II. Brief Introduction of the Project

(*Involved field of science and technology, main technological contents, authorized patents, technical and economic indicators, application and promotion, talent cultivation, and promotion of science and technology progress in Macao*)

Background

Macau is a city that has a high demand on the translation between Portuguese and Chinese. Even after the handover in 1999, Portuguese is still considered as an official language in addition to Chinese. Due to the lack of people with bilingual proficiencies, besides handling day-to-day translation tasks, which include official documents, bulletin statements, news, technical specifications, proposals, etc., there are still many old documents waiting for translation. Although much is said that the Chinese language is becoming more important after the handover, one cannot overlook economic and diplomatic relationships that Macau has forged with other Portuguese-speaking countries with the use of the Portuguese language. Since the establishment of Special Administration Region, Macau has been maintaining and consolidating its implicit relations with those countries as well as continuing to exert its role as a bridge of exchange.

Problems Statements : Portuguese and Chinese Language Deviations

Since Portuguese and Chinese languages come from different language families, they are very different in terms of writing and grammar. As an example, consider the figure below.
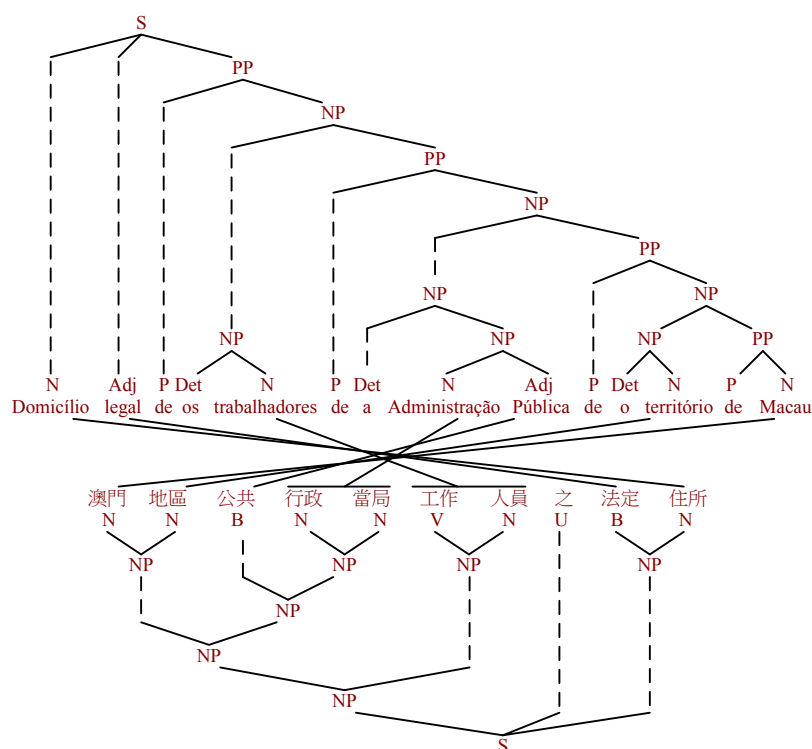
Many words in Portuguese may appear in a different sequence order in Chinese, represented by the crossing lines between the sentences. Moreover, linguistic expressions can vanish or appear in the translation. For example,

the preposition "的" (of) often will not be translated even the Portuguese sentence has the preposition "*de*" (of), like "*território de Macau*", and "*trabalhadores de a Administração Pública*". There are also cases in which the words in Portuguese are composed by more than one Chinese word for the translation. As an example, the verb "*emprestar*" (to lend) is composed of "把" and "借給了" (to lend) in the Chinese translation. In the syntax level, crossing relationships are even more complicated and difficult in modeling their relationships.

Uniqueness and Significance

Although nowadays it is so easy to get translation tools, there isn't any practical and commercial Machine Aided Translation System especially developed for Portuguese and Chinese languages, which shows our **uniqueness** in this field. Moreover, we are one of a few research groups which focus in the R&D of Machine Translation (MT) between Portuguese and Chinese.

As more documents are required to be translated, and due to the lack of translators with bilingual proficiencies, the translation task within an acceptable time frame more practical with the support of our powerful Machine-Aided Translation technologies and tools targeted for this language pair in Macau.

Contributions

Below lists out the main achievements and contributions in this project:

1. Kernel technologies had been grasped by local talents of Macau:
   - Translation Corresponding Tree (TCT) Annotation Schema
     - Novel representation in associating the relationship between bilingual examples for language pairs which do not belong to the same family
   - Constraint Synchronous Grammar (CSG)
     - Modeling the relationship between source and target syntactic structure and sense ambiguity based on feature constraints
     - Different from traditional transfer based MT paradigms, parsing the grammar and generating the translation are performed at one-stage
2. Application of kernel technologies in the development of several systems and prototyping, including:
   - PCTDict (Portuguese Chinese Bilingual Dictionary)
     - Facilitate user from searching word meanings by using the mouse tracking function
     - Ease the learning of the pronunciation of Portuguese and Chinese words
     - Include not only basic meanings of the words, but also phrases used in Macau
   - PCTAssist (Portuguese-Chinese Machine Aided Translation system)
     - Hybrid system combined with Translation Memory and MT technologies
     - Design workflow fits the traditional translation life cycle of professional translators.
     - Translation stylist for each user can be captured
   - Network-based PCTAssist
     - Knowledge and Vocabulary sharing is possible for users connected to the system
     - Users can reuse the translation contribution from predecessors in order to increase the output of the translation work
   - English to Chinese MT prototyping system, Vision to Text translation prototyping system
   - Some of these received local and national awards from different organizations

3. Advancement of the research and development (R&D) of MT and related technologies has made a big and rapid progress in Macau

4. We are one of a few research groups which focus on the R&D of MT between Portuguese and Chinese

Social Benefits

1. PCTDict and PCTAssist have been transformed into commercial products, and had been used by the Macau society.

2. There are more than 700 registrations, and the users are mainly from schools, law firms, companies, and more than 30 government departments.

3. Launching ceremonies are organized and they are open to the public in reporting our research achievements and in sharing how Machine-Aided technologies can reduce the translation workload.

4. Direcção dos Serviços de Administração e Função Pública (SAFP) as a representative of all government departments, has established a Memorandum Of Understanding (MOU) with University of Macau for pursuing the use of our developed PCT systems in the departments of the Macau government.

Talent Cultivation

Many interdisciplinary students were trained and cultivated, ranging from Bachelor to Ph.D. levels. They are trained to have strong background knowledge and technical skills in the fields of Machine Translation, Computational Linguistic, Speech and Natural Language Processing.

Authorized Trademark

For distinguishing our R&D products provided from other entities, trademark registration was applied in the Macao SAR Economic Services and effective from 21st May, 2010. The English name "PCT" (N/045766), Chinese name "中葡通" (N/045767), and the logo of the products (N/045768) are included in the trademark. This can better identify our uniqueness in the society, distinguish our products from others, and it can protect any unauthorized use of the trademark through legal procedures.

Development History

In order to response to the huge demand of translation between Portuguese and Chinese in Macau, since 1998, University of Macau together with "Instituto de Engenharia de Sistemas e Computadores" (INESC-Macau) and Tsinghua University started the research and development (R&D) on this field. Under the financial support firstly from the Fundação para a Cooperação e o Desenvolvimento de Macau and then from the Research Committee of University of Macau, a series of systems have been developed, including "Portuguese-Chinese Bi-Directional Word-by-Word Translation System" (PCTDict v1.0, v2.0, and v3.0), and "Portuguese-to-Chinese Sentence-by-Sentence Machine Aided Translation System" (PCTAssist v1.0 and v2.0). In 2005, our research has been defined as one of the University Key Research Areas in University of Macau. Starting from year 2005, the project got strong support from "Fundo para o Desenvolvimento das Ciências e da Tecnologia" (FDCT) and University of Macau to further investigate our research work.

Science Promotion

We are fostering the use of Portuguese and Chinese languages in Macau through the use of computer

technologies, and preserving indirectly a unique characteristic in the world, which is the use of this language pair, a fusion of eastern and western cultures for more than 400 years in the history of Macau.

(Not exceed 1,200 words)