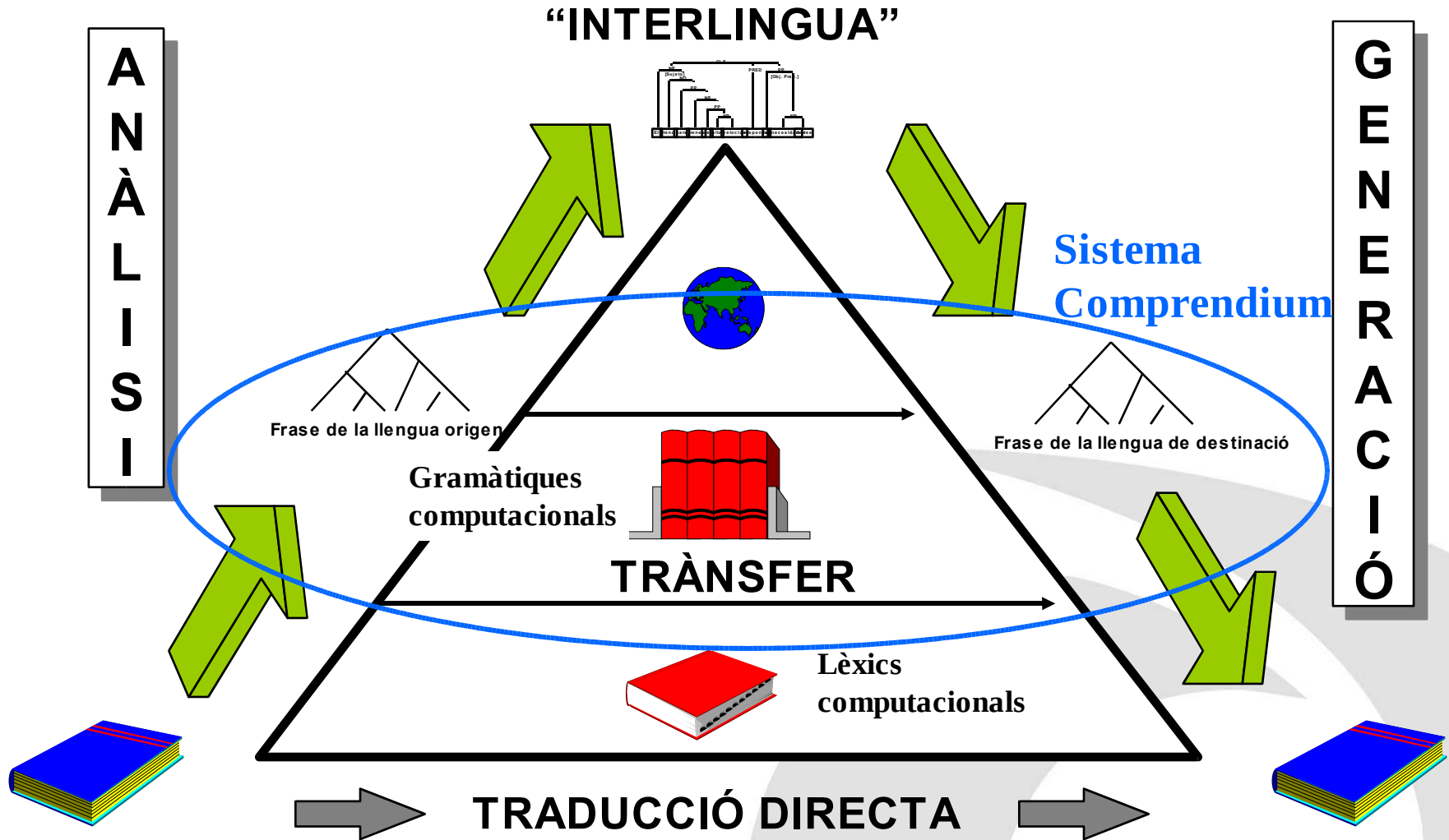
alemany <> rus

alemany <> espanyol

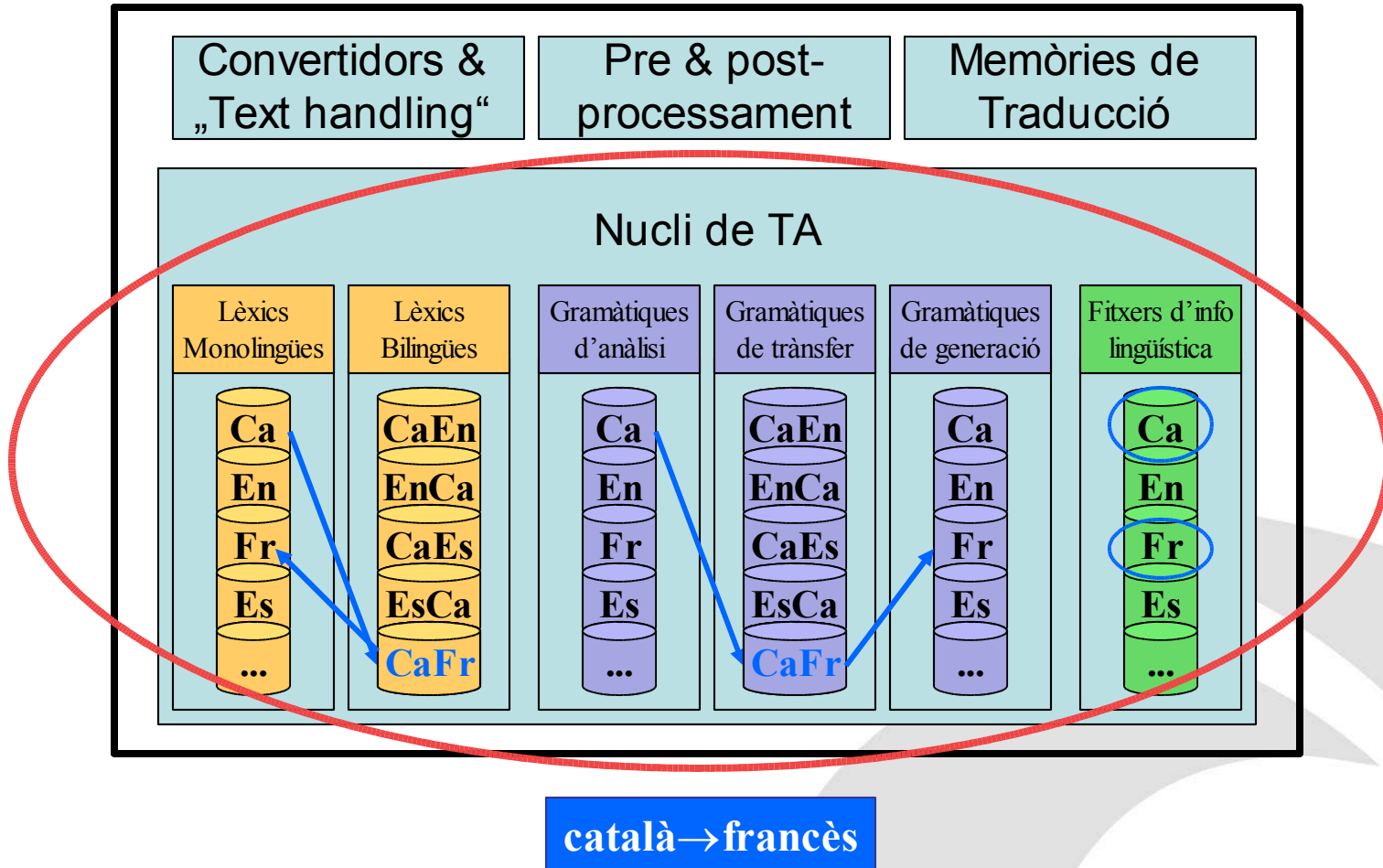
francès <> espanyol

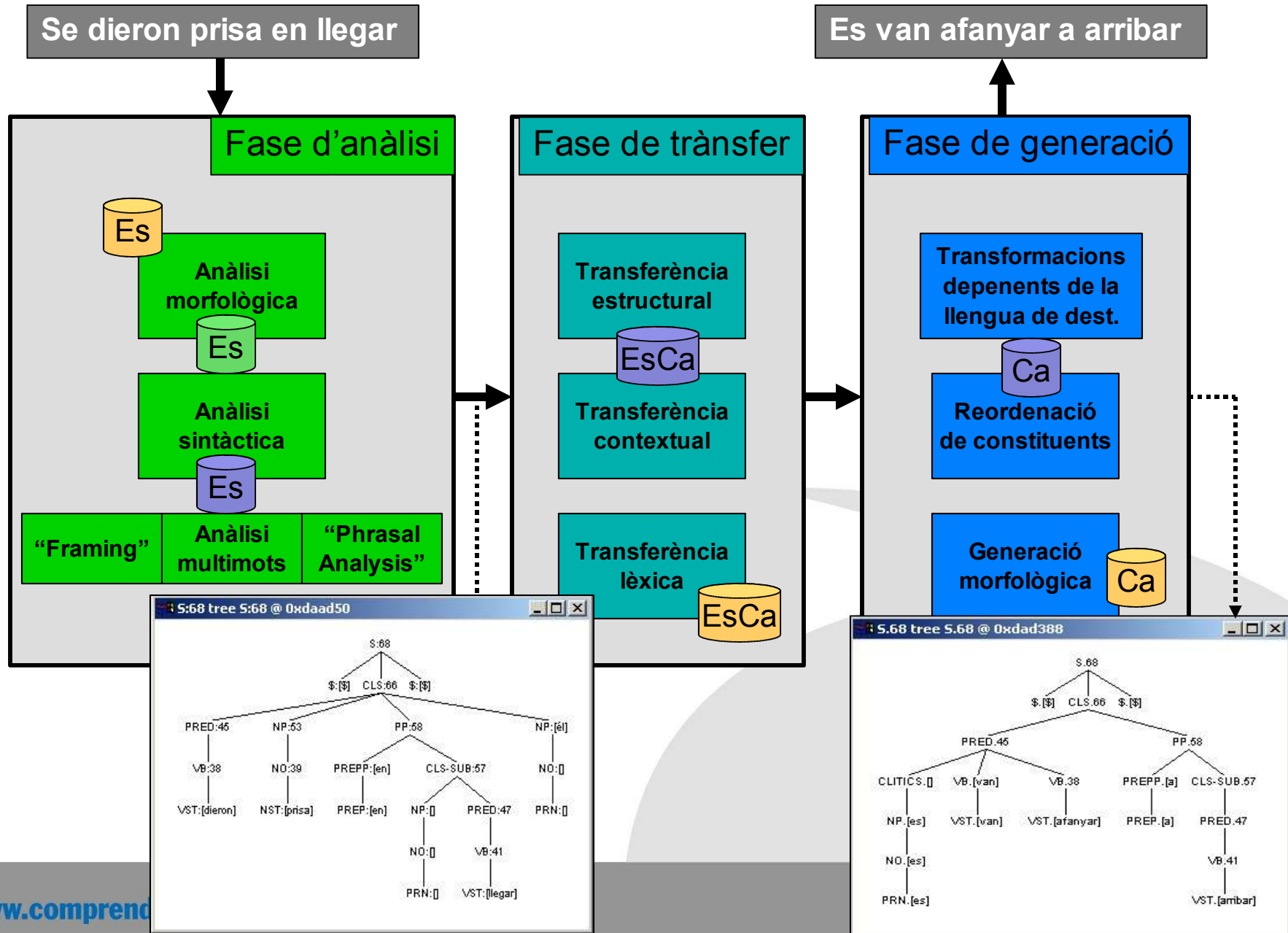
espanyol <> català

Un enfocament de transferència



Els components lingüístics





Potencialitat del component lingüístic

☞ Qualitat de traducció

☞ Sistema modular:

- Reaprofitament dels mòduls per a nous parells de llengües
- Uniformitat en la cobertura

☞ Ús dels components per al desenvolupament d'altres eines lingüístiques

- Interfícies en llenguatge natural
- Eines d'autoaprenentatge per al català
- Correctors intel·ligents ortogràfics i gramaticals
- Diccionaris de valències verbals
- Eines intel·ligents d'extracció i cerca d'informació
- ...

☞ Estudis computacionals del català

☞ Posem el català al nivell computacional de les altres llengües europees

Usos dels STA

- **Ús “informatiu”:** per superar les barreres de l’idioma
- **Ús “productiu”:** per aconseguir traduccions professionals més ràpides i econòmiques

Correu electrònic

COMPRENDIUM
ENTERPRISE SOFTWARE TECHNOLOGIES

English Deutsch Français Español **Català**

Traductor en línia **Text** Documents

Traducció de textos de fins a 4096 caràcters:

Direcció de traducció: Català-Anglès

Àrea temàtica: Vocabulari general

Text:

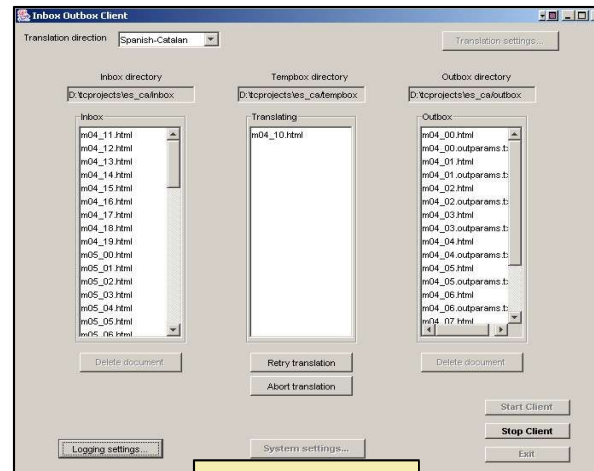
Traducció

Difusió de documentació

Institucions catalanes participen en projectes interestatals o europeus



Barcelona
(català)



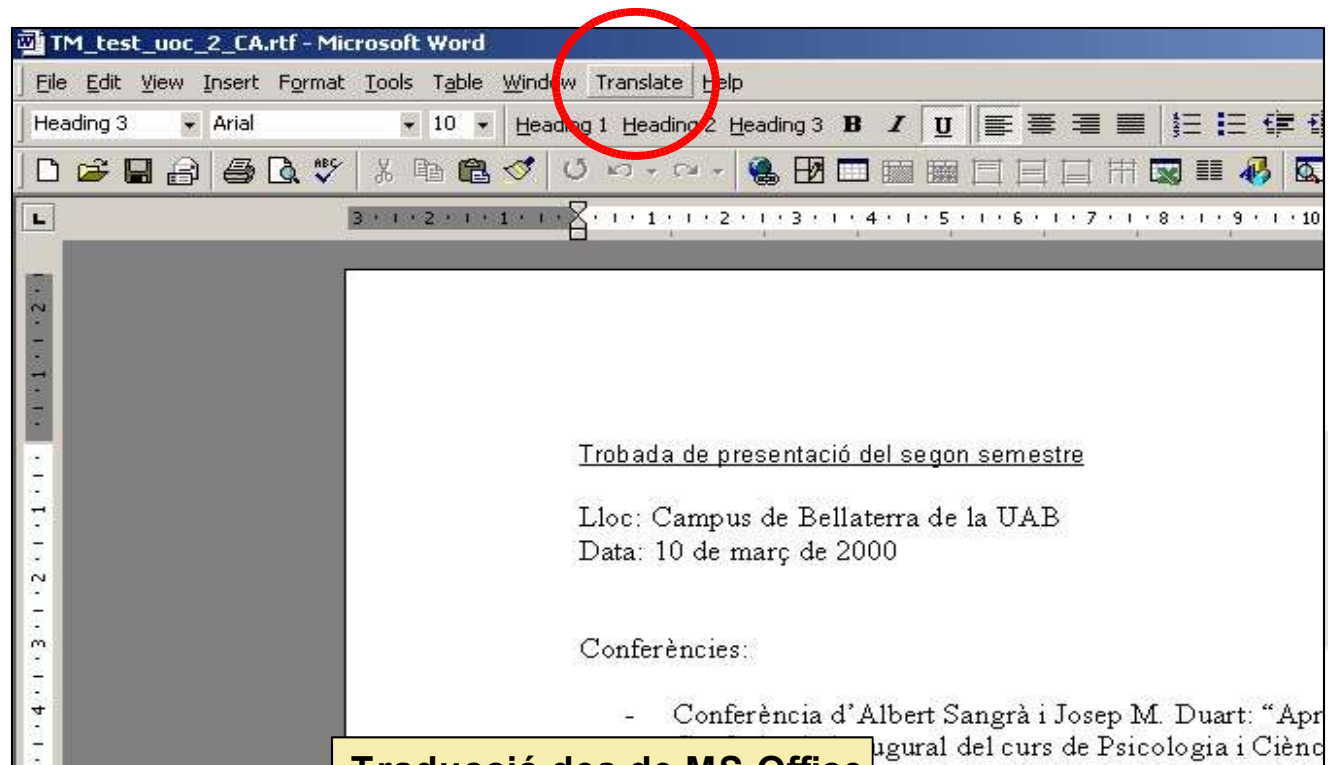
Bústia E/S



Madrid
(castellà)

Generació de documentació multilingüe

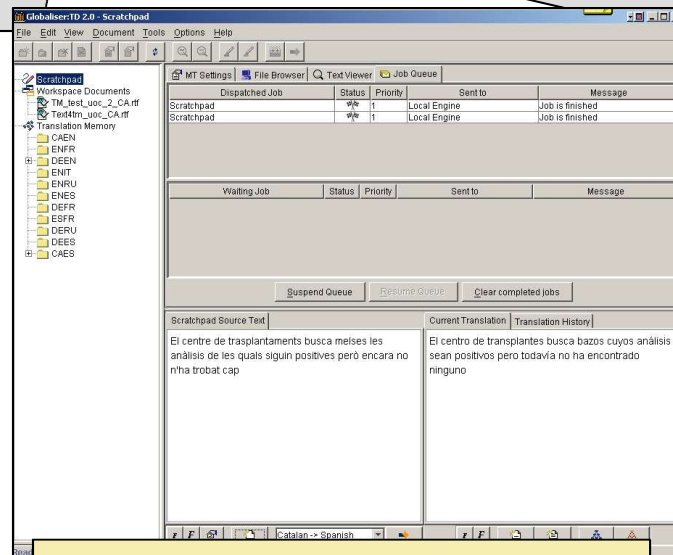
Una empresa internacional on es generen documents en diferents idiomes



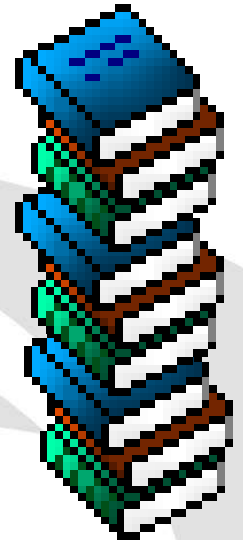
Traducció des de MS-Office

Traducció de documents

Un traductor professional vol reduir el temps i l'esforç de producció i abaratir costos



Client de traducció professional



Navegació per Internet

Volem navegar en una llengua que entenguem

The image displays two side-by-side screenshots of the Russian news website 'IZVESTIA.RU' in its English interface. The left screenshot shows the homepage with a red circle highlighting the article 'НАТО ПРОЯВИТ ВОЕННУЮ СДЕРЖАННОСТЬ' (NATO will show military restraint). The right screenshot shows the same article with a red circle highlighting the title 'THE NATO WILL SHOW THE WAR SELF-CONTROL'.

Left Screenshot (Home Page):

- Header: Новости · On-line · Форумы · Рассылка · Финансовые Известия
- Logo: ИЗВЕСТИЯ.RU
- Navigation: ВСЕ НОМЕРЫ, РАЗДЕЛЫ, ЦИТАТА ДНЯ, НОВОСТЬ ЧАСА
- Article Highlight: **НАТО ПРОЯВИТ ВОЕННУЮ СДЕРЖАННОСТЬ** (NATO will show military restraint). Subtitle: А российская армия не войдет в Афганистан. Author: Иван АТЛАНТОВ.
- Image: A photograph of soldiers in military gear.

Right Screenshot (Article Page):

- Header: The news of · On-line · the Forums · Distribution · Financial Izvestias
- Logo: ИЗВЕСТИЯ.RU
- Navigation: THE WHOLE NUMBER, THE PARTS, THE QUOTATION OF THE DAY, THE NATO WILL SHOW "THE WAR SELF-CONTROL", IN THE NUMBER: TODAY
- Article Highlight: **THE NATO WILL SHOW "THE WAR SELF-CONTROL"**. Subtitle: And the Russian army will not enter Afghanistan. Author: A henry ФИЛДИНГ, the English writer.
- Image: A photograph of soldiers in military gear.

Traductor en línia rus → anglès

BBC i CATEGORIES TV RADIO COMMUNICATE WHERE I LIVE INDEX SEARCH

Text only | MAKE THIS YOUR HOMEPAGE **BBC i** About the BBC | Friday 6th June 2003

SEARCH THE WEB

☒ the Web ☐ BBCi only ☐ BBC News

Popular searches right now are:
Why Search the Web with BBCi?



D-Day anniversary

Today marks 59 years since D-Day. Find out more about it and the people who remember it.

WWII in depth
People's WWII experiences
Listen to soldier's stories
Trace your family tree

NEWS

Audio | Video | Mobile



- Blix criticises coalition over Iraq weapons
- Mail trains to be scrapped

News in 43 languages فارسی|عربی

SPORT

Audio | Video | Mobile



- Barca claim Beckham deal
- England aim for big total
- Vickery to lead England

TV

- BBC TV schedules
- Digital TV
- A-Z of BBC programme websites

RADIO

- Launch Radio Player
- 1Xtra, 6 Music, BBC7
- Asian Network
- Radio 1/2/3/4/5Live

WHERE I LIVE

For local information, enter your UK postcode or town name

e.g. HR4 9AR or Hereford

Weather for London



cloudy
min 14°C max 19°C

View the 5 day forecast

National sites:

- England
- Scotland
- Wales/Cymru
- N. Ireland



BBCi DIRECTORY

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Business & Money

Small Business, Money, News, Programmes, Work & Careers...

Lifestyle

Antiques, Food, Gardening, Homes, Competitions...

Children's

CBBC: Xchange, Cartoons, C-Agents, CBeebies: Fimbles, Games, Stories...

Music

News, Reviews, TV/Radio, Gigs/Concerts, Listen...

Education & Learning

Schools, Colleges, Revision, Languages, Subject Listing...

News

UK, World, Business, Science, Entertainment...

Entertainment

Collective, Cult, Films, Games, Soaps...

Science & Nature

Animals, Birds, Dinosaurs, Space...

WORLD SERVICE

Traductor en línia anglès → ...

BBC i CATEGORIES TV RÀDIO COMUNIQUIS ON VISC ÍNDEX RECERCA Go

Text només | FACI AIXÒ LA SEVA PÀGINA INICIAL **BBC i** Sobre el BBC | Divendres 6 de juny de 2003

FACI UNA CERCA AL WEB

 SEARCH

☒ el Web ☐ BBCi només Notícies de ☐ BBC

Les recerques populars ara mateix són:
Per què Fer una cerca al Web amb **BBCi**?



Aniversari D-Day

Avui marques 59 anys de llavors ençà **D-Day**. Esbrini més sobre això i la gent que el recorden.

WWII a fons
Les experiències de **WWII** de la gent
Escolti les històries dels soldats
Localitzi el seu arbre genealògic

NOTÍCIES

Audio | Vídeo | Mòbil



Blix critica coalició sobre armes d'Iraq
Trens correu per ser desballestats

Notícies en 43 llengües **فارسي|عربي**

ON VISC

Per a la informació local, introdueixi el seu codi postal de Regne Unit o nom de ciutat

 GO

e.g HR4 9AR o Hereford

GUIA de BBCi

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Negoci & Diners
Empresa petita, Diners, Notícies, Programes, Treball & Carreres ...

Nens
CBBC: Xchange, Còmics, Agents de centigrad
CBeebies: Fimbles, Jocs, Històries ...

Educació & Aprenentatge
Escoles, Escoles, Revisió, Llengües, Llistant Tema ...

Diversió
...

Estil de vida
Antiguitats, Menjar, Jardineria, Cases, Competicions ...

Música
Notícies, Ressenyes, TV/Ràdio, Cabriolés/Concerts, Escollar ...

Notícies
Regne Unit, Món, Negoci, Ciència, Diversió ...

Ciència & Natura
Animals, Ocells, **Dinosaurs**, Espai, Humà Cos ...

ESPORT

Audio | Vídeo | Mòbil



Quantitat de **Beckham** de declaració **Barca**
Anglaterra aspirar a total gran **Vickery** dirigir Anglaterra

Temps per a Londres



núvol
màx. de 14°C mínim 19°C

Vegi la previsió de dia de 5

TV

Horaris de **BBC TV**
TV digital
A-Z de llocs web de programa de **BBC**

RÀDIO

Intèrpret de Ràdio de **Llanxa**
1Xtra, 6 Música, BBC7
Xarxa A
Ràdio 1
Live

Llocs nacionals:
Anglaterra
Escòcia
Wales / Cymru
N. Irlanda



Teatre, mes ...

Traductor en línia ... → català



El català al nivell computacional de les altres llengües europees



Gràcies per la vostra atenció

Comprendium España S.L.
Balmes 114, 5è 08008-BCN
Tel: 93 492 01 60
Fax: 93 492 01 61
www.comprendium.biz