

“LEXICOGRAPHIE, TERMINOGRAPHIE ET TRAITEMENT AUTOMATIQUE DES CORPUS”

(LEXICOGRAPHY, TERMINOGRAPHY AND CORPUS PROCESSING)



OBJECTIVES

The aim of the second-year professional Master's degree in Lexicography, Terminography and Corpus Processing (“Lexicographie, Terminographie et Traitement Automatique des Corpus” – “LTTAC” for short) is to provide students with high-level skills in lexicography, terminography and the language industries. The course, which has been going for more than fourteen years, is taught by many professional lexicographers, terminographers and specialists from leading companies – e.g. publishing houses producing both general and specialized dictionaries, or companies producing language-oriented software (search engines, translation memories, etc.).

JOB PROSPECTS

- **Lexicography:** jobs pertaining to all aspects of bilingual and monolingual lexicography, both traditional and computer-assisted, with private or public dictionary publishers.
- **Terminography:** jobs with official institutions, leading companies and private publishing houses carrying out projects in the field of monolingual, bilingual and multilingual terminography, both traditional and computer-assisted.
- **Language industries:** jobs related to the design and development of computer tools focusing on the lexicon (electronic dictionaries, programs for machine or computer-aided translation, information retrieval software, spell- and grammar-checkers, linguistic databases, knowledge databases, etc.).

STUDENT PROFILES

Initial training

Prerequisite: first-year Master's degree or equivalent.

Because of the diversity of the skills that are taught, the LTTAC Master's degree is open to students trained in a wide range of disciplines: French literature and language, linguistics, French as a foreign language, modern or ancient languages, philosophy, information science, computer sciences, etc., as well as to specialists in all kinds of scientific or technical disciplines desiring to learn how to compile dictionaries related to their subject matters. As a consequence, since some students are better prepared than others for the degree, decisive factors for admission include quality of application and motivation.

Further education

The LTTAC Master's degree is open to professional lexicographers or terminographers wanting to add to their knowledge of the theory of their discipline, to specialists in scientific or technical disciplines wanting to learn how to compile terminological dictionaries, and to specialists from the language industries designing and using electronic lexical resources and corpora.

Foreign students

The degree is open to all French-speaking students regardless of their nationalities. We have had many students from all over the world since the launching of the course.

SPECIAL CASES

Students resuming their education

Applicants not holding a first-year Master's degree or having interrupted their university education for over two years are required to apply for an evaluation of their knowledge and experience. This requires a special form in addition to the application file, which can be obtained from the SUVAC, short for “Service Universitaire de Validation des Acquis et des Compétences” (University Service for Prerequisites and Skills Validation).

Students with foreign higher education

Applicants not holding a French first year Master diploma are required to fill in a specific form designed to validate studies pursued abroad. The form comes in addition to the rest of the basic application file. It is available from the DAEVU, short for “Direction de l'Administration des Études et de la Vie Universitaire” (Department for Studies Management and University Life).

ADVICE TO APPLICANTS

- In their first-year Master degree's thesis, applicants should choose a topic that is related to subjects taught in the course. The quality of the thesis will be an important element in the selection process.
- Applicants should have an idea of their job prospects through analysing our programme down to details and gathering information on the jobs they are interested in.

♦ THE STATUS OF THE LTTAC TRAINING COURSE

LTTAC is managed by the Department of “Lettres modernes” (French). In the new organization of courses at Lille 3, it is part of a second-year professional specialization called “Traitement Automatique des Langues, Dictionnaires, Terminologies, Corpus” (Natural Language Processing, Dictionaries, Terminologies, Corpora), itself part of “Sciences Du Langage” (Linguistics) courses, in the “Art, Lettres, Langues et Communication” (Art, Literature, Languages and Communication) Master's degree. Ever since it was created in 1991, the course has had close links with research carried out in the field of linguistics by STL, short for “Savoirs, Textes, Langage” (Knowledge, Text, Language) - previously SILEX, short for “Syntaxe Interprétation LEXique” (Syntax, Interpretation, LEXIS) -, a CNRS UMR (i.e. a joint research unit).

♦ PROFESSORS IN CHARGE

Pierre CORBIN

pierre.corbin@univ-lille3.fr, pierre.corbin@libertysurf.fr

Nathalie GASIGLIA

nathalie.gasiglia@univ-lille3.fr, gasiglian@aol.com

♦ MAXIMUM STUDENT CAPACITY

15 to 20 students + a few professionals.

♦ SELECTION PROCEDURES

Examination of previous university achievements & interview

- Subjects broached:

Previous university achievements or professional career

Jobs prospects and motivation

Knowledge of linguistics, lexicography, terminography, computer sciences and languages

- Selection criteria:

Quality of university achievements and application form

Motivation

Envisaged career path

♦ APPLICATION FORMS

Available from the professors in charge of the course or from the French Department (“UFR de Lettres Modernes”). Also downloadable from the university website. The form should be sent to the professors in charge (see the first address below).

♦ WHERE TO SEND YOUR APPLICATION FORM

Nathalie GASIGLIA – UMR STL (bâtiment B)

Université Charles-de-Gaulle Lille 3

B.P. 60149

F-59653 VILLENEUVE D'ASCO CEDEX

♦ CLOSING DATE FOR APPLICATIONS

First recruiting session: 27/06/08

Second recruiting session: 05/09/08

♦ FOR FURTHER INFORMATION ON CONTENTS OF THE COURSE OR ADMINISTRATIVE DETAILS

Nathalie GASIGLIA – UMR STL (bâtiment B)

Université Charles-de-Gaulle Lille 3

B.P. 60149

F-59653 VILLENEUVE D'ASCO CEDEX

Phone: (33 3 + 03) 20 41 66 61 or (33 6 + 06) 11 44 20 33

e-mail: nathalie.gasiglia@univ-lille3.fr, gasiglian@aol.com

website:

<http://stl.recherche.univ-lille3.fr/siteheberges/LTTAC/index.htm>

SUVAC

Phone: (33 3 + 03) 20 41 66 47

e-mail: suvac@univ-lille3.fr

DAEVU – foreign students unit

Phone: (33 3 + 03) 20 41 66 59

e-mail: cee@univ-lille3.fr

COURSE CONTENTS

The course encompasses all the processes involved in building lexical resources which require the use of a computer intervention at every stage of the development.

S3 UE1 / The production process of dictionaries and editorial project management - 30h - 3 ECTS

- Introduction to lexicography; lexicology and lexicography; terminology and terminography.
- The economics of lexical index production: private vs. public-funded publishing.
- The different aspects of an editorial project.
- Techniques for setting up an editorial project.

S3 UE2 / Structure of dictionaries - 51h - 6 ECTS

- The different types of dictionaries.
- The structure of dictionaries: text and paratext.
- The structure of the nomenclature: selection of headwords according to various parameters (nature of the project, linguistic and extra-linguistic criteria).
- The organization of the word list: simple or complex macrostructures, sublemmatic addressing.
- Access structure and signposting systems: running titles, running initials, cross-references.
- General organization of entries.

S3 UE3 / Monolingual lexicography and terminography: writing articles - 51h - 6 ECTS

- The plurality of writing modes: writing original texts vs. adapting existing products.
- Selecting terms for a particular editorial project.
- Labelling sociolinguistic variations.
- Writing definitions: sense and reference.
- Principles for organizing senses.
- Selection and multi-purpose status of examples.
- Typographical representation of the structure of entries.

S3 UE4 / Linguistics for lexicography - 24h - 6 ECTS

- Morphology or lexical semantics course.

S3 UE5 / Corpora as linguistic and lexicographic documenting resources: principles for building and exploring such resources - 30h - 3 ECTS

- The different kinds of documenting resources (printed corpus vs. electronic corpus, written sources vs. oral sources, dictionary sources).
- Problems in building electronic corpora.
- Getting started with existing extraction tools for corpus mining (morpho-syntactic analysers, concordance software, term extraction tools).

ORGANIZATION OF THE COURSE

The course consists of 360 hours of classes, a technical project, and a work placement of at least three months. Classes begin in September and end in April.

ASSESSMENT

Assessment of S3 UE 1 to 6 and S4 UE 1 to 3 consists of oral presentations and/or reports.

The work placement (S4 UE 4) is assessed through a report written by the student's tutor (based on an evaluation grid) and a formal presentation.

Assessment of the technical project (S4 UE5) is based on the quality of the report and presentation.

The eleven teaching units (=UE) represent 60 ECTS credits.

S3 UE6 / Computer-based structuring and tagging of dictionaries - 54h - 6 ECTS

- The theory of document tagging: SGML/XML and XSL.
- Tagging dictionary entries in SGML and XML: building DTDs, compiling marked-up entries, post-processing marked-up texts for style editing, or for the generation of other documents.
- The retroconversion of printed dictionaries.

S4 UE1 / Bilingual lexicography and learner dictionaries - 39h - 3 ECTS

- Which dictionaries for foreign learners (bilingual, monolingual, bilingualised)?
- The characteristics of bilingual dictionaries : L1 and L2, dictionaries for comprehension, for expression, and intermediate solutions.
- How to compile a bilingual dictionary.
- Principles for organizing and compiling entries.
- The characteristics of monolingual learners' dictionaries: methodology, defining vocabulary.

S4 UE2 / Programming for corpus processing - 48h - 6 ECTS

- Basic algorithmics.
- Writing regular expressions, automata, transducers for corpus mining and text manipulation.
- Programming in Awk or Perl for corpus mining with applications in lexicography or terminography.

S4 UE3 / Relational databases for storing linguistic and lexicographical information - 33h - 3 ECTS

- Building relational databases for storing, organizing and documenting lexicographical or terminographical data.

S4 UE4 / Work placement in a private company or public institution - 3 months minimum - 6 ECTS

- Carrying out tasks in keeping with course contents.
- The work placement should take place in a private company or public institution that develops products in relation to lexicography, terminography, or corpus processing.

S4 UE5 / Technical project report - 9 ECTS

- Writing a report on an original project in lexicography, terminography, or corpus processing.

ECTS = European Credit Transfer System.

The ECTS credit system was designed to improve cooperation between universities by fostering mobility for students and increasing communication between institutions in the European Union.

WORK PLACEMENT

It takes place in a private company or public institution that develops products in relation to lexicography, terminography or corpus processing.

EXAMPLES OF WORK PLACEMENT PROGRAMMES

- Compiling entries for bilingual or monolingual dictionaries, printed or electronic.
- Proof-reading of dictionary material: typography, checking spelling, and computer-based encoding of cross-references.
- Creation and expansion of neologism databases.
- XML conversion of existing lexical resources.
- Participation in a corpus-building project.
- Contribution to integrating semantics in search engine systems.

PARTNERSHIPS

Abroad: Aristote University (Thessaloniki, Greece); Edigeo (Milan, Italy); Ella Associates (Oxford, U.K.); eMagin (Rabat, Morocco); HarperCollins Publishers (Glasgow, U.K.); Institut des langues vivantes (Leuven, Belgium); Institut für maschinelle Sprachverarbeitung - IMS (Stuttgart, Germany); Joint Research Centre - Language Technology, European Commission (Milan, Italy); Katholieke Universiteit Leuven Campus Kortrijk; K Dictionaries (Tel Aviv, Israel); KULAK (Courtrai, Belgium); Langenscheidt Fachverlag (Munich, Germany); Microsoft (Redmond, U.S.A.); Van Dale Publishing House (Utrecht, Netherlands)

In France : Académie Française (Paris); As An Angel (Paris); Auralog (Montigny-le Bretonneux); Clé International (Paris); DataLing [ex-Anglo-Institute] (Lannion); Décathlon (Villeneuve d'Ascq); Délégation Générale à la Langue Française et aux Langues de France - DGLGLF (Paris); Dictionnaires Le Robert (Paris); EDIT (Paris); Éditions Bonneton (Paris); Éditions Garnier (Paris); Éditions Philippe Auzou; Éditions Nathan (Paris); ExeQuo (Paris); Generalia Multimedia (Sartrouville); Hachette Éducation (Paris); Hachette Livre (Paris); IBM (Noisy-le-Grand); Institut de Recherche en Informatique de Toulouse - IRIT UMR 5505 du CNRS (Toulouse); Institut de Recherche pour le Développement - IRD (Orléans and Cayenne, French Guiana); Institut National de la Langue Française [INaLF] (Nancy); Laboratoire d'Informatique de Paris-Nord - LIPN UMR 7030 du CNRS (Villetaneuse); Laboratoire LIDILEM Grenoble III Stendhal University; Larousse (Paris); Microsoft (Paris); Redon (Marsanne puis Paris); Séditec, Groupe Aéroconseil (Toulouse); Sélection du Reader's Digest (Bagneux); Sinequa (Ivry-sur-Seine); Softissimo (Paris); Systran (Paris); Temis (Paris); Thales Communication - Kalima (Massy-Palaiseau); Union Latine (Paris)