

# Structuring an Ontology of the Basic Vocabulary of Tourism

Patrícia Tosqui-Lucks and Bento Carlos Dias Da Silva

**Abstract**—In an attempt to fulfill the needs of a student of Tourism to master the basic vocabulary of tourism and to contribute to the semantic-conceptual study of the lexicon, this paper presents an ontological structuring of the basic vocabulary of tourism which, on the one hand, constitutes a linguistic and pedagogical resource and, on the other hand, can be integrated to specific lexical data bases. Firstly, we present the different kinds of sources which helped us establish the concepts and from which the lexical items were extracted and were constituted by dictionaries, technical books of Tourism and didactic books of English for tourism students. Then we present some problems which were posed during the structuring of the ontology and some possible solutions for them. In the end, we presented a sample of the ontology using the ontology editor Protégé 3.3. We have selected, from each of the different types of sources, lexical and conceptual information relevant to the assembly of both the ontology and vocabularies.

**Index Terms**—Basic vocabulary, ESP, tourism, ontology, Protégé Ontology Editor

## I. INTRODUCTION

In this work we present some results of our Doctoral dissertation [1], in which we built a proposal of ontology for the Tourism domain. An ontology can be considered the “specification of a conceptualization” or, in other words, “a specific artifact designed with the purpose of expressing the intended meaning of a shared vocabulary”[2]. An ontology describes the concepts and relationships that are important in a particular domain, providing a vocabulary for that domain as well as a computerized specification of the meaning of terms used in the vocabulary. This paper presents part of the process of building an ontology of Tourism in order to assist students of English for specific purposes (ESP). The aim of this research was to help Brazilian students of English who are been prepared to work in the different areas of Tourism - such as hotels, travel agencies, restaurants, airports, among others - understand the concepts of this domain and improve their knowledge of the vocabulary used in professional situations. This is especially relevant because Brazil, which is an emerging economic powerhouse, will be the host country of two major international sports events: FIFA World Cup in 2014 and the Olympic Games in 2016, and expects a great increase in the number of international visitors.

Manuscript received May 20, 2012; revised June 14, 2012.

P. T. Lucks is with São Paulo State University (UNESP), Brazil (e-mail: patricialucks@uol.com.br).

B. C. D. D. Silva is with Institute of Air Traffic Control (ICEA), Brazil

## II. BUILDING THE ONTOLOGY

### A. The Sources of Concepts

In order to elaborate this ontology of tourism, we have consulted different kinds of sources to enable the understanding and the collecting of the concepts related to the specialized domain of “Tourism”, and the corresponding vocabulary, in the form of lexical items that express these concepts, both in English and in Portuguese. To extract the concepts that would consist the ontology, the following sources were consulted: technical texts about tourism, books of English for Tourism students (ESP), onomasiological dictionaries of English for learners, and the semantic nets Princeton WordNet and Berkeley Framenet. The selected dictionaries were: *Longman Language Activator (1997)*, *Longman Essential Activator (2005)*, *Cambridge Word Routes English-Portuguese (1999)* and *Longman Lexicon of Contemporary English (1981)*. We have chosen these dictionaries because they are all designed for English learners and they organize the lexical entries according to the concepts they express, and not in alphabetical order.

### B. Delimitating the Scope of the Ontology: Some Problems and Possible Solution

After collecting the concepts and corresponding lexical items from the sources, we started structuring the first version of the ontology, selecting classes and subclasses, which are represented with capital letters. As in any ontology, the possible relations between the concepts are: subordination and superordination, part-whole and inclusion, identity, similarity, contrast or opposition [3].

During this task, we realized that similar concepts could be expressed in different domains. For example, the concept FOOD AND DRINK is related to the concept RESTAURANT which, in turn, also occurs in the ontology in other relationship, i.e., subordinated to the concept FACILITIES, which is subordinated to the concept HOTEL. How to formally represent the relationships that are established at different inter and intra levels? This was the first problem we had to solve. Besides this kind of issue, we noted that both concepts and lexical items could establish different types of semantic relations. We have, for example, a super ordination relationship between the concepts TOURIST and KIND OF TOURIST, but the concept TOURIST also involves subconcepts grouped by the labels THINGS TOURISTS CARRY and THINGS TOURISTS BUY. How to improve the description so that there is as much uniformity as possible between the concepts? Other questions have arisen as the ontology was developed. For example, the concept PLATFORM is similar in TRAIN STATION and BUS STATION. The concept CABIN can be

part of the concepts TRAIN and AIRPLANE, but with differences in meaning. How to represent and relate similar concepts in different subclasses? How to explain very different concepts which are expressed by the same lexical item? These questions led us to the hypothesis that to develop the ontology of tourism for our purposes, it would be necessary to establish fully and formally understandable relationships between concepts. To help us resolve issues of this nature, we analyzed the operation of sources that are structured formally from semantic relations: the WordNet and FrameNet. These networks provide lexical-conceptual information, because they organize lexical items according to the concepts lexicalized by them, and specify the relationships that are established between the concepts.

### C. The Concepts

In our ontology, the concepts are structured as follows:

- LEISURE TOURISM, MASS TOURISM, ECOTOURISM, EVENTS TOURISM, BUSINESS TOURISM and CULTURAL TOURISM are KINDS OF TOURISM.
- TOURISM BUSINESS includes TRAVEL, TOURIST INFORMATION OFFICE, BROCHURE, TOUR OPERATOR and TRAVEL AGENCY.
- TOURIST INFORMATION OFFICE includes TOURIST INFORMATION OFFICER.
- TOUR OPERATOR includes TOUR COMPANY REPRESENTATIVE.
- TRAVEL AGENCY includes PACKAGE TOUR, FARE, SEASON TRAVEL AGENT, RESERVATION and CANCELLATION.
- HIGH SEASON and LOW SEASON are opposite kinds of SEASON.
- HOLIDAY, HONEYMOON, EVENT, BUSINESS and ECOTOURISM are kinds of MOTIVATION.
- TRAVELLER, HOLIDAYMAKER, HONEYMOONER, BUSINESS TRAVELER, EVENT PARTICIPANT and BACKPACKER are kinds of TOURIST.
- BAGGAGE, DOCUMENTS, MAP and TRAVEL GUIDE are kinds of THINGS TOURISTS CARRY.
- BAG and SUITCASE are kinds of BAGGAGE.
- PASSPORT, VISA, VOUCHER, INSURANCE and TRAVELLER S CHECK are kinds of DOCUMENT.
- SOUVENIR, POSTCARD and HANDICRAFTS are kinds of THINGS TOURISTS BUY.
- AIR TRAVEL contains AIRPLANE, AIRPORT and AIRLINE.
- AIRPLANE contains FLIGHT, which contains CLASS and FLIGHT ATTENDANT.
- AIRPORT contains GATE, TERMINAL and CHECK-IN COUNTER. GATE, TERMINAL and CHECK-IN COUNTER are parts of AIRPORT.
- ROAD TRAVEL includes CAR, BUS and BUS STATION. PLATFORM is part of BUS STATION.
- RAILWAY TRAVEL contains TRAIN and TRAIN STATION. BERTH is part of TRAIN and PLATFORM is part of TRAIN STATION.
- WATER TRAVEL contains BOAT, FERRY BOAT, CRUISE, CRUISE SHIP and CRUISE LINE.

- ARRIVAL is opposite of DEPARTURE.
- HOTEL contains LOBBY, FACILITIES, RECEPTION, ROOM and STAFF.
- SWIMMING POOL, FITNESS ROOM. ROOM SERVICE, CONFERENCE FACILITIES, RESTAURANT and BAR are kinds of FACILITIES, which is part of HOTEL.
- CHECK-IN is opposite of CHECK-OUT.
- SINGLE ROOM, DOUBLE ROOM, TWIN ROOM, TRIPLE ROOM, FAMILY ROOM and DORMITORY are kinds of ROOM.
- BED, TELEPHONE, SAFE, MINIBAR, TV and BATHROOM are kinds of ROOM FACILITIES
- BATH, SHOWER, SINK, TOILET and TOILETRIES are contained in BATHROOM.
- PARK, FOREST, NATURE RESERVE, MOUNTAIN, WATERFALL, BEACH, RIVER, CAVE and WILDLIFE are kinds of NATURAL ATTRACTION.
- MUSEUM, ART GALLERY, CASTLE, CHURCH, TEMPLE, ZOO, AQUARIUM, FESTIVITIES, THEATER, RUIN and MONUMENT are kinds of CULTURAL ATTRACTION.
- THEME PARK, AMUSEMENT PARK, CASINO and NIGHTLIFE are kinds of ENTERTAINMENT.
- NIGHTLIFE contains DISCO and BAR.
- TOUR and TRIP are similar.
- CITY TOUR, SIGHTSEEING, WALKING TOUR and EXCURSION are kinds of TOUR.
- TOUR GUIDE is part of TRIP.
- BUNGEE JUMPING, HIKING, TREKKING, SKIING, SNOWBOARDING, CLIMBING, MOUNTAIN-BIKING, DIVING, SCUBA DIVING, SWIMMING, WATER SKIING and RAFTING are kinds of SPORT.
- RESTAURANT contains MENU, MEAL, KITCHEN, ROOM, TOILET ROOM, STAFF and CLIENT.
- MENU contains APPETIZER, DISH, BEVERAGE and DESSERT.
- BREAKFAST, LUNCH and DINNER are kinds of MEAL.

These were the concepts that composed the first version of the ontology. As it was being constructed, other concepts were necessary and some relations were modified, in order to make the ontology more complete and coherent. The next step, which will not be presented thoroughly in this paper because of space constraints, was to provide the lexical items that corresponded to each concept, both in English and Portuguese, that is, the bilingual vocabulary anchored to the Tourism ontology. In the next item, we will present some examples of how we used the ontology editor Protégé 3.3 [4] to organize the concepts and their relations and to present their lexicalization in English and Portuguese.

### D. The Ontology Editor Protégé

In order to build the ontology we used the free, open-source ontology editor and knowledge-base framework Protégé 3.3. This platform was chosen because it offers a suite of tools that support the creation, visualization, and manipulation of ontologies to construct domain models and knowledge-based applications [5, 6]. The ontology was built with around 200 concepts, listed above, which have, as

instances, the corresponding lexical items in English and their equivalents in Portuguese.

We begin our TOURISM ontology by inserting the class and its main subclasses. The main subclasses are: KINDS OF BUSINESS, BUSINESS TOURISM, MOTIVATION, TOURIST, TRANSPORTATION, ACCOMMODATION, ATTRACTIONS, ACTIVITIES and FOOD AND DRINK.

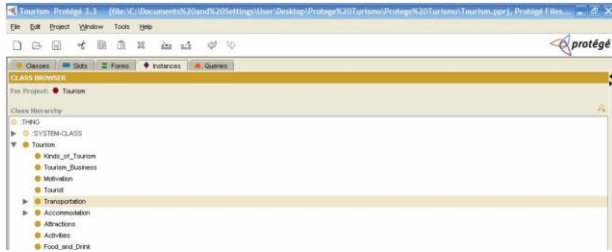


Fig. 1. Insertion of the class TOURISM and its main subclasses.

We also specified the properties. For example, in Figure 2, we represent the relation of opposition between CHECK IN and CHECK OUT, which are contained in RECEPTION. The last example is the representation of the bilingual vocabulary anchored to the ontology. In Figure 3, we represent the lexical items of the concept RECEPTION in English, which represent similarity (*reception, front desk and reception desk*) and in Portuguese (*recepção*). It is possible to notice that there is no identity in the numbers of lexical items used to express the same concept in both languages.

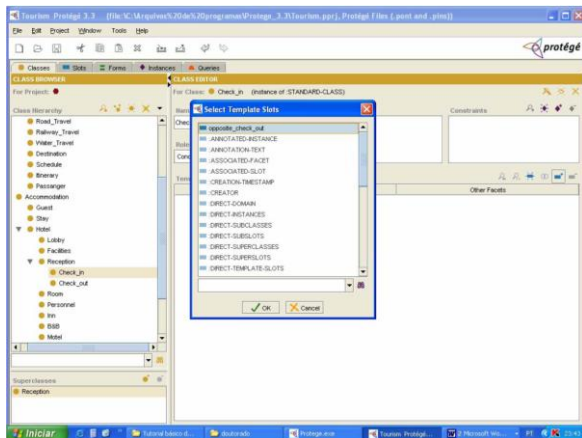


Fig. 2. Relation of opposition between CHECK IN and CHECK OUT.

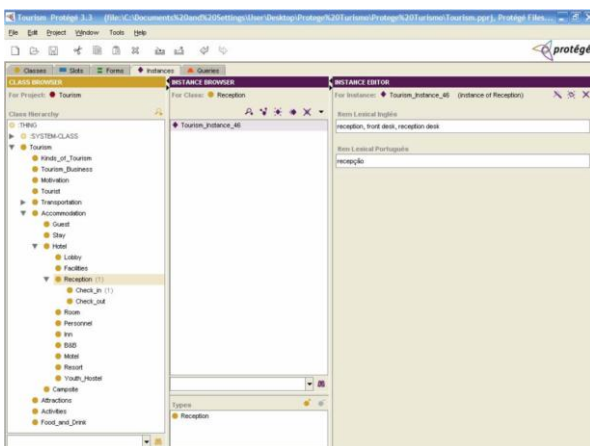


Fig. 3. Insertion of the lexical items in English and portuguese.

### III. FINAL CONSIDERATIONS

In this paper, we have presented the construction of a basic ontology of the field of tourism with a bilingual English-Portuguese vocabulary anchored to it. Firstly, we presented the different kinds of sources which, for their complementary nature, helped us establish the concepts and from which the lexical items were extracted. Then we presented some problems which were posed during the structuring of the ontology and some possible solutions for them. In the end, we presented a sample of the ontology using the ontology editor Protégé 3.3. We have selected, from each of the different types of sources, lexical and conceptual information relevant to the assembly of both the ontology and vocabularies. If the books specialized in tourism taught us how to classify the different branches that make up this activity, the specific books ESP contributed fundamentally to the selection of lexical items. The lexicographical sources, in turn, proved to be reliable and versatile because if on the one hand, they represented the authority expected of a reference work on the other hand, they proved that paper dictionaries are not in their final days, as already advocated, but have adapted to new realities and needs of learners of a foreign language. The last type of source of research information shows the contribution of computational tools to assist in tasks that would be delayed or impeded if performed manually. Networks WordNet Princeton and Berkeley FrameNet revealed a world in which the speed of the consultation, the amount of information available and the possibilities for manipulation, analysis and data processing could become a pleasurable activity that would, without doubt, be impossible to be performed only with traditional dictionaries.

The motivation for conducting this research arose from the need to develop a material that could serve as an aid to teaching English language courses for Tourism students. To do this, we recurred to the concept of ontology - a term borrowed from philosophy, which means the investigation of "all things" and that was redefined by the Artificial Intelligence and its related fields to describe studies that attempt to explain and organize formally the concepts that represent entities and processes in the world in the task of building a simulacrum of the artificial environment of computers. A pretentious task, no doubt, full of complexities - the same as they are complex relations between world, mind and language - but embraced by those who see the computer capable of performing an action that attempts are still a challenge to man. There are many possible future developments to continue our work, such as: to expand the concepts of the ontology and, consequently, the vocabulary; to insert other languages; to make very specific ontologies of some concepts, such as ACCOMMODATION or TRANSPORTATION, for example, exploring other concepts and relations inherent to them.

### ACKNOWLEDGEMENTS

We acknowledge CNPq for providing financial support for the presentation of this paper at ICKET 2012.

REFERENCES

- [1] P. T. Lucks, "Construction and Ontological Anchorage of the Basic Bilingual Vocabulary of Tourism for Pedagogical Use," Doctoral dissertation, São Paulo State University, 2007.
- [2] G. Hirst, "Ontology and the Lexicon," In: S. Staab and S. Studer (eds) *Handbook on Ontologies*. Berlin: Springer-Verlag, pp. 209-229, 2004.
- [3] S. Nirenburg and V. Raskin, *Ontological Semantics*. Cambridge: MIT, 2004.
- [4] *PROTÉGÉ*. [Online]. Available: <http://protege.stanford.edu/>
- [5] N. F. Noy and D. L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*. [Online]. Available: [http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html)
- [6] E. Sachs. *Getting Started with Protege-Frames*. [Online]. Available: [http://protege.stanford.edu/doc/tutorial/get\\_started/get-started.html](http://protege.stanford.edu/doc/tutorial/get_started/get-started.html)