（別紙）

# 専攻分野及び研究計画

Field of Study and Study Program

Name in full, in your native language

（姓名（自国語））　Lecailliez (Family name/Surname), Louis (Given name), (Middle name)

Name in Roman capital letters

（姓名（ローマ字））　Lecailliez (Family name/Surname), Louis (Given name), (Middle name)

Nationality

（国　籍）French

Proposed study program in Japan（Outline your field of study on this side and the specifics of your study program on the reverse side of this sheet. This section is one of the most important references for selection. The statement must be typewritten or written in block letters. Additional sheets of paper may be attached if necessary. If plagiarism or fraud is discovered after selection, the selection will be cancelled retroactively.)

（日本での研究計画；この研究計画は、選考及び大学配置の重要な参考となるので、表面に専攻分野の概要を，裏面に研究計画の詳細を具体に記入すること。記入はタイプ又は楷書によるものとし、必要な場合は別紙を追加してもよい。なお、採用後に不正、盗用等が判明した場合は遡って採用を取り消す。）

If possible, please write your response in Japanese.（相当の日本語能力を有する者は日本語により記入すること。）

1　Present field of study（現在の専攻分野）

Field of study: Computer Science > Natural Language Processing (NLP) > Electronic Dictionaries.

Natural Language Processing Master Thesis: Approaches for a Good Quality OCRing of a Vietnamese-French Dictionary Containing Nôm Characters.
Japanese Studies Master Thesis: Towards Graph Modeling of Japanese dictionaries.
Other related work: currently implementing a web-based Japanese dictionary for a group of 18 researchers.

2　Your research topic in Japan: Describe articulately the research you wish to carry out in Japan.
（渡日後の研究テーマ：日本においてどういった研究がしたいかを明確に記入すること）

**Context**

The availability of personal computing devices has made possible the creation by researchers and enterprises of Intelligent Computer-Assisted Language Learning (ICALL) software that aims to help people learning a foreign language. “Context-aware” systems try to acquire data about the user and the learning situation to present adapted educational content. However the vast majority of the research systems fail to address a crucial point: student motivation. Nearly all platforms require the learner to use a given website or application, which is not proved to be actually done in the long run i.e. longer than a few weeks or months of experimentation.

**Issue**

I plan to address the issue by investigating the architecture of a recommender system which work where the learner has already his habits: on the Web. Rather than competing with existing websites, social networks and content providers for the user full attention on the language-learning task, the system under research will make use of various analytics to detect learner’s interests and current knowledge in the studied language. It can then make relevant recommendations about vocabulary, grammar points or cultural knowledge suited to the user level and curiosity.

**Research Questions**

This raises the questions of structuring such a system and modeling student knowledge. The classical client-server model approach will be followed. All the intelligence of the system is in the server tier: it analyzes web pages, constructs and maintains a model of the learner’s knowledge and gain new insights by comparing students behavior. In addition to tailoring existing CALL architectures to the needs of this new system it requires composing various state of the art natural language processing (NLP) and machine learning (ML) technics to perform the expected operations.

3　Study program in Japan: (Describe in detail and with specifics — particularly concerning the ultimate goal(s) of your research in Japan)

（研究計画：詳細かつ具体に記入し、特に研究の最終目標について具体的に記入すること。）

**Research Impact**

The research ultimate goal is to lay the foundation of a non-intrusive Intelligent Tutoring System for language learning that can create an emotional-like relationship with the student by taking in account contextual information such as the learner's knowledge and interests. The system can then provide him language-related recommendations just beyond the edge of his current knowledge to improve its language proficiency. Fine-grained feedback on user progress can be given as well. The main difference and advantage over existing systems is the fact it doesn't require the learner to use a specific website or application, and that it acquires most contextual information automatically.

Building a minimal working prototype (needed for experiments) is in the scope of the research. Commercial applications of results are only a few years ahead after thesis publication. Implementations are expected to help learners building vocabulary and cultural knowledge by small-intensity long-running exposure to the foreign language under study. Such software would be useful for students and self-learners of countries with mild foreign language education success.

**Methodology**

A non-comprehensive literature review was already done as part of this proposal about the CALL field and its subfield Mobile Assisted Language Learning (MALL) but it needs to be completed. The first year[1] or so of research will also be devoted to reviewing the state of the art in language knowledge representation and webpage or article summarization. Common applications of machine learning that are relevant to the research project such as clustering, recommendation algorithms and Educational Data Mining will be studied during this period.

The second research step will be focused on drafting the architecture of the whole system and making an informed decision on how the Student Model must works. The choice will be put under experimentation with a software implementation, as this is one of the most critical components for effectiveness of the whole system. Multiples models working under different theoretical frameworks may be implemented and tested if it is clear they do not fit the expected behavior. At this point, using a lexical graph is considered as the first modeling method to try.

The final format of the thesis (either a monograph or a publication compilation) as well as the timeframe devoted to its writing will be discussed with the supervisor after arrival in Japan, as it is not yet determined. While a complete overview of the subject is a project of its own, the privacy question is important for any real world application or large-scale experimentation. As such, a minimal overview of Japan and Europe legislations about personal data usage and technical means of protecting privacy (e.g. differential privacy algorithm) should be written. This summary will be delivered as a chapter of the final thesis.

**Intermediate Publications**

Work done during the system design phase and experiment results could be shared with the research community at domestic or regional events. Japan hosts the JALTCALL conference each year and the ASIACALL conference is held in various places in Asia. Some results might fits in more general conferences such as COLING that focus on Computational Linguistics.

**Detailed Proposal with References**

A more comprehensive doctorate proposal that give more details about the context, research questions, approach and which lists a few references is available as document ③-2.

---

1 This schedule does not include post-arrival period as a non-degree seeker research student before taking university's entrance examination. That interval of time will be used mainly for entrance examination preparation as well as learning more Japanese.