Doctoral School of Philological Studies
Faculty of Letters
“Alexandru Ioan Cuza” University of Iași

Abstract

Terminology and Translation of General Use
ICT Texts: A Romanian Perspective

Doctoral student: Sorina CIOBANU (married POSTOLEA)
Doctoral supervisor: Professor Rodica DIMITRIU

2014
This thesis puts forth a detailed, both descriptive-qualitative and quantitative analysis of the main norms that govern the translation of general use information and communication technology (ICT) texts from English into Romanian. In Romania, the language of ICT and its relationship with English has been discussed by various scholars. However, most of the studies published up to now have focussed on only one product of this relationship, i.e. terminology, leaving many other aspects untapped, e.g. the phraseology, the syntax, or the general translation strategies used in the texts produced in this field. Thus, one of the novelties brought about by this thesis is that it proposes an integrated analysis of the Romanian general use ICT discourse, exploring the manner in which terminological and phraseological units, sentences, as well as texts, as discourse macro-units, are usually transferred from English, the source-language, into Romanian.

Taking into account that, as a rule, English is the source language of ICT communication and translation the main vehicle through which the new linguistic entities produced in this sector are introduced into Romanian, this research starts from the general premise that, at all levels (terminology, phraseology, morpho-syntaxis, etc.), Romanian ICT discourse is based on the imitation of English ICT discourse. This general hypothesis is tested through three specific premises: 1) at the terminological level, borrowing from English is the main pattern used in the creation of new terms in the Romanian language of ICT; 2) at the phraseological and syntax level, direct/literal strategies prevail in the translation of English ICT discourse; 3) at the textual macro-level, the translation of general use ICT texts is mainly based on the imitation of the specific functions and structures of English source-texts.

In order to tests these working hypotheses, this research is based on an up-to-date methodology, widely used in applied linguistics studies abroad, but almost absent in Romania: corpus-based analysis. In this respect, an important contribution brought by this thesis consists in the collection, processing, and analysis of a corpus of electronic texts created specifically in order to test our initial hypotheses. More precisely, this study draws on a parallel corpus formed by 275 pairs of English source-texts and Romanian target-texts, or, otherwise put, 275 original texts and their translations. The 550 texts comprise a total number of 563,601 words (about 1500 standard A4 pages) and belong to four textual genres frequently used in general use ICT communication: ICT news articles, ICT press releases, ICT product descriptions, and ICT user manuals. Compiling this corpus was in itself a consistent research effort: first, we identified
the Romanian texts that seemed to be translations from English; then, through successive searches on the Internet, we identified their English sources.

The innovating nature of this thesis does not reside only in its use of an electronic corpus created by the author and analysed drawing on the most recent theories of corpus linguistics (Stubbs, 2002; Tognini Bonelli, 2010; Reppen, 2010; McEnery & Hardie, 2012), but also in its use of three corpus-analysis tools. More precisely, the working hypotheses set down in this thesis are tested with three different programs: ParaConc, a parallel concordancer specifically created to allow for parallel corpora to be studied in a contrastive manner; UAM CorpusTool, a program which allows for corpora to be manually annotated in order to highlight and analyse relevant textual aspects; AntConc, an electronic concordancer created for the study of monolingual corpora. In order to use these electronic tools, all the 550 texts included in the corpus were submitted to a lengthy process of textual processing and alignment. Even though abroad, in countries such as Great Britain, the United States of America, France or Spain, corpus-analysis tools are widely used in applied linguistics and even in literary studies, to my our knowledge, Romanian research in this field does not benefit from a corpus-analysis tool specifically designed to be used with Romanian texts, a language with particular features. From this viewpoint, this research is one of the first attempts to use international electronic text-analysis tools on a corpus which includes English texts as well as Romanian texts; some of the difficulties encountered during this process are described in the thesis.

Thus, starting from the newest concepts from the fields of text and corpus-based linguistics, this research aims: a) **to identify the main translation patterns** used in the interlingual transfer of English ICT discourse, starting from the basic level of terms, going through the level of phraseological, grammatical, and syntactic units, in order to reach, in the final chapter, the level of texts as translation macro-units. In turn, this process triggers the need b) **to create an interdisciplinary methodology in terminological and translational studies**, able to combine and apply, within the numerous analyses carried out in this thesis, various notions and theories from the fields of corpus linguistics, corpus semantics, applied and contrastive linguistics, genre and register analysis, as well as elements of pragmatics – which is what we attempted to do.

Moreover, c) the numerous patterns revealed with the help of corpus analysis tools are used as starting points 1) **to theorise and generalise the norms** usually applied in the translation
of the linguistic units analysed, in the field at hand, b) **to highlight potential sources of problems**, both for the translators who work in this field and, in general, for the Romanian translators who work with English source-texts, and c) **to critically review some traditional concepts and notions** used in the literature (the couple *general language – general translation*, *direct/literal* strategies and the concept of *literalness* in translation etc.).

Drawing on the analyses carried out throughout, the thesis also puts forth d) **a series of new theoretical notions**; each of them could be developed and furthered in subsequent studies: *general use communication*, a comprehensive typology of Romanian ICT terminology, a classification of ICT loans, the category of *pseudo-calque* in the translation of words and terms, *verbatim translation* as a pattern in the transfer of sentences, as well as a framework for the analysis of translation at the macro-textual level, based on the identification of genre and register features of the source-texts and of the manner in which they were rendered/transfered in the target-texts.

*Chapter I* places our research within a broader theoretical framework. Thus, the first subchapter defines the main features of translation in the field of ICT as a type of specialised translation, starting from one of the basic dichotomies defined in the literature, i.e. the distinction between *general language* and *specialised language*. The relationship between these concepts and the criteria used to differentiate them in the literature are critically revised, whereas the language specific to the field of ICT is defined as a specialised language based on its specific features. The first part of this chapter continues with a presentation of the main branches involved in the study of specialised languages and translation; it reveals a series of problematic areas in the theorisation of several concepts, e.g. the distinction between *general translation* and *specialised translation*, highlighting the latter’s modest status within the discipline. The second subchapter discusses and defines the main **features of communication in the field of ICT**. Starting from a series of recent statistics regarding the magnitude of the informational phenomenon and new technologies, the analysis highlights the discourse communities active in the field of ICT, the main information vectors ad the main types of texts produced in this sector. The features of general use ICT communication are defined and analysed in relationship to the four textual genres included in the corpus. The chapter ends with a presentation of the way in which the tools provided by the field of ICT have influenced terminology and translation studies.

The methodological principles and the materials used in our research are described in
detail in Chapter II. The first part of this chapter is devoted to a discussion and a presentation of
the main notions used in the field of corpus linguistics, which also represent the basic
methodological framework of the investigations conducted in this thesis. Our analysis starts with
a presentation of the concept of corpus as defined in the latest works published in this field and
continues with a classification of the main types of corpora currently used in applied linguistics.
Modern analysis tools and the main notions that define this analysis are described in the next
section. Thus, at this stage, with the help of actual examples, a series of key-terms used
throughout the research are explained in detail (type, token, type-token ratio, keyword etc.) and
the various ways in which concordancers and corpus annotation tools can be used in language
analysis are thoroughly described.

The second part of the chapter describes at length the corpus-based methodology used in
our research. In this section, readers are offered details about the selection criteria and the
methodology used to collect textual materials, about the three computer programs used in our
analyses, as well as a detailed description of the overall corpus structure and of each of its sub-
corpora (one parallel sub-corpus for each textual genre).

The next three chapters are devoted to the identification and the analysis of the main
translation patterns used in the texts included in our corpus. Focussing on terminology, Chapter
III pursues two major goals. On the one hand, the analyses carried out in this chapter aim to
provide a comprehensive terminology of Romanian ICT terms; on the other hand, another goal is
to test the first specific hypothesis mentioned initially in the thesis and in this abstract, according
to which borrowing from English is the main strategy and source in the creation of Romanian
ICT terms. In order to provide a unitary theoretical framework, the chapter starts with a critical
analysis of the definitions given to the concept of term in the literature (Pearson, 1998; Cabré,
1999; Temermann, 2000; L’Homme, 2005; etc.), then it continues by defining primary and
secondary term formation processes (Sager, 1990; ISO 704:2009(E); etc.) and by discussing the
main translation theories and categories which focussed on words as minimal translation units

The second chapter provides a detailed analysis, backed by statistical data, of the main
types of terms used in the Romanian language of ICT. Thus, the terminological units present in
our corpus are classified according to three different criteria. From the viewpoint of their form
and structure, Romanian ICT terms are further divided into plain terms, derived terms (formed
with prefixes, suffixes, and parasynthetic terms), compound terms (with prefixoids), multi-word terms (made up of at least two word), initialisms, acronyms, abbreviations and clippings. Each of these term-types is analysed at length, with numerous examples retrieved from the corpus. According to the second criterion of classification, the function performed in texts, Romanian ICT terms are divided into nouns, adjectives and adverbs, and verbs. In this section, the top 100 most frequent terms in our corpus are analysed according to their relative distribution. The results of this study show, on the one hand, the important proportion of initialisms and proper names in the corpus and, on the other hand, the fact that nominal terms (common nouns, proper nouns, and initialisms) represent no less than 70% of the lexical content of the texts at hand. The last part of this chapter offers a typology of Romanian ICT terms according to their semantic content.

The first specific hypothesis of our research is only partly confirmed by the analyses carried out in the last part of Chapter III. In this subchapter, Romanian ICT terms are analysed from the viewpoint of the patterns used in their translation from English into Romanian. Thus, the 171 borrowings identified within the top 1000 most frequent terms in the corpus are classified into full loans, acclimatised loans, and assimilated loans. Further on, our analysis takes into account multi-word loans and hybrid terms (made up of a Romanian and an English word). The pattern of calque, another translation technique included by theorists into the group of direct procedures, is also analysed in detail. Our research takes into account both one-word calques, focussing on analogous and homophonous terms, and multi-word calques. At this stage of our research, we advance a new category to be included into the group of direct translation procedures, i.e. pseudo-calque, defined as an intermediate stage between calque and adaptive translation patterns; the latter are analysed in the last part of this chapter.

The actual data provided by the corpus at hand show that in the Romanian language of ICT only initialisms and proper names are borrowed consistently. In what regards the other types of terms, the decision to loan seems to be motivated 1) by the morpho-syntactic features of the source-term (its adaptability to the structures of the Romanian language) and 2) the perceived level of specialisation of the source-term (only the terms that display a high degree of specialisation within the field are borrowed consistently). Corpus analysis also shows that from a quantitative viewpoint, analogy-driven translation, through calque and pseudo-calque is the prevailing pattern in the transfer of ICT terminological units.
In Chapter IV, our investigations take into account a higher level of analysis, focusing on the main translation patterns used in the case of phraseological units, grammatical structures, and sentences. Drawing on a series of recent studies in this field (Martin, 2008; Gries, 2008; Nascisione, 2010; etc.), the chapter starts by defining the notion of phraseological unit (phraseme) in relation with two adjacent concepts, term and collocation. The next section puts forth a typology of the most frequent non-ICT specific phraseological units found in the corpus and analyses the main patterns used in their translation. The research continues by defining the notion of specialised phraseological unit (Aguado de Cea, 2007), studying the translation patterns used in the interlingual transfer of the most frequent units of this kind in our corpus; in turn, the latter are classified based on morpho-syntactic criteria. The analysis carried out at this level confirms the hypothesis according to which direct/literal translation patterns prevail in the translation of these units. Borrowing is much less frequently used in the case of ICT phrasemes whereas calque and pseudo-calque are prevalent. The use of these procedures often has a negative impact on the naturalness of the issuing target texts, being a source of clumsy and atypical expressions in Romanian. Moreover, the existence of several competing translations for a single source-phraseme leads to a dissipation of phraseological ranges in the target texts.

Although obligatory in nature, the choices made in the translation of source grammatical structures are often accompanied by added nuances in the target-texts. An example in this respect is the translation of English imperatives and 2nd person verbal forms. In our corpus, as a rule, these two English grammatical forms are translated by 2nd person plural forms, and this choice is accompanied by a “specification” of register, which from neutral in English becomes polite and formal in Romanian. Our analysis also shows that, at the level of grammatical structures, English “-ing” forms engender various translation patterns in their transfer into Romanian and may become a source of problems for translators.

Chapter IV also reveals the main patterns used in the transfer of sentences from the source-texts into the target-texts. At this level, verbatim translation, which sticks closely to the structure of the source-segment, sometimes at the expense of clarity and naturalness in the target-language, may be seen as a norm, confirming the hypothesis set down initially. Of a total number of 20336 sentences in the source-texts, only 229 (~1.1%) were subject to structural changes in the course of translation; the rest were reproduced on a one-to-one basis in the target-
texts. This faithfulness to the structure of source-sentences is often a source of errors, and the thesis provides numerous examples in this respect. On the other hand, the use of compressing patterns, in which two or more source-sentences are compressed into just one target-sentence, seems to be motivated by generic and stylistic criteria. The same conclusion holds in the case of expansive patterns, in which a single source-sentence is expanded into two or more target-sentences.

The final chapter analyses the translation of the 275 source-texts included in the corpus from a broader perspective, envisaging them as discourse/communicative macro-units. At this level, our research starts by defining the main concepts used in modern discourse analysis, dividing them into two categories: surface textual features and inner textual features. Drawing on several recent works in this field (Bex, 2001; Bhatia, 2004; Biber & Conrad, 2009; etc.), a series of key-notions, such as context, purpose, register, and genre are analysed within the former category; the latter category takes into account other defining textual features: subregister, generic/register features and markers, cohesion, coherence. All these concepts are then used in the analysis of the four textual genres included in the corpus. The chapter continues with a presentation of the two main translation studies trends which focussed on text as a primary unit of analysis, i.e. the textual approach (Neubert & Shreve, 1992; Trosborg, 1997) and the functionalist approach (Reiss, [1971]2004; Nord, 1997a, 1997b). After having discussed two translation assessment models put forth in the literature and based on the notions previously discussed, we put forth a unitary framework for the analysis of the translations included in the corpus. This model is mainly based on the identification of the generic/register features of the source-texts, followed by an analysis of the manner in which they were rendered/transferred in the target-texts.

In the second part of Chapter V this translation-assessment model is used in order to study the translation patterns that prevail at the level of each genre. Thus, in the case of ICT news articles, these patterns differ according to the type of source-text. According to our analyses, although they were not marked as such when they were published in the Romanian ICT magazines from which they were retrieved, the 140 Romanian ICT news included in the corpus are either translation of source press releases or translations of foreign ICT news articles. The investigations conducted in the last chapter show that the translation of English press releases as Romanian ICT news articles is usually accompanied by the omission of the qualifying
lexis (adjectives and adverbs) used in the source-texts. The prevailing pattern used in the translation of English ICT news articles as Romanian ICT news articles is reduction in size.

Because they have the same producer (ICT companies) and pursue the same communicative goals in the source and the target context, the two promotional genres included in the corpus show similar translation patterns. In their case, one may talk about the transfer of source-text features from the source into the target-language, with slight adjustments (e.g. the substitution of time and place deixis elements) in order to better integrate the translation into the target-context of reception. The translation of ICT user manuals shows the highest degree of faithfulness to the structures of source-texts; the changes made in the translation of these texts are minimal. The analyses carried out in this chapter confirm only in part the hypothesis set down initially: considered in bulk, the translation patterns that prevail in the translation of the four textual genres included in the corpus are situated along a cline, which spans from the highest degree of imitation in the case of user manuals to the highest degree of adaptation, in the case of source press releases translated as Romanian ICT news articles.

Although exploratory in many respects, we hope that the numerous analyses carried out in this thesis are just as many proofs which show the potential and usefulness of corpus-based analysis and modern corpus-analysis tools in terminological, translation, and linguistic research. Each of the investigations conducted in this thesis may be further developed in subsequent studies and many other aspects remain to be found and analysed in our corpus. Moreover, the collection of texts compiled for this research may be used to create a bilingual ICT glossary. Finally, the methodology put forth in this thesis may serve as a model for other similar studies in the fields of terminology and specialised languages.

Bibliography


Borin, L. (2002). ... and never the twain shall meet? In L. Borin (Ed.), *Parallel corpora, parallel worlds: Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April, 1999* (pp. 1-47). Rodopi.


de Beaugrande, R. (1987, December). Special Purpose Language and Linguistic Theory. UNESCO Alsed LSP Newsletter, 10 (2 (25)).


Fiola, M. (2004). The ultimate goal of university translation curricula: training specialized translators or translation specialists? In K. Ahmad, & M. Rogers (Ed.), Proceedings of the
14th European Symposium on Language for Special Purposes - 2003 (pp. 431-436). Surrey: University of Surrey.


**Electronic references**

[18] http://w3techs.com/technologies/overview/content_language/all [accessed on February 3rd 2014]