

# The Evolution of the European Machine Translation Programme at the European Patent Office

**Patrizia Biani**

Country coordinator

European Affairs/Member States

European Patent Office

Saint Raphaël, 28 May 2010



# The European Patent Office



## Mission

As the patent office for Europe, we support **innovation, competitiveness and economic growth** across Europe through a **commitment to high quality and efficient services** delivered under the European Patent Convention.

- Second largest intergovernmental institution in Europe
- Not an EU institution
- Self-financing, i.e. revenue from fees covers operating and capital expenditure



## MT services are relevant to the EPO because they...

- Provide access to patent information to enterprises, researchers and technically qualified users in Europe
- Support the London Agreement
- Serve as a contribution to resolving the translation/language issue related to the Community patent
- Enable examiners to search prior art

# The dawn of MT for patents at the EPO: 2004

- Approval of the European Machine Translation Programme (EMTP) by the Administrative Council of the EPO
- Objective: Provide an automated translation service of a sufficient quality to make the technical content of a patent document understandable to a technically qualified person
- Study and Call for tender: only rule based engine bids received
- Quality assessment: EPO selected WorldLingo (using Systran)
- Technical approach used: rule-based engine, hierarchical technical dictionaries built with IPC-based patent terminology

# An insight in the creation of technical dictionaries

1. Select, scan and OCR patent documents to acquire matching text in source and target language (NPO & EPO).
2. Align source and target texts on sentence or paragraph level (EPO).
3. Automatically extract terms and their translations from aligned text (external provider).
4. Select term candidates for inclusion in technical dictionaries (EPO).
5. Validate final set of dictionary terms (translation, grammatical information) (external provider).
6. Build bi-directional dictionaries (EPO).
7. Test in Test environment (NPO & EPO).
8. Deploy in Production environment (translation engine provider).

## In the meantime...

- 2008: first language pairs, EN-ES/ES-EN and EN-DE/DE-EN, entered into production.
- 2008/9: two further language pairs, EN-FR/FR-EN and EN-IT/IT-EN, entered into production - but improvement still ongoing (quality not satisfactory)
- As per 1 July 2008 IT/EN translation service used for "WOIT" files - enables EPO examiners to carry out prior-art searches and prepare written opinions for Italian files
- 2009: high-quality dictionaries created for SE and PT - interaction with engine delivers poor quality - implementations on hold
- 2010: a SMT (Language Weaver) selected for the translation of Italian files due to the persistency of insufficient quality

# Some figures

	German		Spanish		French		Italian		Portuguese		Swedish	
	DE-EN	EN-DE	ES-EN	EN-ES	FR-EN	EN-FR	IT-EN	EN-IT	PT-EN	EN-PT	SE-EN	EN-SE
No. documents (5-50 pgs/doc)	871.000		168.046		871.000		108.500		84.885		200.493	
No. created XML files	250.137		42.366		147.972		63.781		32.789		N/A	
No. aligned sentences	7.000.000		5.768.314		4.567.825		6.069.820		3.782.037		N/A	
No. Dictionary terms/words	386.204	332.681	274.979	274.995	213.602	182.933	795.854	764.664	118.071	126.675	1.385.439	1.378.751
Human acceptance score for translation in production	scale (3-9) score: 6		scale (3-9) score: 6		(1-5) score: 4,3	(1-5) score: 3,25	(1-5) score: 2,89	(1-5) score: 2,82	N/A	N/A	N/A	N/A

- The scores for French and Italian language results from EPO internal acceptability test
- In dictionaries the same terms appeared in (for example in 5) different IPC-dictionaries are counted 5 times.
- The score 6 on the scale (3-9) is close to the score 3 on the scale (1-5)



## EPO MT services are available...

- to the public via esp@cenet (abstract, descriptions and claims)





<http://ep.espacenet.com>



- to the EPO examiners via SEA Viewer from Epoque

# Lighting rail for use in kitchen or bathroom comprises rectangular section profile receiving square support plates for electrical fittings

- Bibliographic data
- Description
- Claims
- Mosaics
- Original document
- INPADOC legal status

**Publication number:** DE19848637 (A1)  
**Publication date:** 2000-03-02  
**Inventor(s):** VERON CHRISTIAN [FR] +  
**Applicant(s):** M S A DIFFUSION Z A LES GENETS [FR] (M.S.A. DIFFUSION Z.A. "LES GENETS", SAINT RAPHAEL)  
**Classification:**  
- international: *F21S4/00; F21V21/34; F21S4/00; F21V21/34; (IPC1-7): F21S4/00; F21S8/02; F21V19/00*  
- European: *F21S4/00A*  
**Application number:** DE19981048637 19981022  
**Priority number(s):** FR19980010200 19980807

**Also published as:**  
 FR2782146 (A1)  
 ES2152179 (A1)  
 ES2152179 (B1)  
 ITTO980889 (A1)

**Cited documents:**  
 DE19511042 (A1)  
 GB288053 (A)

[View all](#)

[View INPADOC patent family](#)  
[View list of citing documents](#)

[Report a data error here](#)

Abstract not available for DE 19848637 (A1)  
Abstract of corresponding document: **FR 2782146 (A1)**

[Translate this text](#)

The square supports have offset attachments, and may be rotated to achieve required positioning. The lighting rail comprises a C-shaped rectangular section profile (1) with a base (11) and two sides (12, 13). Folded edges (14, 15) lie parallel to the base



Select target language for the translation of the description of document 2782146

| [French](#) | [German](#) | [Italian](#) | [Spanish](#) |



Abstract of FR2782146

[Print](#)

## Result Page

Notice: This translation is produced by an automated process; it is intended only to make the technical content of the original document sufficiently clear in the target language. It is not intended to replace professional translation services. The esp@cenet® Terms and Conditions of use are also applicable to the use of the translation tool and the results derived therefrom.

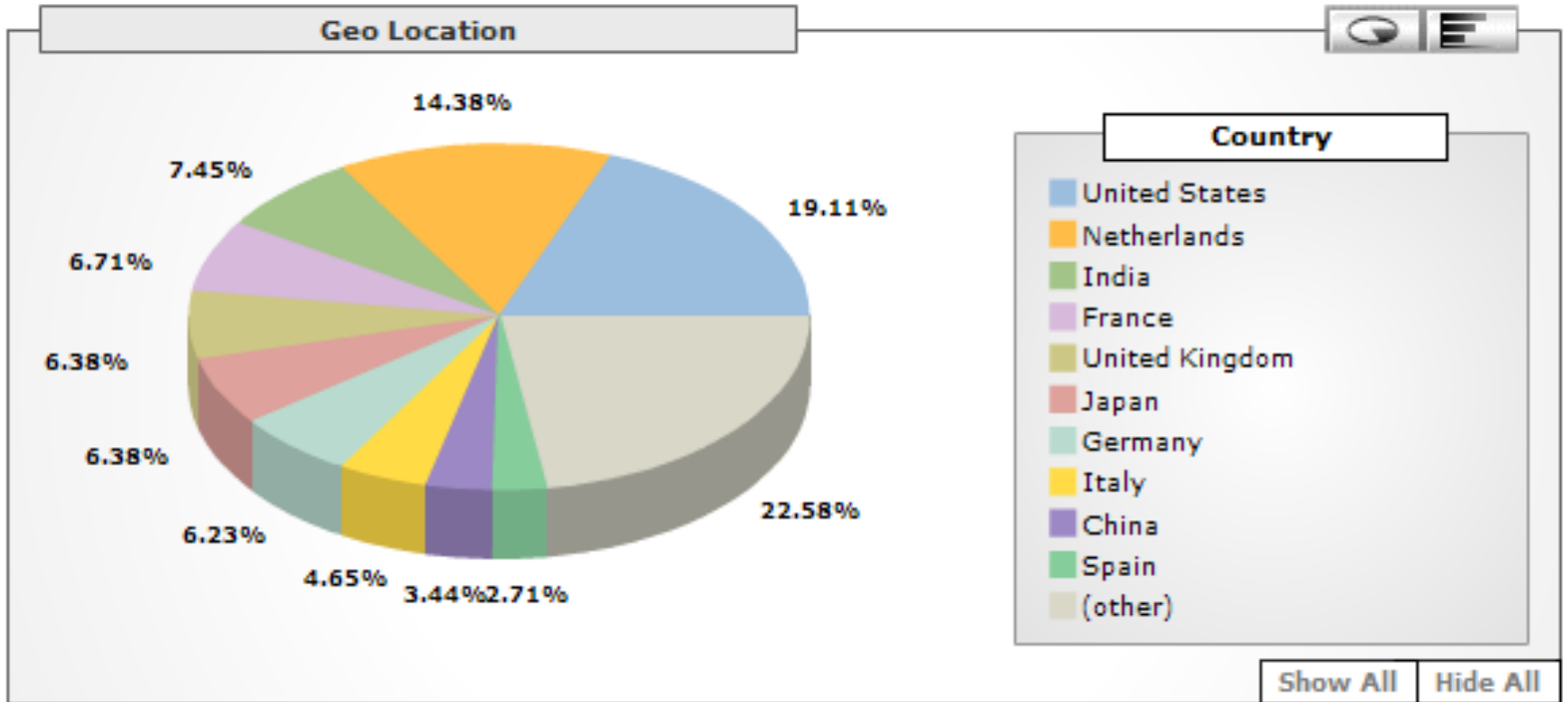
Los soportes del cuadrado tienen fijaciones desviadas, y se pueden girar para lograr posicionado necesario. El carril de la iluminación comprende A.c. - perfil con dos lados (12.13). 14.15 mentiras plegada de los bordes (paralelas a las bases y completas la sección rectangular C-conformada. Cada accesorio eléctrico se correderas el carril, retenido por los bordes plegados. Un receptor se monta en cada placa para alojar el accesorio eléctrico. Este receptor es w.r.t. desplazado. e placa de base en una de cuatro diferentes orientaciones, el accesorio eléctrico puede adoptar una variedad de las diferentes posiciones w.r.t. el carril.

## Number of esp@cenet translation requests (Jan-Apr 2010) ...

- ES → EN: ca 1.800
- EN → ES: ca 33.000
  
- DE → EN: ca 127.000
- EN → DE: ca 20.000
  
- FR → EN: ca 52.000
- EN → FR: ca 50.000
  
- IT → EN: ca 3.500
- EN → IT: ca 20.000

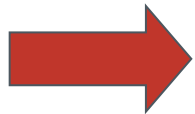
# ... and their geographical origin

## Countries



## Technical limits of the current approach reached

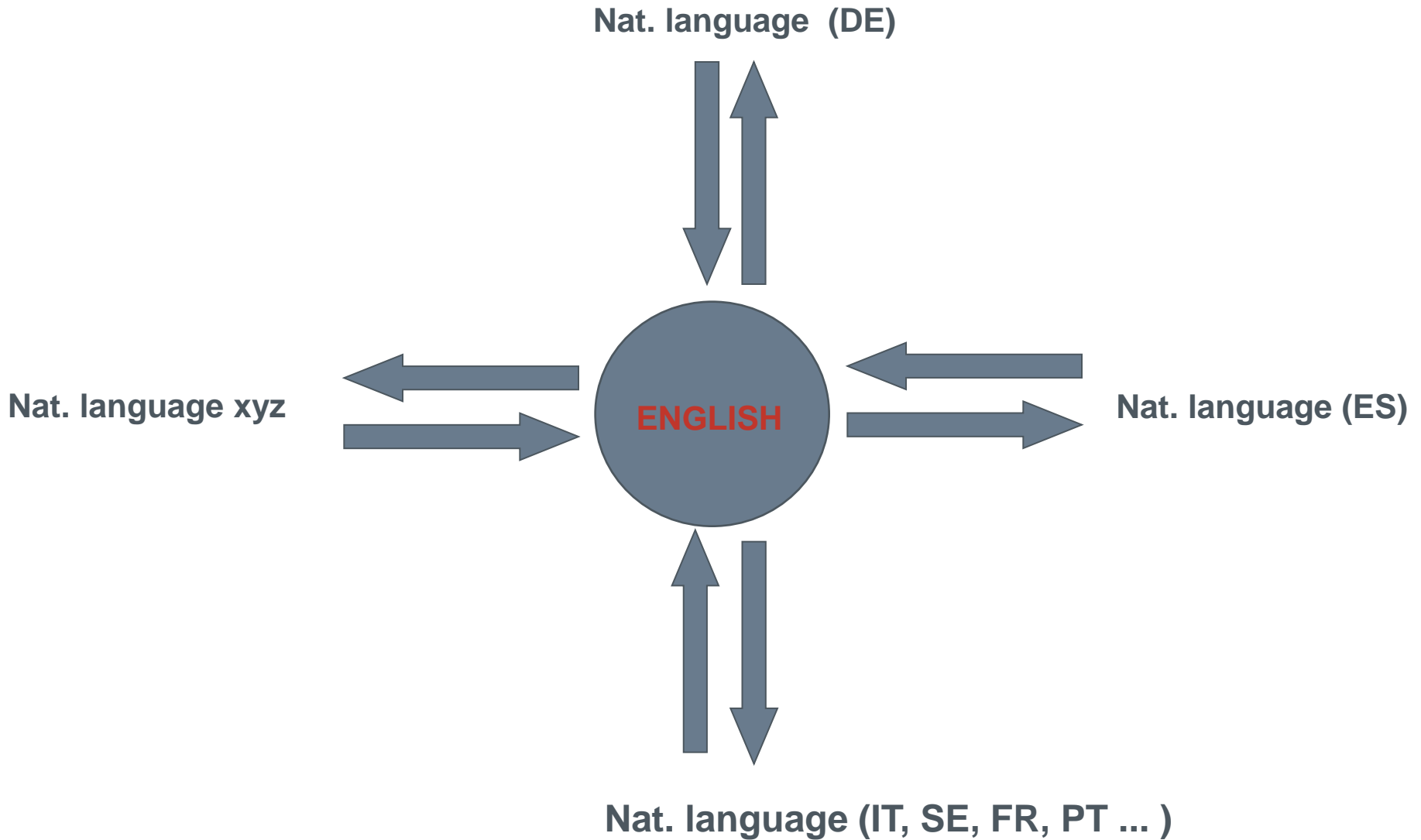
- Implementation of further language pairs on hold due to:
  - insufficient quality of current engine / technical approach
  - no suitable rule-based translation engines for certain EPO languages (e.g. RO)



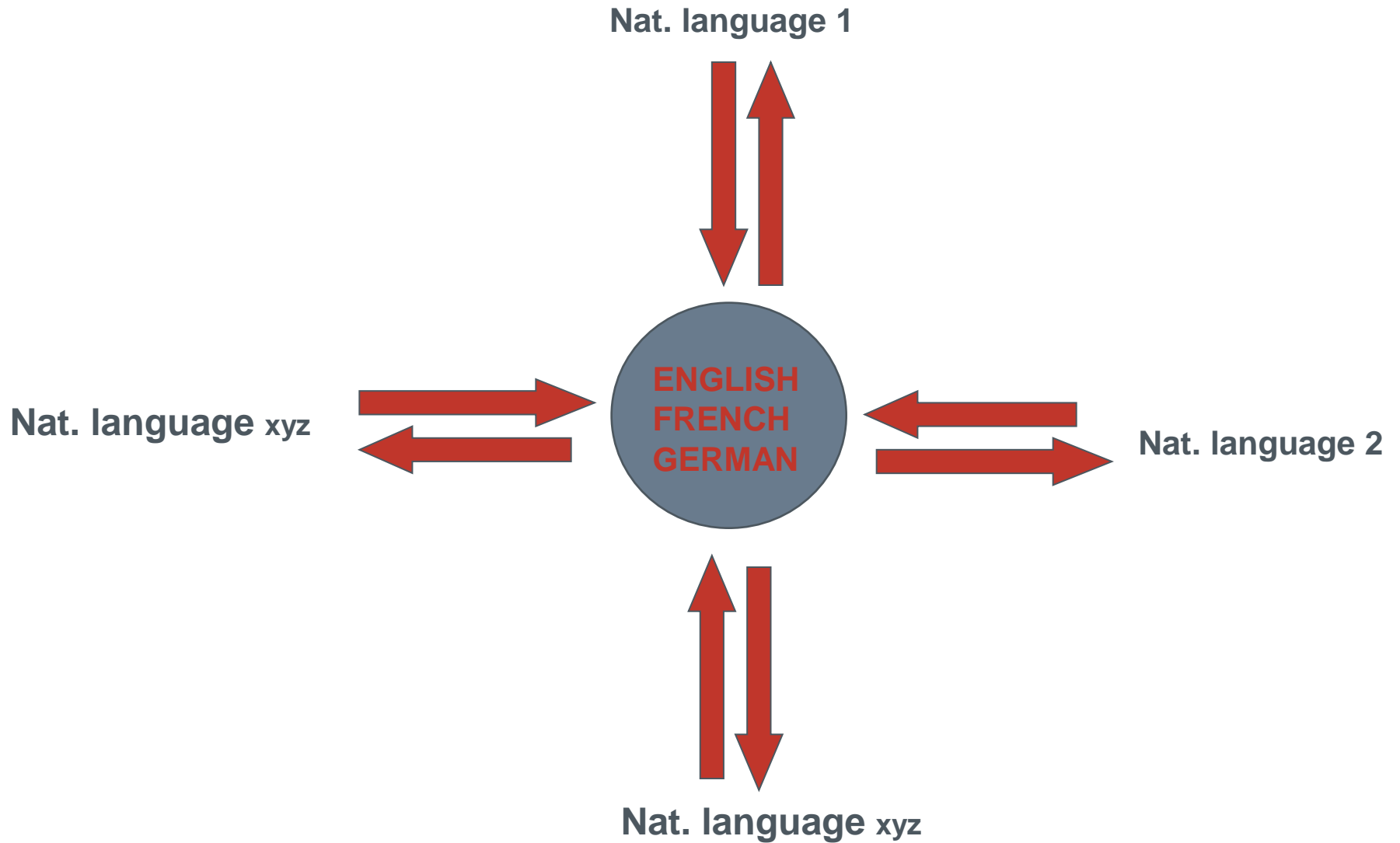
**need to move on to a new concept**

- Identify the most suitable MT technologies and approaches in order to offer, internally and externally, MT solutions which:
  - are tailor-made for translating patent documents providing the necessary accuracy
  - offer scalable machine translation services from the three EPO official languages into **ALL** member states' languages

# What we have today...



# ... and what we will need in the future





## Result: a new MT co-operation programme

- a new comprehensive MT programme is drafted
- to be approved by the EPO Administrative Council  
(objective: October 2010)
- within the co-operation framework existing between EPO and member states
  - it serves the users of the European patent system
  - it needs broad support and effort

## Elements of the new MT co-operation programme

- Contribution from the national patent offices of the member states (patent documents)
- MT technology monitoring (interaction with leading MT research groups and business entities)
- Possibility to use more providers in parallel, also for the same language
- Application of effective methods for identification of best suited technology solution for a particular language pair
- Exploring of quality enhancement measures
- Integration of the MT service into other patent tools and services

**Thank you for your attention**