# COPPA, CLIR and TAPTA: three tools to assist in overcoming the Patent language barrier at WIPO

**Bruno Pouliquen** **Christophe Mazenc**
World Intellectual Property Organization
34, chemin des Colombettes
CH-1211 Geneva 20
Bruno.Pouliquen@wipo.int Christophe.Mazenc@wipo.int

## Abstract

WIPO has built three tools (one resource and two online web interfaces) to help users and researchers to overcome the language barrier when searching patents published in different languages. WIPO has a huge corpus of translated patent applications which has been released in a new product called COPPA (Corpus Of Parallel Patent Applications). As part of its freely-available search engine PATENTSCOPE, WIPO has built a tool called CLIR (Cross Language Information Retrieval) to assist users in querying patent applications in various languages. The third tool, called TAPTA (Translation Assistant for Patent Titles and Abstracts) is a statistical Machine Translation tool aimed at helping the user to understand a patent application written in a language he does not know, TAPTA and CLIR, for their English-French data, both rely on the COPPA corpus.

## 1 Introduction

WIPO is a specialized United Nation Agency responsible for Intellectual Property and one of its activities consists in translating patent application[1] titles and abstracts into both English and French. WIPO has an extensive parallel corpus of manually translated patent documents collected over time, especially for the language pair English-French (more than 1.7 million documents). Patent applications are published on the PATENTSCOPE search engine[2], which contains various national and international collections in different languages with some texts (titles, abstracts, descriptions and claims) not translated, therefore WIPO is investigating techniques for overcoming the language barrier: with cross-language retrieval and machine translation.

Automatic translation of patents is catching international attention as a means to overcome the language barrier, for example: the Patent Machine Translation Task at NTCIR-9[3], the European project Pluto (Tinsley et al. 2010) and the collaboration between the European Patent Office and Google translate (Täger 2011), etc. One well-known approach to machine translation is Statistical machine translation which “learns” a translation model from parallel texts (Koehn 2010).

In order to boost research in this field, we recently decided to release the PCT English-French corpus in an easy-to-use format TMX in a product called COPPA (Version 1.0 released in July 2011), described in detail in section 2.

The fact that WIPO has access to parallel data (e.g. a patent application title available in English, French and German) conducted us to build a product (called CLIR) to assist users when searching for terms in foreign languages. This product is described in section 3.

[1] Also called PCT application, see WIPO (2010)

[2] http://www.wipo.int/patentscope/search

[3] This task proposes to participant to train patent machine translation tools on the same parallel corpus in English, Japanese and Chinese, and then compare the various techniques and results, see http://ntcir.nii.ac.jp/PatentMT

The COPPA corpus has been fed into an open-source-based statistical machine translation tool (called TAPTA: Translation Assistant for Patent Titles and Abstracts). It can translate texts from English to French (and Chinese) and vice-versa. This product is described in section 4.

# 2 COPPA: Corpus Of Parallel Patent Applications

Following requests from the academic world, WIPO has released a new product called "WIPO-Corpus Of Parallel Patent Applications". This new product initially includes a bilingual English-French corpus of more than 8 million parallel segments (translation units) in a format that is easy to use in building machine translation systems.

The segments were obtained by aligning the sentences of the abstracts and titles of the Patent Co-operation Treaty (PCT) applications with their translations (for the applications published between 1990 and 2010 inclusive), the translations having been performed by professional patent translators. It is therefore a gold mine for linguistic research such as terminology extraction, translation memory building and research on Machine Translation.

With the goal of supporting open innovation, WIPO offers the product free of charge to academic and private research institutions for research purposes only; in return those institutions commit to share their published results with WIPO.

WIPO hopes that the wide availability of this sizeable corpus will actively contribute to progress in building more accurate machine translation systems for patent texts with the ultimate goal of lowering the linguistic barrier for inventors and the general public and of improving the efficiency and the accessibility of the international patent system.

## 2.1 Statistics

The corpus contains more than 180 Million words, for comparison (only for English-French), the UN-corpus (Rafalovitch & Dale 2009) contains about 3 Million words, the JRC-acquis (Steinberger et al. 2006) is about 35Million words, Europarl (Koehn 2005) is about 50 Million and MultiUN (Eisele & Shen 2006) is 370 Million. See appendix A for detailed statistics about the WIPO corpus.

## 2.2 Technical details

The widely used TMX format[4] was chosen and each document contains title and abstracts available in both languages.

The first export format contains one “translation unit” for the title and one for the abstract with no further text processing. However, it must be noted that the abstracts usually contain more than one sentence, therefore we applied home-made tokenization, segmentation and alignment to build fine-grained translation units, the result of this processing is available in a second export format. This format may be more suitable for statistical machine translation as the translation units are shorter than 80 words.

Each document contains, in addition, the IPC classification, which can be used to train “domain-aware” tools (as with CLIR and TAPTA, see sections 3 and 4).

## 2.3 Availability

The corpus is available for free for research purposes and for a reasonable fee for other purposes, order form and details are available at: http://www.wipo.int/patentscope/en/data/products.html#coppa

# 3 CLIR: Cross-lingual Information Retrieval in Patentscope search engine

CLIR is aimed at improving patent application searches. With CLIR, a user can search patent applications by entering search term(s) in any language (see below) in the search box. The system will suggest variants and translations of the term(s). Variants and translations are domain-dependant (using IPC classification and an automatic domain detection tool[5]). It allows the user to search patent documents which were disclosed in a foreign language. CLIR improves recall without sacrificing precision (it provides not only translations of terms but also pertinent synonyms from the patent domain).

[4] http://www.lisa.org/tmx/

[5] We built 32 domains which are high level categories of patents (Transport, Medicine, Energy, Foods, Chemistry, etc.). Each IPC classification belongs to one or more domain(s).

CLIR is a component of PATENTSCOPE, it builds an enriched query (a Boolean query containing all term variations in different languages and the IPC classification corresponding to the domain).

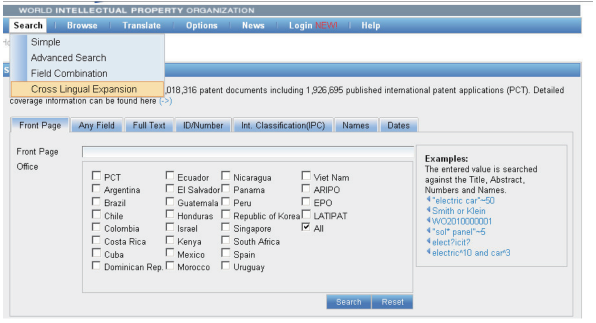
### 3.1 Background

Historically, CLIR was the first multi-lingual product produced by WIPO. We compiled a huge list of titles available in two languages and used our own method to extract a bilingual terminology without any further linguistic input (we only provide lists of "non-significant" words like 'the', 'a', etc…).

The tool runs Mgiza++ and "learns" terms that are often translations of each other in different domains. It then combines the various bilingual terms to build a multilingual domain-aware terminology knowledge base, currently available in English, French, German, Japanese, Spanish, Chinese, and Korean (with some terms in Portuguese and Russian too).

### 3.2 Availability

This tool is available, free of use, as part of the general PATENTSCOPE search engine at http://www.wipo.int/patentscope/search/clir/clir.jsp



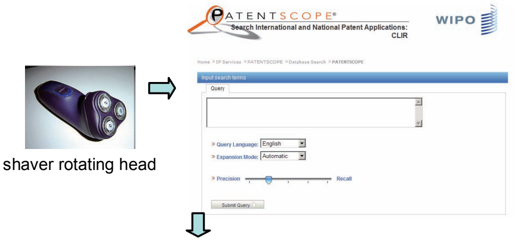
,018,316 patent documents including 1,926,695 published international patent applications (PCT). Detailed coverage information can be found here (->)

Examples:
The entered value is searched against the Title, Abstract, Numbers and Names.
"electric car"~50
Smith or Klein
WO2010000001
"sol* panel"~5
elect?icit?
electric^10 and car^3

Figure 1: CLIR access on Patentscope

### 3.3 Supervised/unsupervised mode

A first Graphical User Interface, the so-called "unsupervised mode", offers the use the possibility to search for terms without any further input[6] as shown on figure 2.



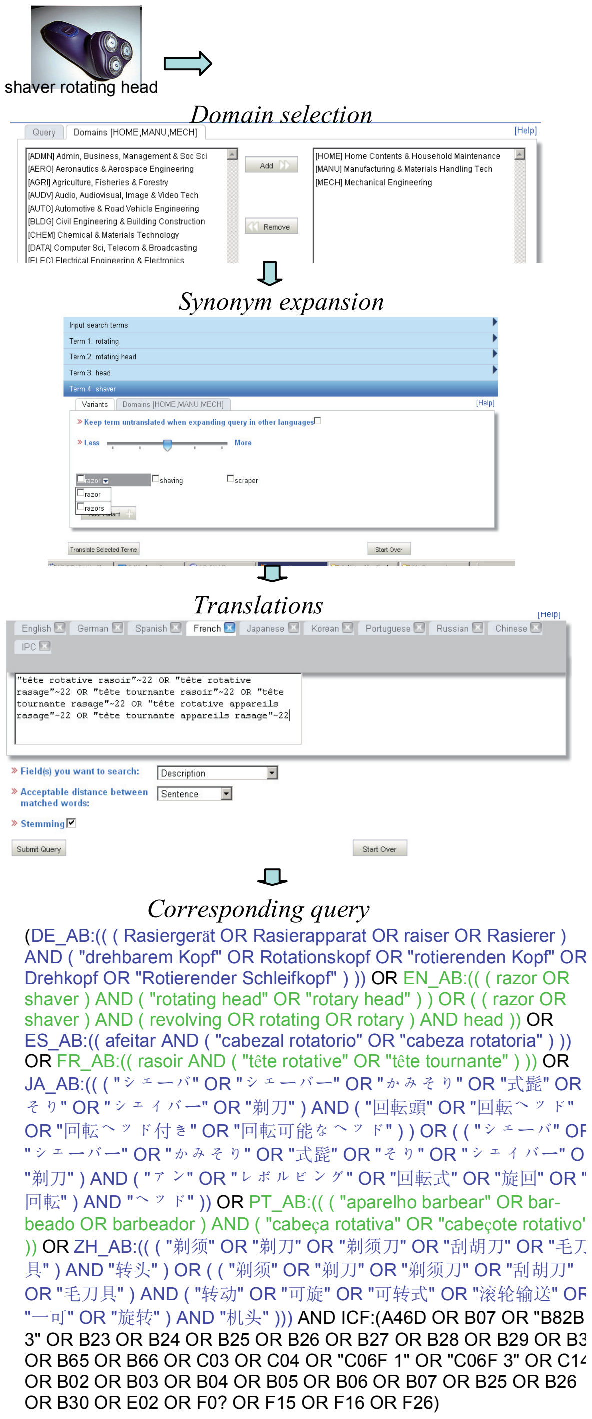
(DE_AB:(( ( Rasiergerät OR Rasierapparat OR raiser OR Rasierer ) AND ( "drehbarem Kopf" OR Rotationskopf OR "rotierenden Kopf" OR Drehkopf OR "Rotierender Schleifkopf" ) )) OR EN_AB:(( ( razor OR shaver ) AND ( "rotating head" OR "rotary head" ) ) OR ( ( razor OR shaver ) AND ( revolving OR rotating OR rotary ) AND head )) OR ES_AB:(( afeitar AND ( "cabezal rotatorio" OR "cabeza rotatoria" ) )) OR FR_AB:(( rasoir AND ( "tête rotative" OR "tête tournante" ) )) OR JA_AB:(( ( "シェーバ" OR "シェーバー" OR "かみそり" OR "式髭" OR "そり" OR "シェイバー" OR "剃刀" ) AND ( "回転頭" OR "回転ヘッド" OR "回転ヘッド付き" OR "回転可能なヘッド" ) ) OR ( ( "シェーバ" OR "シェーバー" OR "かみそり" OR "式髭" OR "そり" OR "シェイバー" OR "剃刀" ) AND ( "アン" OR "レボルビング" OR "回転式" OR "旋回" OR "回転" ) AND "ヘッド" )) OR PT_AB:(( ( "aparelho barbear" OR barbeado OR barbeador ) AND ( "cabeça rotativa" OR "cabeçote rotativo" ) )) OR ZH_AB:(( ( "剃须" OR "剃刀" OR "剃须刀" OR "刮胡刀" OR "毛刀具" ) AND "转头" ) OR ( ( "剃须" OR "剃刀" OR "剃须刀" OR "刮胡刀" OR "毛刀具" ) AND ( "转动" OR "可旋" OR "可转式" OR "滚轮输送" OR "一可" OR "旋转" ) AND "机头" ))) AND ICF:(A46D OR B07 OR "B82B 3" OR B23 OR B24 OR B25 OR B26 OR B27 OR B28 OR B29 OR B3? OR B65 OR B66 OR C03 OR C04 OR "C06F 1" OR "C06F 3" OR C14 OR B02 OR B03 OR B04 OR B05 OR B06 OR B07 OR B25 OR B26 OR B30 OR E02 OR F0? OR F15 OR F16 OR F26)

Figure 2: CLIR, unsupervised mode

The second mode, the so-called "supervised mode", allows the user to further disambiguate the query terms by entering the domains and providing additional synonyms. Each selected domain will help the system to disambiguate between possible translations: e.g. "*automatic translation*" is translated in French as "*traduction automatique*" in computer science, while it could be "*translation automatique*" in mechanical engineering.

Each selected synonym also helps the system to disambiguate, the resulting query, after user supervision, is therefore more precise.

[6] Except a sliding bar to balance between precision and recall



Query | Domains [HOME,MANU,MECH] [Help]

[ADMN] Admin, Business, Management & Soc Sci
[AERO] Aeronautics & Aerospace Engineering
[AGRI] Agriculture, Fisheries & Forestry
[AUDV] Audio, Audiovisual, Image & Video Tech
[AUTO] Automotive & Road Vehicle Engineering
[BLDG] Civil Engineering & Building Construction
[CHEM] Chemical & Materials Technology
[DATA] Computer Sci, Telecom & Broadcasting
[ELEC] Electrical Engineering & Electronics

Add | Remove

[HOME] Home Contents & Household Maintenance
[MANU] Manufacturing & Materials Handling Tech
[MECH] Mechanical Engineering

Input search terms
Term 1: rotating
Term 2: rotating head
Term 3: head
Term 4: shaver

Variants | Domains [HOME,MANU,MECH] [Help]

» Keep term untranslated when expanding query in other languages □

» Less — More

□ razor □ shaving □ scraper
□ razor
□ razors

Translate Selected Terms | Start Over

English | German | Spanish | French | Japanese | Korean | Portuguese | Russian | Chinese | IPC [Help]

```
"tête rotative rasoir"~22 OR "tête rotative
rasage"~22 OR "tête tournante rasoir"~22 OR "tête
tournante rasage"~22 OR "tête rotative appareils
rasage"~22 OR "tête tournante appareils rasage"~22
```

» Field(s) you want to search: Description

» Acceptable distance between matched words: Sentence

» Stemming ☑

Submit Query | Start Over

```
(DE_AB:(( ( Rasiergerät OR Rasierapparat OR raiser OR Rasierer )
AND ( "drehbarem Kopf" OR Rotationskopf OR "rotierenden Kopf" OR
Drehkopf OR "Rotierender Schleifkopf" ) )) OR EN_AB:(( ( razor OR
shaver ) AND ( "rotating head" OR "rotary head" ) ) OR ( ( razor OR
shaver ) AND ( revolving OR rotating OR rotary ) AND head )) OR
ES_AB:(( afeitar AND ( "cabezal rotatorio" OR "cabeza rotatoria" ) ))
OR FR_AB:(( rasoir AND ( "tête rotative" OR "tête tournante" ) )) OR
JA_AB:(( ( "シェーバ" OR "シェーバー" OR "かみそり" OR "式髭" OR
そり" OR "シェイバー" OR "剃刀" ) AND ( "回転頭" OR "回転ヘッド"
OR "回転ヘッド付き" OR "回転可能なヘッド" ) ) OR ( ( "シェーバ" OR
"シェーバー" OR "かみそり" OR "式髭" OR "そり" OR "シェイバー" O
"剃刀" ) AND ( "アン" OR "レボルビング" OR "回転式" OR "旋回" OR "
回転" ) AND "ヘッド" )) OR PT_AB:(( ( "aparelho barbear" OR bar-
beado OR barbeador ) AND ( "cabeça rotativa" OR "cabeçote rotativo'
)) OR ZH_AB:(( ( "剃须" OR "剃刀" OR "剃须刀" OR "刮胡刀" OR "毛刀
具" ) AND "转头" ) OR ( ( "剃须" OR "剃刀" OR "剃须刀" OR "刮胡刀"
OR "毛刀具" ) AND ( "转动" OR "可旋" OR "可转式" OR "滚轮输送" OR
"一可" OR "旋转" ) AND "机头" ))) AND ICF:(A46D OR B07 OR "B82B
3" OR B23 OR B24 OR B25 OR B26 OR B27 OR B28 OR B29 OR B3
OR B65 OR B66 OR C03 OR C04 OR "C06F 1" OR "C06F 3" OR C14
OR B02 OR B03 OR B04 OR B05 OR B06 OR B07 OR B25 OR B26
OR B30 OR E02 OR F0? OR F15 OR F16 OR F26)
```

Figure 3: CLIR, supervised mode

## 4 TAPTA: Translation Assistant for Patent Titles and Abstracts

### 4.1 Background

Using the COPPA corpus, a translation model was built with open source Moses (Koehn et al. 2007). Each patent application is classified into one or more IPC class(es), we used this information to simplify the classification in one of our 32 domains.

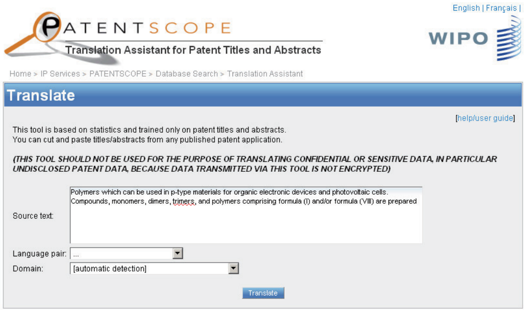
An interactive graphical user interface (using Java Swing) was created that allows users to drive the translation on the fly (selecting the best segments to translate and choosing the right proposal). A significant experiment was conducted with human operators and the tool has been used to help non professional translators to translate patent applications. The output was judged to be successful. All details have been published in the paper "TAPTA: a user-driven translation system for patent documents based on domain-aware statistical machine translation" (Pouliquen et al. 2011).

### 4.2 Availability

A cut-down TAPTA tool is available on the Patentscope website, free of use, at the following address: http://www.wipo.int/patentscope/translate/

### 4.3 Usage

The TAPTA web interface is available via the "Translate" menu in Patentscope.


English | Français |
PATENTSCOPE
Translation Assistant for Patent Titles and Abstracts
WIPO
Home > IP Services > PATENTSCOPE > Database Search > Translation Assistant
Translate
[help/user guide]
Translate

This tool is based on statistics and trained only on patent titles and abstracts.
You can cut and paste titles/abstracts from any published patent application.

*(THIS TOOL SHOULD NOT BE USED FOR THE PURPOSE OF TRANSLATING CONFIDENTIAL OR SENSITIVE DATA, IN PARTICULAR UNDISCLOSED PATENT DATA, BECAUSE DATA TRANSMITTED VIA THIS TOOL IS NOT ENCRYPTED)*

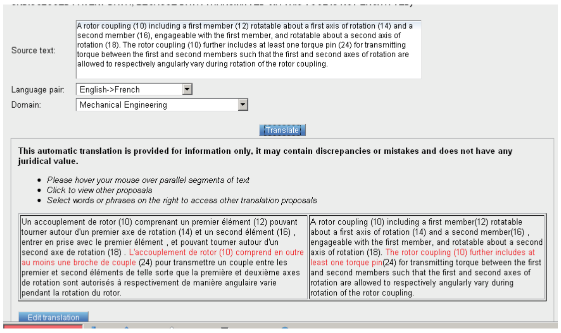
Source text: Polymers which can be used in p-type materials for organic electronic devices and photovoltaic cells. Compounds, monomers, dimers, trimers, and polymers comprising formula (I) and/or formula (VIII) are prepared

Language pair: ...

Domain: [automatic detection]

Figure 4: TAPTA-web, an overview

User can translate any title and abstract from English to French or from French to English (trained on COPPA corpus), additionally English-Chinese (both directions) is also provided.



Source text: A rotor coupling (10) including a first member (12) rotatable about a first axis of rotation (14) and a second member (16), engageable with the first member, and rotatable about a second axis of rotation (18). The rotor coupling (10) further includes at least one torque pin (24) for transmitting torque between the first and second members such that the first and second axes of rotation are allowed to respectively angularly vary during rotation of the rotor coupling.

Language pair: English->French

Domain: Mechanical Engineering

**This automatic translation is provided for information only, it may contain discrepancies or mistakes and does not have any juridical value.**

- *Please hover your mouse over parallel segments of text*
- *Click to view other proposals*
- *Select words or phrases on the right to access other translation proposals*

Un accouplement de rotor (10) comprenant un premier élément (12) pouvant tourner autour d'un premier axe de rotation (14) et un second élément (16) , entrer en prise avec le premier élément , et pouvant tourner autour d'un second axe de rotation (18) . L'accouplement de rotor (10) comprend en outre au moins une broche de couple (24) pour transmettre un couple entre les premier et second éléments de telle sorte que la première et deuxième axes de rotation sont autorisés à respectivement de manière angulaire varie pendant la rotation du rotor.

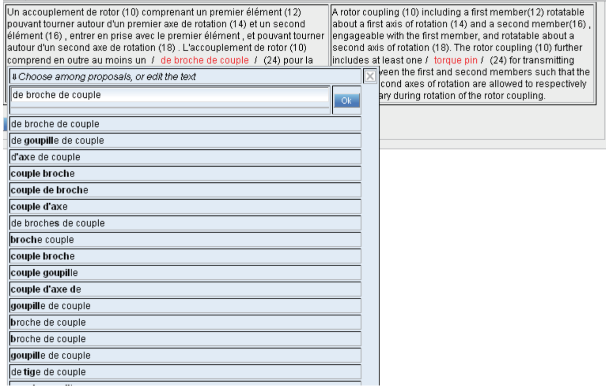
A rotor coupling (10) including a first member(12) rotatable about a first axis of rotation (14) and a second member(16) , engageable with the first member, and rotatable about a second axis of rotation (18). The rotor coupling (10) further includes at least one torque pin(24) for transmitting torque between the first and second members such that the first and second axes of rotation are allowed to respectively angularly vary during rotation of the rotor coupling.

Figure 5: TAPTA translation result page

The text is segmented and sentences and corresponding translation are highlighted in a two-column format. Clicking on a sentence will show more proposals.

When the user wants more proposals for a particular phrase, he can select the source phrase, the phrase will be segmented from the rest of the sentence and more proposals will be displayed.

E.g., in figure 5, if the user wants to look for more proposals for "torque pin", he first selects the phrase, waits a second and gets the following display:



Un accouplement de rotor (10) comprenant un premier élément (12) pouvant tourner autour d'un premier axe de rotation (14) et un second élément (16) , entrer en prise avec le premier élément , et pouvant tourner autour d'un second axe de rotation (18) . L'accouplement de rotor (10) comprend en outre au moins un / de broche de couple / (24) pour la

A rotor coupling (10) including a first member(12) rotatable about a first axis of rotation (14) and a second member(16) , engageable with the first member, and rotatable about a second axis of rotation (18). The rotor coupling (10) further includes at least one / torque pin / (24) for transmitting ween the first and second members such that the cond axes of rotation are allowed to respectively ary during rotation of the rotor coupling.

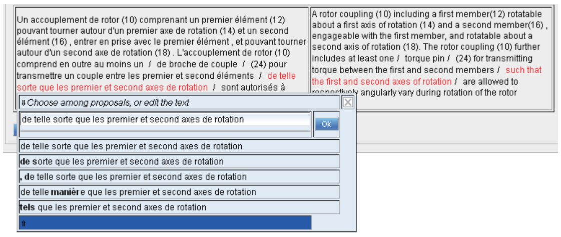
⇓Choose among proposals, or edit the text

de broche de couple

de broche de couple
de goupille de couple
d'axe de couple
couple broche
couple de broche
couple d'axe
de broches de couple
broche couple
couple broche
couple goupille
couple d'axe de
goupille de couple
broche de couple
broche de couple
goupille de couple
de tige de couple

Figure 6: TAPTA, getting more proposals for a segment

This functionality can also be used to further segment long sentences. In the example, the last sentence is too long for the tool to display good proposals, therefore we can select a phrase in the middle of the sentence to split the sentence into three sub-segments (e.g. by selecting "such that the first and second axes of rotation"):



Un accouplement de rotor (10) comprenant un premier élément (12) pouvant tourner autour d'un premier axe de rotation (14) et un second élément (16) , entrer en prise avec le premier élément , et pouvant tourner autour d'un second axe de rotation (18) . L'accouplement de rotor (10) comprend en outre au moins un / de broche de couple / (24) pour transmettre un couple entre les premier et second éléments / de telle sorte que les premier et second axes de rotation / sont autorisés à

A rotor coupling (10) including a first member(12) rotatable about a first axis of rotation (14) and a second member(16) , engageable with the first member, and rotatable about a second axis of rotation (18). The rotor coupling (10) further includes at least one / torque pin / (24) for transmitting torque between the first and second members / such that the first and second axes of rotation / are allowed to respectively angularly vary during rotation of the rotor

⇓Choose among proposals, or edit the text

de telle sorte que les premier et second axes de rotation

de telle sorte que les premier et second axes de rotation
de sorte que les premier et second axes de rotation
, de telle sorte que les premier et second axes de rotation
de telle manière que les premier et second axes de rotation
tels que les premier et second axes de rotation

Figure 7: TAPTA, example of user's segmentation

We also trained the system for the English-Chinese pair. This Web interface is currently used daily (in August 2011, we have about 300 translation requests every day, more than half of the requests come from China, Chinese to English being the most used direction).

## Conclusion and future work

One of the mandates of WIPO is to facilitate access to technical knowledge and information. To achieve this goal, not only does WIPO give access to patent applications in various language through its PATENTSCOPE search engine, but it also provides a way to search across languages (CLIR), to assist users in understanding foreign language text (TAPTA), and also encourages open innovation by providing its corpus of translated patent application (COPPA) free of charge for research purposes.

We would like to highlight here the potential of the "data-driven" approach: using only huge amount of data, were we able to build cross lingual search and machine translation tools that can ease the work of the user when he has to tackle languages he does not know (our TAPTA tool has been adapted to Chinese-to-English translation which is currently more used than the English-French direction).

Future work includes the improvement of existing tools in terms of language coverage (with a special focus on Chinese, Japanese and Korean) and in terms of functionalities (e.g. train translations with claims and/or descriptions).

## Acknowledgments

Special thanks to our colleague Paul Halfpenny for his valuable proof-reading.

## References

Eisele, Andreas & Yu Chen: MultiUN: a multilingual corpus from United Nation documents. LREC 2010: proceedings of the seventh international conference on Language Resources and Evaluation, 17-23 May 2010, Valletta, Malta; pp.2868-2872.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris C. Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, Evan Herbst. (2007). Moses: open source toolkit for statistical machine translation. In Proceedings of ACL 07. Morristown, NJ, USA, 177-180.

Koehn, Phillip. (2010) Statistical Machine Translation. textbook, Cambridge University Press, January 2010.

Koehn, Phillip: Europarl: a parallel corpus for statistical machine translation. MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: the tenth Machine Translation Summit; pp.79-86. [Philipp Koehn: Europarl: a parallel corpus for statistical machine translation. MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: the tenth Machine Translation Summit; pp.79-86.

Pouliquen, Bruno, Christophe Mazenc & Aldo Iorio: Tapta: a user-driven translation system for patent documents based on domain-aware statistical machine translation. [EAMT 2011]: proceedings of the 15th conference of the European Association for Machine Translation, 30-31 May 2011, Leuven, Belgium; eds. Mikel L.Forcada, Heidi Depraetere, Vincent Vandeghinste; pp.5-12.

Rafalovitch, Alexandre & Robert Dale: United Nations general assembly resolutions: a six-language parallel corpus. MT Summit XII: proceedings of the twelfth Machine Translation Summit, August 26-30, 2009, Ottawa, Ontario, Canada; pp.292-299.

Steinberger, Ralf, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaž Erjavec, Dan Tufiş, & Daniel Varga: The JRC-Acquis: a multilingual aligned parallel corpus with 20+ languages. LREC-2006: Fifth International Conference on Language Resources and Evaluation. Proceedings, Genoa, Italy, 22-28 May 2006; pp.2142-2147

Täger, Wolfgang, (2011): The sentence-aligned European patent corpus. *[EAMT 2011]: proceedings of the 15th conference of the European Association for Machine Translation*, 30-31 May 2011, Leuven, Belgium; eds. Mikel L.Forcada, Heidi Depraetere, Vincent Vandeghinste; pp.177-184.

Tinsley, John, Andy Way and Páraic Sheridan, (2010) PLuTO: MT for online patent translation. *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas*, Denver, Colorado, October 31 – November 4, 2010; 8pp

WIPO. (2010). PCT The International Patent System - Yearly review, developments and performance in 2009, WIPO Publication No. 901(E)/09, June 2010

## Appendice A: statistics on the WIPO-COPPA V1.0 corpus

| Year | Size(bytes) | Size(compressed) | Nº translation units | Nº documents | Nº characters | Nº words |
|---|---|---|---|---|---|---|
| 1990 | 46473037 | 9095220 | 81046 | 16055 | 10573333 | 1722269 |
| 1991 | 57468664 | 11207227 | 100300 | 20085 | 12981914 | 2109601 |
| 1992 | 65418799 | 12694222 | 114026 | 22847 | 14783102 | 2398012 |
| 1993 | 74296322 | 14401080 | 129432 | 25968 | 16845464 | 2724730 |
| 1994 | 84817778 | 16375984 | 148049 | 29872 | 19244159 | 3111219 |
| 1995 | 100639428 | 19584649 | 174485 | 35478 | 23027411 | 3720341 |
| 1996 | 120323625 | 23239275 | 208084 | 42012 | 27516077 | 4433081 |
| 1997 | 142763728 | 27413781 | 246961 | 50021 | 32611383 | 5262382 |
| 1998 | 169492474 | 32290339 | 291985 | 59186 | 38860504 | 6247812 |
| 1999 | 192548288 | 36463180 | 330906 | 67545 | 44384533 | 7119477 |
| 2000 | 223035175 | 41953713 | 383722 | 79146 | 51606674 | 8273773 |
| 2001 | 274677688 | 50899788 | 472875 | 97814 | 63446267 | 10131499 |
| 2002 | 282429325 | 52250045 | 488418 | 102495 | 64822714 | 10354019 |
| 2003 | 291862364 | 53725017 | 506822 | 106349 | 66612714 | 10647279 |
| 2004 | 312427052 | 57178666 | 541576 | 113456 | 71487472 | 11458800 |
| 2005 | 347754061 | 63049665 | 604481 | 124079 | 79365357 | 12720470 |
| 2006 | 392615206 | 71210918 | 685938 | 137533 | 89200185 | 14326117 |
| 2007 | 426685344 | 76461897 | 749514 | 148770 | 96389566 | 15481121 |
| 2008 | 458223761 | 80905756 | 811288 | 156985 | 102476871 | 16467721 |
| 2009 | 470180933 | 81704247 | 838687 | 158076 | 104190300 | 16760333 |
| 2010 | 451878162 | 77132105 | 808354 | 151345 | 99855142 | 16043201 |
| **TOTAL** | **4986011214** | **909236774** | **8716949** | **1745117** | **1130281142** | **181513257** |