University Eötvös Loránd
Faculty of Humanities

PHD DISSERTATION THESES

THE EUROTERM

THE CHARACTERISTICS OF THE EUROPEAN UNION TERMINOLOGY, FOCUSING ESPECIALLY ON ITS PARTICULARITIES IN THE ROMANCE LANGUAGES

ÉVA RÁDAI-KOVÁCS

2009

Introduction

Terminology plays a vital role in the administration of the European Union, where the great quantity of documents must be translated into all the official languages of the organisation. This dissertation focuses on the analysis of the EU terminology, trying to answer the following questions: how the EU subject field can be delimited, whether there is an EU specialised language and how it can be characterised, what connection can be identified between Eurojargon and Euroterminology as well as how the euroterms can be linguistically described from a conceptual and formal point of view. Furthermore, we found it important to be able to determine how the euroterms could be clearly distinguished from other lexical elements occurring in EU texts.

The analysis of the terminology of the EU specialised language (Eurolanguage) is carried out in three Romance languages (French, Italian, Spanish), in English and in Hungarian, using linguistic methods. We are dealing with terminology, the EU language and the linguistic analysis in a common framework.

An attempt was made to elaborate a comprehensive methodology for the general linguistic analysis of the terms, which is based on the formal side of the term. The linguistic form was the basis for the lexical, syntactic, morphologic and semantic analysis as well as for the conceptual analysis.

The description of the linguistic characteristics of the terms might be helpful in the automated term extraction on the one hand, in term formation, in terminology management and in the appropriate insertion of the terminological items in an existing linguistic and terminological system on the other hand. They might contribute to the compilation of an EU dictionary or a multilingual online database and they might help in the classification systems to determine which lexical elements can be considered as belonging to the EU subject field.

Subject field, specialised language and terminology

Subject field was defined as a domain of human knowledge or activities, built upon a central organising element, which is delimited from a given point of view, in a given time, with a clearly determined goal in mind.

Specialised communication is realised with the help of specialised languages which can be defined as linguistic tools used for specialised communication between experts, in specialised contexts for the transfer of specialised knowledge. The choice of the linguistic tools is determined by the form of the communication and the level of the specialisation. Within specialised language, oral and written communication as well as official and non-official specialised language usage can be distinguished. In our classification the specialised texts are seen as the main tool of the written form of the official specialised communication, while jargon was considered as the primary form of the oral non-official specialised communication.

Terminology can be interpreted in three different ways: glossary of terms, terminology science and practice. As term is defined as a linguistic unit that designates a

concept belonging to a given subject field, it was concluded that the formal side could be examined with linguistic methods.

The language policy of the EU (the tower of Babel)

The language policy is mainly characterised by an admissive approach: the Member States have the right to define their language policy. However, at the institutional level this policy is governed by the principle of multilingualism.

The translation policy is based on the principle that all documents meant for the public must be translated into all official EU languages. The relation between translation and terminology is of utmost importance in the case of euroterminology, since the official standardised expressions are shaped and developed during the translation process, which is closely linked to the legislative procedures. The parallel texts must not be considered politically as translations, they are seen as linguistic versions. The equality between the versions is guaranteed by an explicit functional and legal equivalence.

The EU subject field is primarily a domain of activity in the centre of which stays the EU as a political, administrative and legal entity and which consists of concepts related to the founding, history and daily functioning of the organisation. When locating the EU among the other specialised fields, the EU appeared as an independent subject field in certain classification systems, e.g. in EUROVOC and Lenoch. The subject field itself could be divided in six subdomains as follows: basic principles, European integration, institutional system, legal framework, budget and policies.

The EU specialised language and Eurojargon

We characterised the EU specialised language as a language parallelly present in the different languages, which forms part of the organisational identity of the EU and plays an important role in the legal unity. Since we focused on the EU specialised texts, we dealt exclusively with official written communication and in this case we talked about Eurolect or Eurolanguage. We reserved the word Eurojargon for the non-official, oral – primarily internal – specialised communication whose main characteristic is the presence of English and/or French calques and borrowings.

The specialised texts generally have specific pragmatic, syntactic, morphologic, lexical and semantic particularities that can also be identified in the case of the EU texts, so we concluded that the EU language has a raison d'être among the specialised languages.

Terminology in the European Union

The terminology appearing in the EU texts is two-folded: on the one hand terminology can be divided into euroterms designating EU concepts in strictu sensu and into other types of terminological units; on the other hand there is a frame containing legal and administrative terms and there is a core with specialised terms that assure the central content

of the texts. We use the word euroterm to designate those legal, administrative and specialised lexical items that describe concepts closely pertaining to the EU and have a clear definition in this subject field.

The standardisation of the terms never takes place through official channels, the euroterms, however, introduce a normative usage by their conceptual and formal definitions in the legislative acts.

The euroterms are conceptually formed at a political/diplomatic level and the primary term formation takes place in one language which is English at a greater extent and French at a lesser extent. The secondary term formation happens in the other languages during the translation process involving numerous actors and institutions. In order to examine the life cycle of the term the concepts of a directive were analysed in five languages, in four different phases of the legislative process. We observed that the form and the content of the terms was less modified than the wording of the definitions. The more frequent modification of the wording of the definitions points out the term-formation role of the translation process: at the different legislative phases translators redraft and change the different text formulations.

The handling of the lexical items in the EU translations

The linguistic approach of terminology is also required by translation oriented aspects: in the texts, translators first meet the structures composed of lexical items and through these they can reach the underlying concepts. Thus, terminology management is based on texts and linguistic forms.

It was argued that every problem to be solved during the translation is a translation problem which we labelled as a translation unit. In order to identify more precisely the euroterms in the texts it was important to distinguish them clearly from the other lexical translation units. To do so we extracted all the translation units from an EU directive and classified them into six different groups: free translation units, specialised translation units, ad hoc translation units, fixed translation units, borrowed terms and euroterms. The distinction between the groups was based on three criteria: whether they belong more to the LGP or to the LSP; at which extent their translation is fixed and at which part of the EU text they belong. According to this classification we stated that euroterms were highly fixed LSP items that could be located at the border of the core and the frame of the EU texts.

Identification of terms in a corpus

Contrary to the earlier practice based on a conceptual term collection, we chose the corpus based method stating that it is the linguistic form which is mainly at our disposal and only after its identification it becomes possible to explore the conceptual system. We identified 985 concepts in a representative corpus containing in each language ca. 350 000 words. Our starting point was that the framework of the EU, the legal basis of its functioning, its principles, institutions and competences are laid down in the founding treaties, so in the first place we extracted the EU terms form these treaties. We were looking for the noun

phrases designating EU concepts in the English language documents, since the majority of euroterms are currently born in this language. Afterwards the French, Italian, Spanish and Hungarian equivalents were collected from the parallel texts. In order to insure representativeness 35 other EU texts were also terminologically analysed which cover the principal activities of the organisation as well as the most frequent document types of the different institutions.

Despite the fact that it is normally not advisable to use translations in a corpus we opted for the inclusion of the different language versions since the formation of the EU terminology is translation oriented. In the term extraction and management our approach was descriptive, which implied that the lexical items were entered into our database in the graphic form of their first appearance. We collected furthermore the term variants belonging to the same language. For the term extraction we chose the manual method arguing that it made it possible to filter the complete terminological forms and their variants with great accuracy.

The terms belonging to one of the EU subdomains were entered into a glossary together with the following data necessary for the linguistic analysis: linguistic forms, definition of the concept and its source and, as the reference, the first appearance of the term.

The linguistic analysis of the terms

The linguistic analysis was based on the two-foldedness of the term, so we took into consideration that terms had a conceptual and a linguistic side, and the linguistic form itself was composed of two parts: a base and a modificator. We only dealt with linguistic characteristics that might be relevant for terminology and translation, for terminology management and for term identification. Accordingly, the formal side was examined from a morphologic, syntactic, lexical and semantic point of view, as for the conceptual side the central element of the term structure (the base) was identified.

In the linguistic examination a distinction was made between the analysis of the complete terminological units (term forms, semantic and syntactic analysis) and the analysis of the individual lexical items (lexical units, word classes, word formation mechanisms). The analysis of the complete terminological units took place with the help of syntax, semantics and terminology, which was accompanied by the analysis of the individual lexical constituents by means of morphology and lexicology.

Each linguistic analysis started with the classification of the terminological units according to their structures, based on which we differentiated between simple, complex and short form terms. The collection of the equivalents and the term variants made it possible to examine terminological synonymy, i.e. how many concepts are designated with more than one term. The study of the relations between the term variants lead us to the distinction of four major (and 12 minor) groups: the variants were formed by short forms, lexical modification, structural modification and borrowing.

The comparison of the terminological forms pointed out the presence of polysemy. This phenomenon is an important issue in the translation-oriented terminology since the

starting point is the linguistic form appearing in a text; it is thus essential to determine which meaning of the term we are dealing with.

As we saw every term as a structure composed of a base and a modificator we could distinguish four different groups according to the linguistic forms of the terminological units: complex terms, compounds, simple terms and short forms (acronyms and abbreviations).

Within the framework of the syntactic analysis we examined how many lexical units the terms were composed of, and we described the most frequent structures.

From a lexical point of view the number of the terminological entries (concepts), the number of the terms, the total number of the lexical units and the number of the individual units were determined. The first 15 most frequent nouns and the first 10 most frequent adjectives were identified. In the case of the individual lexical items a word class classification was carried out.

In the morphological part, firstly the affixes were identified: the nominal and adjectival prefixes and suffixes occurring in the individual lexical items, and then the compounds were analysed from a structural point of view.

The conceptual classification

In the conceptual analysis not only was it important to identify the base, but it was also indispensable to describe precisely the underlying concept, so the examination was carried out in three steps: definition of the concept, grouping of all the concepts into 20 basic conceptual categories and the representation of the conceptual system.

The basic categories made it possible to determine the main concepts behind the euroterms, which can be defined as designations of legal acts, regulatory frameworks, principles, symbols, policies, persons, fora, bodies, procedures, processes, behaviours, instruments, programmes, systems, actions, places, documents, statuses, relations and goals. It could be stated that the EU conceptual system was related to a programme-oriented administrative and legal entity that strives to achieve general goals with unique instruments and persons, by specific actions, in a unique place. The uniqueness is highly reflected at terminological level: 28% of all terms designate individual concepts.

The linguistic characteristics of the euroterms

1/5 of all euroterms (18%) have more than one terminological designation, which is most frequently an acronym. Polysemy is present in all the languages, but in a lesser extent than synonymy: the number of polysemous elements is between 6 and 9.

The English and Italian term structure contains on average three, the French and Spanish three or four, the Hungarian two or three lexical items. The most frequent term structure is composed of two items in all languages. The most typical syntactic structure in English and Hungarian *Adj N*, in the Romance languages *N Adj*. The characteristic three-element terminological unit in English and in Hungarian is the *Adj Adj N* noun phrase, while in French, Italian and Spanish *N Adj Adj* and *N Prep N*.

An individual lexical item occurs on average 3.5 times in English, 4 times in French and Italian, 4.5 times in Spanish and 3 times in Hungarian. The most frequent noun is *Community* in English*, politique*, *politica, política* in the respective Romance languages and *bizottság* in Hungarian. The adjective that appeared the most was identical in all languages: *European, européen, europeo, Europeo, európai.* Overall, it can be stated that the most frequent lexical items relate to the EU, the bodies, the documents, the comprehensive and unique character. As for the word classes nouns form 2/3 of the lexical items (68%) and adjectives 1/4 of them (26%). These two grammatical categories can thus be considered as the main constituents of the euroterms.

Most terminological units are complex terms: in English and the Romance languages 85%, in Hungarian 78% are complex structures. More than 12% of the terms, i.e. every eighth euroterm is an acronym. The most frequent nominal compound in English, French, Italian and Hungarian is *N+N,* in Spanish *V+N.* The most typical adjectival compound in English has the *Adj+N* structure, while in Hungarian the *N+Adj* structure. The Romance languages did not generally have adjectival compounds as constituents of the terminological units.

In terms of morphology, except for Hungarian, the most frequent nominal and adjectival suffixes and the nominal prefixes in all other languages corresponded to each other. In the order of English, French, Italian and Spanish, the most typical nominal prefixes were: *Re+X; Re (Ré, R)+X; Ri (Ra)+X; Re+X;* the most typical nominal suffixes were: *X+ion; X+ion; X+ione; X+ión;* the most typical adjectival suffixes were: *X+al; X+al, X+el* and *X+é; X+ale; X+al.* In Hungarian the nominal prefix that occurred mostly was *Elő+X* and *Szak+X,* the most frequent adjectival suffix was *X+i,* and the most typical nominal suffix was *X+ás(és).* The adjectival prefixes showed the greatest variability in the different languages: the most typical elements were *Non+X* in English, *Inter+X* in French, *Co(Com, Con)+X* in Italian, *A(Ad)+X* and *Re+X* in Spanish and *Köz+X* in Hungarian.

Conclusion

This analysis can be seen as a first step in a more comprehensive compilation and description of euroterminology. The results might be used as a basis for the collection of the EU core terminology. To this end, based on the identified linguistic characteristics a greater corpus (in our opinion 2 000 000 words) could be used for an automated term extraction. This operation would also be a test of the utility of the results in the electronic applications. For such an extraction all the actual and potential graphic variants should also be taken into consideration. The term candidates should then be examined one by one in terms of frequency and appropriateness, and only the terminologically most accepted variants should be selected. Based on the data coming from this selection, a multilingual EU dictionary or rather an online EU terminological database could be compiled.

Related publications

A terminológia szerepe és alkalmazása a fordítói munkában, In: Fordítók és Tolmácsok Őszi Konferenciája, Előadások szövege, Budapest, 2003. (társsz.: Erdős J.)

Terminológiakurzus a fordítóképzésben, In: Porta Lingua, Szaknyelvoktatásunk az EU kapujában, Debrecen, 2004.

Az EPSO fordító-versenyvizsga tapasztalatai, In: Fordítók és Tolmácsok Őszi Konferenciája, Előadások szövege, Budapest, 2004.

Az EU-ban dolgozó fordítók internetes és számítógépes segédeszközei, terminológiai problémák számítógépes megoldása, In: Fordítók és Tolmácsok Őszi Konferenciája, Előadások szövege, Budapest, 2006.

Transzliterációs és toldalékolási problémák a fordításban, In: Fordítók és Tolmácsok Őszi Konferenciája, Előadások szövege, Budapest, 2007.