



## **Abu-MaTran: automatic building of machine translation**

**Marie Curie IAPP project FP7-PEOPLE-2012-IAPP**

**24-month recruitment of a postdoctoral researcher**

### **Overview**

Prompsit<sup>1</sup> is a research-shaped company created in 2006 inside the Transducens<sup>2</sup> research group at the Department of Software and Computing Systems<sup>3</sup> (Universitat d'Alacant - Spain). It's a leading company in the development of machine translation, specially linguistically-motivated systems such as Apertium<sup>4</sup> rule-based systems or linguistically-augmented Moses<sup>5</sup> statistical systems. The company activity in R&D is intense both as an industry-driven activity or by participation in public national and international R&D programs.

Abu-MaTran<sup>6</sup> (*Automatic Building of Machine Translation*) is an IAPP-FP7<sup>7</sup> project in which the company is currently involved. The project aims at increasing the hitherto low industrial adoption of machine translation by identifying crucial cutting-edge research techniques (automatic acquisition of corpora and linguistic resources, pivot techniques, linguistically augmented statistical translation and diagnostic evaluation) and preparing them to be suitable for commercial exploitation.

Besides Prompsit as a central node of interaction, the project involves four top research institutions (Dublin City University<sup>8</sup> - project coordinator, Universitat d'Alacant<sup>9</sup>, University of Zagreb<sup>10</sup> and Institute for Language and Speech Processing<sup>11</sup>). At Prompsit, the project will be led by researcher Sergio Ortiz Rojas, responsible for most of the code of the Apertium MT platform, Prompsit's linguistically-augmented Moses system, a modular version of the Bitextor<sup>12</sup> parallel text collector and other natural language processing tools for information extraction or opinion analysis.

The position involves research, development and participation in outreach activities to achieve the goals of the Abu-MaTran project as well as collaboration with all researchers in the project.

---

1 <http://www.prompsit.com>

2 <http://transducens.dlsi.ua.es/>

3 <http://www.dlsi.ua.es/>

4 <http://apertium.org>

5 <http://www.statmt.org/moses/>

6 <http://www.abumatran.eu/>, more info at [http://cordis.europa.eu/projects/rcn/106334\\_en.html](http://cordis.europa.eu/projects/rcn/106334_en.html)

7 [http://ec.europa.eu/research/mariecurieactions/about-mca/actions/iapp/index\\_en.htm](http://ec.europa.eu/research/mariecurieactions/about-mca/actions/iapp/index_en.htm)

8 <http://www.dcu.ie>

9 <http://www.ua.es>

10 <http://www.unizg.hr/homepage/>

11 <http://www.ilsp.gr/>

12 <http://bitextor.sourceforge.net/> (see *prompsit* branch in source code)

# Job Description

## **Main Duties and Responsibilities**

- Investigate, in collaboration with the partners, better techniques to automate:
  - monolingual and bilingual general and domain-focused corpora acquisition
  - monolingual and bilingual terminology extraction
  - automatic induction of transfer rules
  - building of pivot or linguistically-augmented machine translation systems
  - machine translation automatic evaluation
- Implement the techniques for each of the previous points
- Carry out experiments to evaluate their performance.
- Release the output as free/open-source tools with appropriate interfaces to use them.
- Write the appropriate documentation for each of the work lines: technical documentation for developers, academic-oriented (papers, posters, etc.) publications, and tutorials or manuals for developers.
- Attend project-related conferences and meetings
- Present the results at relevant conferences and scientific meetings
- Review work plan with the collaborators according to project intermediate milestones and results.
- Get involved and give support to outreach activities (linguistic olympiads<sup>13</sup>, FreeRBMT workshop<sup>14</sup>)

## **Person Specification**

Applicants should provide evidence in their applications that they meet the following criteria.

The staff in charge of this recruitment process, will use a range of selection methods to measure candidates' abilities in these areas including reviewing your application, seeking references, inviting shortlisted candidates to be interviewed, and other forms of assessment action relevant to the post.

## **Criteria**

**Qualifications (compulsory):** PhD in Computer Science (or at least 4 years of full-time research experience) and less than 10 years of full-time research experience.

**International procedure (compulsory):** the candidate cannot have worked or lived for more than 12 months within the last 3 years in Spain.

### **Experience in:**

- Natural language processing, particularly in machine translation (compulsory).
- User-level or developer-level experience in Apertium, Moses, OpenMaTrEx<sup>15</sup>, Bitextor, ILSP Focused Crawler (ILSP-FC)<sup>16</sup>, ccLexExtractor<sup>17</sup> and DELiC4MT<sup>18</sup> (desirable)
- Data acquisition (desirable)
- Terminology extraction (desirable)
- Machine learning (desirable)
- MT evaluation (desirable)
- Creation of user interfaces and software releasing/sharing (desirable)

---

13 [http://en.wikipedia.org/wiki/International\\_Linguistics\\_Olympiad](http://en.wikipedia.org/wiki/International_Linguistics_Olympiad)

14 <http://www.chalmers.se/hosted/freerbmt12-en>

15 <http://www.openmatrex.org/>

16 <http://nlp.ilsp.gr/redmine/projects/ilsp-fc/>

17 <http://www.nljubesic.net/resources/tools/cclexex/>

18 <http://www.computing.dcu.ie/~atoral/delic4mt/>

**Programming languages:** C++, Python, PHP (compulsory). JAVA (desirable).

**Multilingual skills:** Good level of English (compulsory). Basic knowledge of Spanish or Catalan (desirable). Knowledge of the South Slavic Languages targeted in the project use case -- Croatian, Bosnian, Serbian or Montenegrin and Slovenian (desirable).

**Good writing and communication skills:** ability to intercommunicate with people and to communicate results, ideas, etc. (compulsory).

**Collaborative working skills:** ability to take and delegate responsibilities (compulsory).

**Experience in free/open-source software development:** participation in free/open-source software development projects as user or, better, as contributor (desirable).

**Experience in transfer of knowledge between the industry and the academy:** interaction between industry and academy in previous positions is highly valued (desirable).

**Creativity and flexibility skills:** ability to be open to different ideas or opinions, to analyse and solve problems and to make decisions (desirable).

**Active research skills:** ability to follow state-of-the-art research lines associated with the project, to learn and acquire new skills relevant to the project, to write scientific works, and to meet deadlines (desirable).

## Further Information

This post is fixed-term and full-time at Prompsit (Elx/Elche, Spain). The starting date is January 2014 and duration is 24 months. Splits are not possible.

### ***Terms and conditions of employment:***

Terms and conditions will be according to the stipulations of the IAPP program. The recruited candidate will have a full-time contract with full social security coverage subject to the Spanish laws and taxes.

### ***Salary:***

For an Experience Researcher with less than 10 year experience the stipulated salary will be a €57,154 per year gross salary corresponding to living allowance and additionally €683/€977 per month for mobility allowance (depending on family charges).

### ***Closing date:***

1st June 2013.

### ***Informal enquiries:***

For informal enquiries about this job contact us at [info@prompsit.com](mailto:info@prompsit.com).