

Postdoctoral Research Position

The Laboratory ATILF offers a 12 month postdoctoral research position in Computational Linguistics under an ANR financing scheme (MSH Lorraine—Inalco—CNRS ATILF).

The researcher will work as part of the team of the ALIENTO research project (Linguistic and intercultural analysis of short sapiential statements and of their transmission East/West, West/East)

The contract begins on October 1st 2015.

Description of the research project and main focus of the postdoctoral researcher

Aliento is an interdisciplinary innovatory research project in the recent field of e-humanities. It integrates various types of scientific experts in texts, linguistics, paremiology and computer science. This dialogue of a common object (the circulation of brief medieval sapiential statements) opens up new methodological and conceptual research perspectives. Thus ALIENTO plays a fundamental role in the scientific treatment, both methodological and conceptual, of sapiential texts thanks to the nature of its corpora and the way in which it treats them, the multilingual sources, and the annotation system as well as for the originality of our qualitative approach to corpus exploitation (man-computer interoperability).

• **Aims of the research project**

The aim of **the Aliento project** is the creation of a research tool (a piece of software transferable to other similar texts) to be used with an important reference corpus made up of eight comparable texts having circulated in the Iberian Peninsula in the Middle Ages (in Latin, Arabic, Hebrew, Spanish and Catalan), that is 582 pages of texts for a total number of 9,570 sapiential units. This study will allow the extraction of the key elements in the annotated corpora in relation with the sources, the transmission, the circulation and the posterity of the sapiential statements in the Iberian Peninsula from the ninth to the twelfth century between the Three Cultures.

We aim to provide the key (not yet available for lack of ad-hoc software tools) to the treatment of comparable multilingual corpora, which are, to a great extent, not aligned. By the collective creation of a piece of software capable of automatically matching the sapiential multilingual data (including Arabic and Hebrew), we intend to offer medievalists the possibility of accessing one another's knowledge beyond the language barrier and their specific disciplines.

The phases of the work are the following:

- the annotation of texts by using an XML/TEI tagging system on the sense, the form, the structure and the moral of each brief sapiential statement found in order to match and cross the sapiential statements and develop, in close collaboration with the computational researchers, a robust encoding system protocol for this type of items which can be transferable to other corpora (formalisation, normalisation, creation of a common theory applicable to this type of statements)
- the creation of a software tool capable of performing the matching of the sapiential units, i.e. a trilingual interface that will allow the production and observation of the results, namely the links between the brief sapiential statements.
- The analysis of the matching results.

- The communication of the scientific results in the field of humanities that will allow in its turn the study of the circulation of brief sapiential statements, their transformations and their degree of acculturation based on their own degree of intertextuality.

The current research project ANR ALIENTO is the result of four consecutive years of research.

For further information: www.aliento.eu

Focus of the postdoctoral researcher:

The postdoctoral researcher must have training in Computational Linguistics. His/her missions will be to propose and develop a methodology using all the researchers' added annotations (meaning, form, structure, multilingual translation) in order to calculate the sapiential statements matching. The objective is therefore to develop a tool capable of detecting automatically the similarities between statements to be linked together in order to facilitate the subsequent access. The work on the multilingual aspects will be a real gain on the expected results.

S/he will work in Nancy (Laboratoire ATILF, CNRS/Université de Lorraine) in close collaboration with the research group ALIENTO (Paris-Nancy).

• Background

- Doctorate in Computational Linguistics.
- Good knowledge of encoding languages (XML).
- Good knowledge of probabilistic approaches in Linguistics

• Terms of contract

Length: 12 months

The postdoctoral researcher can be of any nationality. English can be the working language for researchers not speaking French.

The contract begins on October 1st 2015.

Applications must be submitted by August 31st 2015.

Gross monthly salary: 2525€ to 2913€ /month depending on experience

Please send three-page curriculum vitae, a summary of your dissertation and a letter of application to:

Étienne Petitjean : etienne.petitjean@atilf.fr

Jean-Marie Pierrel : jean-marie.pierrel@atilf.fr