CHAPTER TWENTY-FOUR

CHALLENGES IN MANAGING TERMINOLOGICAL RESOURCES: THE CASE OF AN IATE PROJECT

ELPIDA LOUPAKI AND RODOLFO MASLIAS

1. Introduction

In our globalized multilingual society terminology and its digital processing has attracted increasing attention over the last twenty five years. As Teresa Cabré has already pointed out (1998, 4): "Terminology is […] affected by social changes which [have] a major effect on linguistic needs". The evolution of the information society has changed all aspects of social organization; the continuous developments in the sciences and technology has resulted in the creation of new concepts and terms; the intensified movement of humans and products all over the globe has brought about the need for safety rules and standardization, and these are just a few of the reasons behind the expansion of terminology management.

The aim of this chapter is to investigate the challenges that can be encountered when dealing with terminological resources. To this end, we will start with a general overview of an IATE project undertook by Masters Students at the Aristotle University of Thessaloniki, in cooperation with the TermCoord Unit of the European Parliament. More specifically, we will briefly present some facts and figures about the IATE terminology database; the framework of IATE projects; the participants in the project; the domain of terminology research; the methodology used throughout the project; and the deliverables. In the next section, the concept of terminology management will be explored with particular reference to the domain of migration. Finally, specific examples taken from our corpus will be analyzed in order to underline different issues that

emerged while working on the project. Our approach is descriptive and empirically based.

2. Research Motivation: Presentation of the Project

2.1 IATE, Terminology Database

IATE, an acronym for "Inter-Active Terminology for Europe", is the EU's inter-institutional terminology database. IATE has been used in the EU institutions and agencies since summer 2004 for the collection, dissemination and shared management of EU-specific terminology. The database partners are: the European Commission, European Parliament, the European Council, the Court of Justice, the Court of Auditors, the Economic and Social Committee, the Committee of the Regions, the European Central Bank, the European Investment Bank and the Translation Centre for the Bodies of the EU.

With over 1.5 million concepts and more than 8.7 million terms, IATE is one of the major terminology databases in the world, its aim being to support the multilingual drafting of the EU texts and legal texts, in particular (see http://iate.europa.eu/about_IATE.html).

2.2 IATE Projects

Since 2012, TermCoord has collaborated with terminology or language departments of universities on terminology projects to feed IATE. Students work on terminology projects, in line with the requirements for IATE terminology work and the respective guidelines, researching and documenting terms in a main source language (English or French) and a target language of their choice (among the official languages of the EU). The advantage for the students is that they work on terminology in the same way as the translators of the European Institution who use an interactive version of the database. The findings are checked and verified by European Parliament terminologists and inserted into IATE (information available at http://termcoord.eu/universities/cooperation-with-universities-on-terminology-projects/).

2.3 Participants in the Project

Project partners were from one hand the European Parliament and specifically TermCoord Unit and from the other hand Aristotle University of Thessaloniki and in particular the MA in Translation.

2.3.1 TermCoord

Founded in 2008, the Terminology Coordination Unit of the European Parliament coordinates, as its name suggests, the terminology produced by the translation units by some 1200 translators and assistants in 24 languages through IATE's interactive database. TermCoord undertakes *proactive terminology*, monitoring the legislative procedures to identify the terminology needs of the translation units. Therefore, the team prepares "TermFolders" with reference material on important issues, searches for topic-specific glossaries and ensures the linguistic consistency of the European Parliament's databases. Since 2011, TermCoord has launched an initiative for the creation of an Interinstitutional Terminology Portal. The Portal includes wikis in each language dedicated to terminology cooperation between the translators of all the institutions, it also provides glossaries, information training programmes, e-books and academic articles on terminology.

2.3.2 MA in Translation, Aristotle University of Thessaloniki

The Aristotle University of Thessaloniki is a first-rate university covering a wide range of disciplines in South-eastern Europe. As from 2004 it offers a joint-departmental Master Degree in Professional Translation and Interpretation. In 2014, the Master in Translation received the EMT Label of the DG Translation of the European Commission. The university departments who run the course are the Schools of English, French, German and Italian Language. The course aims to provide students with high level of education in professional translation. The curriculum includes, among other things, a course in Terminology Research and Documentation, in the framework of which this project was undertaken (http://termcoord.eu/universities/universities-projects/).

2.4 Domain

As a domain for terminological research we chose migration and human trafficking. This choice is by no means accidental, as during the last five years 5, 6 million new migrants are estimated to have arrived in Europe (UN, DESA 2009). This number concerns only documented migrants, as irregular migration is, by definition, difficult to measure. According to data from FRONTEX, Greece is the major gateway of undocumented migrants and asylum seekers entering Europe from Asia and Africa (Data available in the webpage of the International

Organization of Migration (IOM), http://greece.iom.int/iom-greece). As argued by UNESCO, by 2001, 44% of the developed countries have adopted policies aiming at lowering migration against 6% in 1976 (http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/SHS/pdf/migr ation_presentation.pdf). This change of direction also leads to the increased undocumented movement of people in different forms, including trafficking and clandestine migration. Even when migrants manage to enter European countries, they often face growing discrimination, racism and xenophobia. At the EU level, migration has been receiving more attention during recent years: the funding of FRONTEX, the strengthening of Europol, the Lisbon Treaty or the Commission's Agenda on Migration are only some of the attempts to manage migration flows in a balanced way.

As is usually the case, new realities give birth to new terms or collocations in language and we are currently witnessing a proliferation of new terms describing the domain of migration and its different aspects. For instance, in English, relatively new terms, such as "human trafficking", "brain drain" and "asylum shopping" are coined, on the other hand, the translation into Greek of all terms related to migration is not always available in IATE.
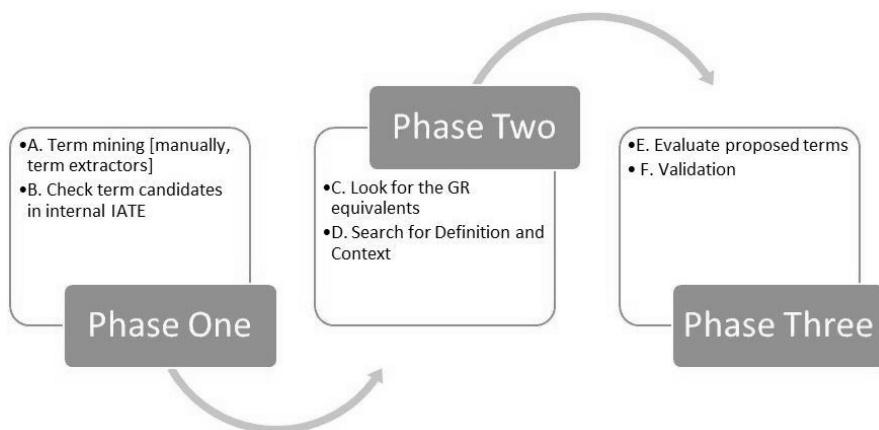
2.5 Methodology

Fig. 24.1 Terminology Project Workflow

As depicted in the Fig. 24.1, we have started by searching for English term candidates in big databases such as *Digital Library on Trafficking*,

the site of the International Organization for Migration, the European Parliament's Think Tank or the Legislative Observatory and of course in Eur-Lex. During phase one, students were asked to find just three terms. One basic requirement in order to retain a term was that it should not be already incorporated in the IATE database for the same domain and for the English and Greek language pair. To this end, students were asked to check all term candidates in the internal IATE terminology database, to which TermCoord had given them access during the project. During phase two, the students were asked to search for Greek equivalents for their three terms. In this phase all Greek equivalents were recorded as our approach is descriptive and not prescriptive. Subsequently, students were asked to search for definition and context. This was the most demanding stage, as lots of terms were not solidly documented and were then discarded. The final stage was that of evaluation: terms were evaluated with reference to specific criteria (see section 4.1 below) and were validated, or else students had to proceed to a new search.

2.6 Deliverables: Data Categories

Our deliverables were 180 terminology records, including most of the information found in IATE. In particular, at Language Level there was the definition of the concept, the reference of the definition and a note (if any). As stated in the instructions given to us by TermCoord:

> Definitions must be similar in all languages and applicable to all terms on the same record. Ideally, a definition would be placed at Language-Independent Level. However, to allow a definition in each language, the definitions in IATE appear at Language Level (internal document).

At Term Level, information available was the term itself, a reference to the term, a context of use and a reference of the context. We were not asked to provide grammatical information nor Language-Independent Level information, such as domain, cross references,[1] problem languages,[2] etc.

After this brief outline of the project, we shall now move on to the concept of terminology management and its application in the domain of migration.

3. Terminology Management and Migration

The notion of Terminology Management was coined in the late 1990s by equivalence to Information Management and because terms such as

terminography or *terminology work* weren't widely accepted (Wright and Budin 1997). According to Wright and Budin (1997, 1), terminology management is "any deliberate manipulation of terminological information". As the authors themselves admit, this is a very general concept that covers a wide range of activities from term extraction and selection to database structuring and corpus-related applications.

Although terminology activity is almost universal, terminology management does not constitute a uniform practice, as different schools with different methodologies have appeared all over the world. However, if we would like to make a major distinction, we should distinguish between: a) Descriptive Terminology Management and b) Prescriptive Terminology Management. The main difference between these two approaches lies in their scope and potential users. For instance, Descriptive Terminology Management prepares terminology products for both subject-field experts and lay practitioners, such as translators, technical writers or students. This is the reason why the descriptive approach does a full inventory of every term used for one concept, without excluding those not preferred. On the other hand, Prescriptive Terminology Management is the normative approach within the framework of standardization bodies such as the ISO. In this approach, one concept must be represented by one term (*univocity principle*). As already stated, our approach is descriptive.

As suggested by Riggs, Mälkiä and Budin, in their article discussing the role of terminology in Social Sciences (1997, 184-196), "The first step in a descriptive terminology project involves identification of a subject field and representative documentation of its literature." During this stage experts or terminologists involved in the project should (ibid.):

1. identify the key works in the field
2. identify the basic concepts of the field
3. find the definition of these concepts
4. find the terms used to describe the concepts
5. select contexts of use.

In general, these were the steps we took while working in the domain of migration. The only exception was that we didn't identify the main concepts of the domain, as our work was to feed the IATE terminology database and some of the basic concepts were already incorporated.

In relation to key work in the field, the situation is quite complicated. For instance, as a social phenomenon migration is multidimensional. This multidimensionality is also reflected in the texts connected to migration. These cover a wide variety of genres and are issued from different institutions, such as the UNHRC, the European Parliament or the national

authorities, NGOs such as the Médecins sans Frontières or Amnesty International, as well as from academic scholars. (For a typology of texts in the domain of migration, see Table 24.1).

A	Legal texts	EU documents (directives, resolutions, regulations, CJEU rulings, conventions) Greek documents (Official Government Gazette)
B	Administrative texts	Public services reports, police instructions, frontier authorities documents, health services guides
C	Academic texts	Academic papers in the subject matter of migration, seen from different scientific angles
D	General public texts	NGOs information leaflets, NGOs reports, glossaries, press articles, university course descriptions, blogs

Table 24.1 Typology of texts in the domain of migration

Texts were grouped into four distinct categories, in relation to their communicative settings (mainly sender, receiver and target). The first three categories are characterized by a higher degree of specialization, while the fourth category contains texts addressed to the general public of a less specialized nature.

The language of migration is therefore not an independent and well defined specialized type of discourse. It could be argued that it forms part of legal/administrative language, as a high number of texts are legal texts; it could also be argued that it forms part of the language of social sciences, as numerous academic papers using sociology, psychology or political sciences as their theoretical background are found in our sample; finally NGOs' reports that mix specialized language and general language or press articles are also an important part of this discourse. In all cases, terminology in the domain of migration does exist and is strongly influenced by the multiplicity of media involved in its production.

As far as terminology management in the domain of migration is concerned, while searching in our *ad hoc corpus*[3] of Greek texts we found some inconsistencies in the use of different terms to describe the same concept. One example is the term "pink card" which is a synonym for the term "asylum seeker's card", and its translation in Greek is "ροζ κάρτα" [back translation: pink card]. However, there is also another type of card,

commonly known in English as "red card" ("special asylum seeker's card") which is also translated into Greek "ροζ κάρτα" [back translation: pink card] (UN Refugee Agency). Another example is the term "undocumented migrant" which is translated into Greek "μετανάστης χωρίς νόμιμα έγγραφα" [back translation: migrant without legal documents] but it also found as "παράνομος μετανάστης" [illegal migrant]. These inconsistencies may suggest that terminology management in the domain of migration needs a more systematic approach, which we tried to offer, albeit on a small scale.

The challenges we met while working on this project will be analyzed in the following section.

5. Challenges in Managing Terminological Resources

5.1 Term Selection

One of the first questions that occurred was: Which terms should be included in the project? To answer this question we consulted both related bibliography and the IATE Input Criteria, as expressed in the handbook *Best Practices for Terminologists* (internal document).

As stated by Sue Ellen Wright (1997, 13), "Terminologists must select the terms that will be documented in their collections based on the objectives and requirements defined for their working environment". Marie-Claude L'Homme extends this argument when she enumerates the criteria a lexical unit should meet, in order to be included (or not) in a terminology project (2004, 57). These are:

- relevance of the domain.
- identity of the user.
- frequency and distribution.

A basic criterion for introducing or not a concept to a terminology collection is the domain's relevance. As explained by L'Homme (*ibid.*), when preparing a collection about IT, terms such as "mouse", "monitor" or "hard disk" should be included. On the other hand, if the domain is Optical Character Recognition (OCR), these terms should be excluded. The identity of the user is also of great importance, as different users have different linguistic needs and expectations. For instance, our choices would be totally different if we were to create a terminology collection for experts or for students.

Frequency is defined as the number of occurrences of a lexical unit in the totality of specialized texts and its distribution. Therefore, a term may

have a high frequency in a corpus of specialized texts, but a low distribution if it occurs only in a small number of them. Of course, as stressed by L'Homme herself, all these numbers are indicative and not absolute because there are cases where a lexical unit mentioned only once can be a very important term in a particular domain.

If we try to apply this approach to the IATE project, we can observe the following: as stated before, the main goal of the IATE terminology database is to provide terminology for legislative activities. On the other hand, this IATE project was undertaken in collaboration with TermCoord, which is a Unit of the European Parliament. From a sociolinguistic point of view, the European Parliament's linguistic needs present some differences in comparison to those of the European Commission or the Court of Justice. In particular, as stated on the EP's site:

> Human rights are among the main priorities of the European Parliament. Parliament is a key actor in the fight for democracy, freedom of speech, fair elections and the rights of the oppressed.[4]

Moreover, documents produced within the EP, such as reports, resolutions, petitions, speeches delivered in the plenary, etc. are not legal texts. Therefore, our final user has greater linguistic needs and the domain of our research should not be confined to legal documents exclusively.

According to the instructions given to us by TermCoord, there are some general input criteria that must be respected when dealing with term candidates. These are as follows:

1. added value. Entries found in IATE must have an added value over other terminological resources found in the Internet. To achieve this goal terminology work should have consistency and coherence (for instance, a document search, the addition of a definition or reference, the designation of the preferred term for consistency reasons, the endorsement of a solution suggested by a translator, etc.)
2. relevance. Terms entering IATE should have a past, present and future within the linguistic environment of the EU. For this reason, priority is given to *proactive terminology*.
3. avoidance of duplicates. If a term already exists, creating a new record would result in inefficient terminology management. To avoid duplicates, terminologists or translators entering data in IATE must carry out a standard check of existing entries.
4. accuracy of data. Data introduced to IATE should be as accurate as possible. In particular, when adding their language to an existing entry, terminologists must check that their term and data match the concept already present for the entry by other languages.

5. single concept. Every entry should deal with one concept only and all data relating to a given concept should be consolidated in one entry.

6. minimum information. Every entry should include as much information as needed to enable unambiguous communication. To this aim, both definition and context should be included in a terminology record.

7. intellectual Property Rights (IPR). Regarding the issue of copyright, terminologists are recommended to adopt a precautionary approach. (*Best Practices*, 5-6).

Having all the above points in mind we checked and evaluated our term candidates and some terms were discarded for one of these reasons. For example, the term "bush-wife" had several occurrences in our corpus but a low distribution, as it was only mentioned in a document related to Sierra Leone; for this reason it was rejected.

5.1 Efficiency of Term Extractors

Term extraction tools, developed over the last decades, have been proven to be of great help in treating significant amounts of field-specific texts in less than a minute. As pointed out by Bernier-Colborne (2012, 15) "Since methodologies for compiling [terminological] resources are increasingly corpus-based, one of the main tools is the term extractor." Term extractors are described by the author as "tools designed to retrieve specialized terms from running text, which play a role in a variety of applications" (*ibid.*).

However, aside from speeding up the extraction procedure, various problems have already been documented in the use of term extractors (cf. Aubin and Hamon 2006, Bernier-Colborne 2012). These are related to term recognition (especially in complex terms, difficulty in distinguishing between terms and nominal phrases); to synonymy (concepts may be denoted by more than one term, existence of variances); noise (problems related to tagging).

In our research, we used open access term extractors, which combine linguistic and statistical models, such as TerMine, Five Filters, Labstranslated.net and WebCorp. The problems we faced can be grouped in two main categories:

- character recognition problems.
- frequency problems.

In the first category, one linguistic element that caused errors was hyphen words. In English, hyphens are commonly used to break single words (especially in PDF files) or compound words. Term extractors used in our research did not recognize hyphen words as one linguistic unit and this led to erroneous results. Another major issue, while working with term extractors, was elements with high frequency (such as conjunctions, or commonly used words) that ranked high in the lists.

5.3 Lack of Equivalence—Lack of Consolidation

Another issue that the students had to deal with while doing the terminology work was the lack of equivalent terms in the Greek bibliography or the lack of consolidation. Examples of terms that did not have any documented equivalence in Greek texts are the terms "child slave" and "bush wife". In this case, students had to abandon these terms and search for new ones. Terms not yet consolidated in Greek may sometimes contain the English term in parenthesis. This is the case of "asylum shopping" and "brain drain", as shown in the examples below:

"(…) βοηθά να αποφεύγεται η "άγρα ασύλου" (asylum shopping) σε άλλα κράτη μέλη (…)" [Eurodac][5]
Back translation: [this measure] helps to avoid asylum shopping in other states (...)

"(…)την τελευταία τριετία και γίνεται συχνά λόγος περί διαρροής ταλέντων (brain drain)"[6]
Back translation: (...) these last three years we often speak about brain drain.

5.4 Reliability and Relevance of Resources

Finally, examples listed in this section are related to resources used as reference for either definitions or contexts. Selecting a reliable source of information is perhaps one of the most important tasks in terminology management. Resources found may either be irrelevant, as in the case of the term "migratory patterns" for which the definition proposed from one of the students referred to Caribbean legislation; or insufficient, as in the context chosen for the term "chain migration".[7]

6. Conclusions

As stated in the beginning of this paper, our aim was to investigate the challenges that one can face when dealing with terminological resources, in the context of a terminology project. To do so, we proceeded to a full presentation of our project (participants, domain, workflow and deliverables) and we then tried to approach the concept of terminology management from different theoretical angles. Finally, specific examples taken from our corpus were analyzed in order to highlight different issues that emerged while working on this project.

Some of the conclusions that can be drawn from this experience are as follows:

- Descriptive Terminology Management constitutes a coherent theoretical framework that can be used in terminology projects, in a university context.
- Experts have a critical role to play from the very beginning of the project and should not only be assigned the role of checker at the end of the project.
- Despite being a time-consuming activity, terminology projects are very important for students, due to the real-life experience they offer.

Notes

[1] *Cross references* are called links to other IATE entities, they can be broader narrower, related, or antonym.
[2] *Problem language* is an anchor to which all the other languages in the entry will be attached. Most of the times, this language is English or French.
[3] *Ad hoc corpora*, also known as *Do-It-Yourself*, or *Custom Made* are collections prepared by a particular teacher or translator to address specific needs of a teaching/translating context, compiled by spotting and retrieving relevant texts either on the Web or locally (http://www.igi-global.com/dictionary/custom-madedo-it-yourselfad-hoc-corpus/37833, last accessed on June 2016.)
[4] http://www.europarl.europa.eu/aboutparliament/en/20150201PVL00015/Human-rights. Last accessed on June 2016.
[5] https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Supervision/Eurodac/09-06-24_Eurodac_report2_summary_EL.pdf
[6] http://crisisobs.gr/2013/10/anastasia-papakonstantinou-figi-neon-epistimonon-sto-exoteriko-brain-drain-brain-gain/ Last accessed on June 2016.
[7] For the totality of text cf.
http://news.kathimerini.gr/4dcgi/_w_articles_civ_3_17/04/2011_439059
Last accessed on June 2016.

References

Aubin, Sophie, and Thierry Hamon. 2006. "Improving Term Extraction with Terminological Resources. In *FinTAL 2006, LNAI 4139*, edited by Tapio Salakoski et al., 380–387.

Bernier-Colborne, Gabriel. 2012. "Defining a Gold Standard for the Evaluation of Term Extractors". In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. 15-18. Accessed June 2016.
http://www.lrec-conf.org/proceedings/lrec2012/index.html.

Best Practice for Terminologists. 2008. Accessed June 2016.
https://iate.cdt.europa.eu/iatenew/help/best_practice.html.

Cabré, Maria Teresa. 1998. *Terminology. Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.

IATE Input Manual. 2008. Accessed June 2016.
https://iate.cdt.europa.eu/iatenew/help/input.html.

L'Homme, Marie-Claude. 2004. *La terminologie: principes et techniques*. Montréal: Les Presses de l'Université de Montréal.

Riggs, Fred W., Matti Mälkiä and Gerhard Budin. 1997. "Descriptive Terminology in Social Sciences". In *Handbook of Terminology Management*, edited by Sue Ellen Wright and Gerhard Budin, 184-196. Amsterdam/Philadelphia: John Benjamins.

Wright, Sue Ellen and Gerhard Budin. 1997. *Handbook of Terminology*. Vol. 1. Amsterdam/Philadelphia: John Benjamins.

Wright, Sue Ellen. 1997. "Term Selection: The Initial Phase of Terminology Management" In *Handbook of Terminology Management*, edited by Sue Ellen Wright and Gerhard Budin, 13-23. Amsterdam/Philadelphia: John Benjamins.