A Multidisciplinary Approach to Specialized Writing and Translation Using a Genre Based Multilingual Corpus of Specialized Texts

Isabel García Izquierdo
& Anabel Borja Albi
GENTT Research Group, Universitat Jaume I [1]
Spain

1. Introduction

One of the skills that is most evidently required to be a good professional linguistic mediator (writer/translator) is textual and communicative competence, which includes the ability to understand and produce texts that conform to the conventions of the specialized languages being working with. This skill is particularly significant in the case of specialized writing or language for specific purposes where the fact that the translator/writer has to work with texts that require a much deeper conceptual and contextual knowledge determines the way in which textual competence should be acquired and applied.
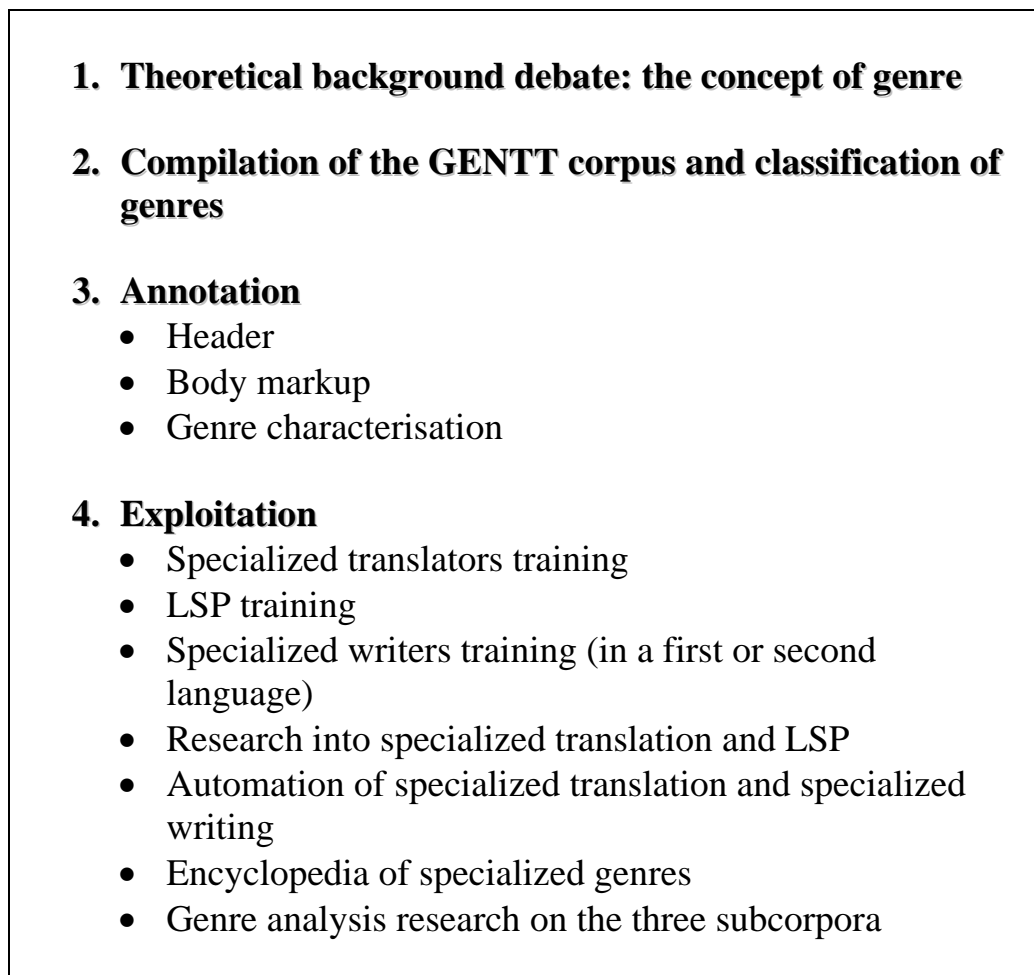
One of the main objectives of the GENTT Research Group (Textual Genres for Translation, *GÉNeros Textuales para la Traducción* in Spanish) is to provide translators and writers of professional texts with text models and patterns that they can use for textual, conceptual, linguistic and terminological reference. The corpus design also intends to create a knowledge management system, similar to terminological knowledge representation systems but structured around the notion of genre, both for translators and writers of professional texts (Borja, 2005; Dam - et alii, 2005).

The GENTT project team works on the multilingual study of genres in professional legal, medical and technical contexts to identify genre conventions (lexicon, syntax, style, rhetorical structure and devices, etc) by means of a hypothetic deductive approach using corpus technologies (García Izquierdo, ed., 2005). The project started in 2000 and has focused on the compilation of a multilingual (Catalan, English, German, Spanish and French) annotated corpus of specialized discourse texts for research and teaching purposes.

In this context, this contribution presents the evolution of the project from the initial discussion of its theoretical background to the compilation and annotation phases, and puts forward some ideas for its future exploitation in the fields of specialized translation and LSP training.

---

1. **Theoretical background debate: the concept of genre**

2. **Compilation of the GENTT corpus and classification of genres**

3. **Annotation**
    - Header
    - Body markup
    - Genre characterisation

4. **Exploitation**
    - Specialized translators training
    - LSP training
    - Specialized writers training (in a first or second language)
    - Research into specialized translation and LSP
    - Automation of specialized translation and specialized writing
    - Encyclopedia of specialized genres
    - Genre analysis research on the three subcorpora

Figure 1. The GENTT Project Phases

2. Theoretical background

In order to be able to provide linguistic mediators (specialized writers and translators) with tools to facilitate their task we advocated the adoption of a

multidisciplinary approach, which would bring together the virtues of various ways of analyzing communicative transaction phenomena. Starting from this idea in 2000, the GENTT team set out to design a theoretical model of analysis for specialized texts that would combine linguistic considerations, socio-professional features and textual genre analysis, bringing together aspects of three disciplines that had separately demonstrated their usefulness in order to try to find points of intersection. And we found that point of intersection which could serve as a point of departure for our analysis precisely in the concept of *text genre*, used both in LSP and translation and which combines formal, communicative and cognitive elements, as we shall see.

One of the most active lines of current research in genre theory is its application to teaching languages and, in particular, teaching languages for specific purposes ('*genre-based approach to ESP*' and '*EAP genre-school*'). The works of Bazerman (1988, 1994, 1994a), Berkenkotter and Huckin (1995), Bhatia (1997, 2004), Dudley-Evans (1997), Engberg et alii (2007), Swales (1990, 1994) and Weissberg and Buker (1990), amongst others, fall into this category, underlining the importance of the fact that those learning a language need to understand specific communicative codes of the culture of the specialty and the structure of its genres in order to become expert users of the language in a particular field of knowledge.

But with the growing challenge of globalization (of legal relations, medical advances and technical progress) many referential texts require translating into conventionalized genres with very specific professional functions and uses (e.g., a *Merger Agreement*, a *Medical Research Article* or a *Patent Description & Claims*). Hence the relevance of researching specialized genres for translation purposes too. The notion of genre began to gain strength within the field of Translation Studies as a semiotic notion related to the intercultural nature of translation in the 70s and 80s (Bassnett and Lefevere, 1990). Yet, in our view, the Translation Studies approach that best represents the complexity of the phenomenon of translation is the so-called *Textual Approach*.

The GENTT research group proposes an eclectic and multidisciplinary concept of *text genre* that is based on systemic functional linguistics (García Izquierdo, 2000), the theory of genre applied to translation (Trosborg, 1997; Borja, 2000; Montalt and García Izquierdo, 2002) and the sociology of professions (Monzó, 2002). Thus, the research concentrates on studying the formal aspects of genre, and analyzing it as a linguistic transaction, but it also applies a sociological or socio-professional approach that attempts to develop the communicative aspect of the concept and to incorporate the needs and

insight of professionals who work with the genres under study. In addition to this, we have recently begun to give more thought to the cognitive side of genre and more particularly to its correlation with the concept of *translation competence* (Borja, 2005; Montalt, Ezpeleta and I. García Izquierdo, forthcoming).

In terms of research methodology we try to apply *methodological triangulation* combining different qualitative and quantitative methods and techniques **for obtaining genre information**. On the one hand, we use **quantitative studies** based essentially on statistical techniques. The GENTT team has incorporated the applications developed in the field of computer linguistics in recent years to construct a corpus management tool specifically designed with specialized writers and translators in mind (see García Izquierdo, 2005a; Borja, 2007b; or www.gentt.uji.es). Nevertheless, and as pointed out by Wilson (2000), quantitative methods sometimes focus on the amount of data accumulated without taking into account the theoretical and descriptive context on which they should be based. This is why we try to complement the process with **qualitative research**.

According to Strauss and Corbin (1998) qualitative research is "any type of research that produces findings not arrived at by statistical procedures or other means of quantification". In *genre analysis* it means collecting data about the context in which communicative events take place, that is to say, to discover the significant subjective aspects of the particular communicative situation under study by means of interviews, focus-groups, etc., with the participants. We work with medical practitioners, lawyers, engineers, etc., who have to write and read these texts in their daily work, and also with specialized writers such as journalists, and translators of legal, medical, technical documents.

From this rather eclectic perspective, the GENTT group understands the concept of **genre as a product**, i.e., as a successful achievement of a specific communicative purpose (Bhatia, 1993), which is characterized by its conventionality, restrictions (Swales, 1990) and typicality (Bazerman, 1999). Following Bekenkotter and Huckin (1994), we also understand genres as devices to package our speech and make it a recognizable response to the demands of the situation. And as the recognizable shape through which *participation* is enacted and understood (Bazerman, 1999).

But we also see **genre as a process**, a recognizable communicative event (Swales, 1990); a social construct that gradually emerges, evolves and disappears depending on communicative needs (Bathia, 2004); it is a window of opportunity for realizing certain kinds of *activities, meanings,* and

relations (Bazerman). In short, following Hatim and Mason, 1990, as 'conventionalized forms of texts' which reflect the functions and goals involved in particular social occasions as well as the purposes of the participants in them. For these authors, genre is conceived as a notion that includes formal aspects (*conventionalized forms*), sociocultural aspects (*social occasions*) and cognitive aspects (*purposes of the participants*). This idea of three dimensions is extremely important to be able to understand the complexity of genres in different languages and cultures (See also Schäffner, 2004).

Genre therefore becomes a culture-specific category that allows us to observe the different ways languages conceptualize reality (García Izquierdo, 2002). In short, as members of a cultural community, we are capable of recognizing that a certain text belongs to a genre (from a particular socio-professional domain) thanks to the features of prototypicality and recurrence, which are displayed in different micro- and macrostructural categories. The GENTT analysis of genres seeks to explain and systematize these features from a monolingual perspective and the differences (in terms of process and product) observed in genres that could be considered "equivalent" in different languages/cultures (a "Power of Attorney" and a *"Poder Notarial"*, for instance) from a multilingual approach.

Although we started our research using a semiotic static definition of genre, the results of our work show that genre is a dynamic concept, and both the intrinsic complexity and the variations in genres from one culture to another sometimes make it difficult to establish the limits of a genre. We have thus shifted from the traditional semiotic approach to a more dynamic definition of genre, in order to link process with product connecting the communicative, the social and the professional approaches (García Izquierdo, 2007). In this new definition, we would like to draw attention to four aspects that will help us understand the complex, dynamic and hybrid nature of genres:

2.1 Dynamism

The mapping of genres in professional domains needs to be updated to take into account constant changes in the social context and those involved in specialized writing and translation must be aware of these (structural and conceptual) changes. For example, in recent years quality policies have been imposed on all institutions' processes and we have observed that the genres related to quality standards are changing at a dramatic rate, reflecting the changing needs of society. The same is true for brokerage agreements in the field of finances for instance. So, in describing and characterizing a genre, in addition to generic integrity (which preserves its prototypical elements), it is essential to take into account the characteristics of flexibility, creativity and

innovation, together with lexical-grammatical patterns and discursive organization (Bhatia, 1999: 21).

2.2 Hybridation

Because the boundaries between some genres are fuzzy (less conventionalized genres such as a *letter of complaint* or an *Acknowledgement of receipt*) and some genres are difficult to ascribe to a professional domain (a *Will* is clearly a legal field genre, but a *Living will* could be considered both a medical genre and a legal genre) the specialized writer or translator has to know the context in which they occur in order to determine their ascription, the people they are addressed to and the purposes and functions they are meant to fulfill. In fact, even after doing this we find that some genres share features from different socio-professional fields or of different categories of genres.

2.3 Genre Systems

We have introduced the concept of *System of genres* (Bazerman, 1994) into our research. This notion refers to the existence of interdependent genres that appear as certain typical sequences and which form relations with one another and have interacting purposes and forms (Montalt, 2005; Montalt and González 2007, I. García Izquierdo, forthcoming). They are collections of genres that, despite having different characteristics, are related to each other in some way, either because they need each other or complement each other in a specific communicative context within a specific discursive community. In the field of maritime transport, for example, the *Charter Party* is part of a system of genres together with the *Maritime Insurance*, the *Bill of Lading*, the *Packaging List*, etc… When a vessel is chartered, either as a time charter or a voyage charter, all these genres will be interacting in a complex communicative situation. This concept is extremely useful, then, for recognizing the relations between genres present in the same process within an area of specialty, understanding the specific functions that each of them fulfils, and using a homogeneous terminology and style.

2.4 Transgenre

The concept of *transgenre* (Monzó, 2002: 251) is also useful. This is taken as referring to a genre that is exclusive to translation and which includes the three cultural, cognitive and discursive aspects of genres. This specific type of genre (a translated genre) displays differences with respect to analogous original texts in the source and target cultures and at the same time shares characteristics with them. The degree of similarity/differentiation will vary depending on the strategy and method of translation adopted by the translator, which may be more or less *foreignizing* or *domesticating* (Venuti,

1995). In our view, for translating purposes, the emphasis is on *generic competence* rather than linguistic competence.
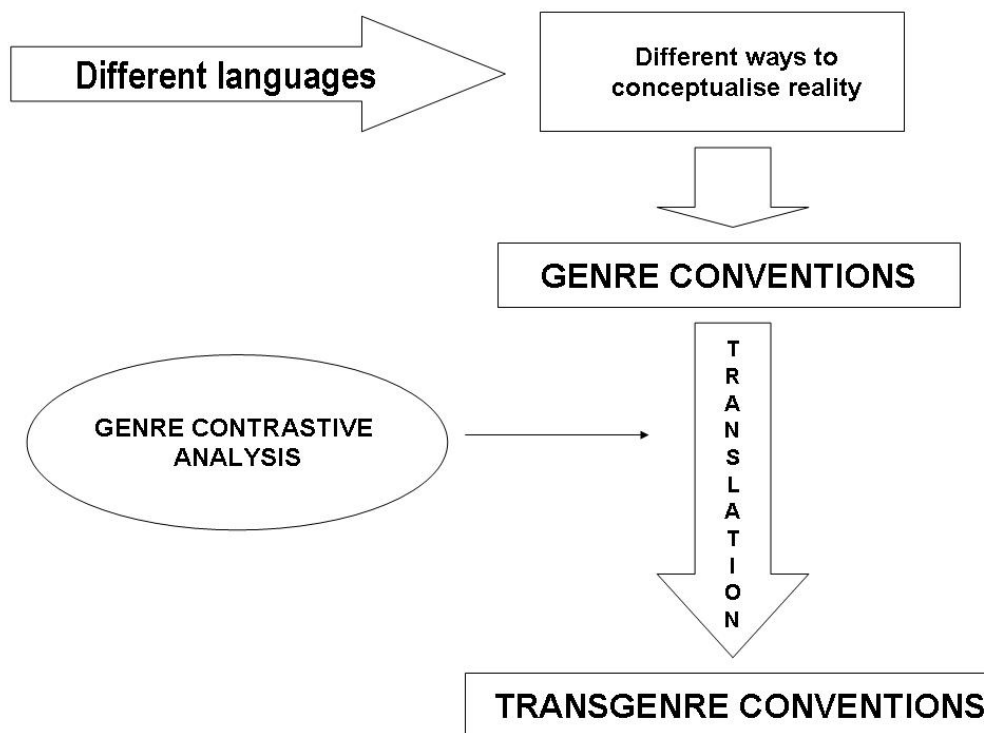
Figure 2. GENTT Corpus contrastive analysis for translation

In very specialized fields (i.e. drafting legal documents and legal translation) translators are "outsiders" (García Izquierdo and Montalt, 2002) in both source genres and target genres because they do not belong to the professional community in question (lawyers, medical practitioners, engineers…). As a result they will have to focus all their attention on the genre as a whole (and not on specific aspects such as terminology or specialized contents) and understand not only the communicative habits, limitations and possibilities of the genre, but also how the different genres interact in the source and target languages and cultures (*contrastive genre knowledge* or *genre competence).* That is, they will have to take the concepts described above into consideration, which shape the basis of our proposal, as we shall now see.

3. The GENTT specialized genres corpus

In 1999, the GITRAD research group of the Universitat Jaume I (Castellón) began to compile an English-Spanish corpus of legal documents starting from

a modest collection of Spanish-English legal tests compiled by Borja for her PhD dissertation (1998); this was then extended in a subsequent research project in which Monzó, Jennings and Borja herself took part. The idea of compiling this corpus arose from the observation that the comparison of legal texts in the source and target languages is something that specialized translators do all the time, and legal translators need to be conversant with the textual typology used in their field of specialization to ensure that they are observing the necessary textual, social, and in our case legal, conventions. (Borja 2000; Borja and Monzó, 2001).

From 2000 onwards, the GITRAD group became part of the GENTT research group and began a new, much more ambitious, project consisting of the compilation and subsequent analysis of a data base, in the form of a corpus of original texts, comparable texts (texts that fulfill the same function in the culture of the source language and that of the target language but are not translated texts but original texts in different languages) used in legal, medical and technical contexts, classified by genres. At the present stage, we are extending the perspective to add parallel texts (that is, the original and its translation into one or more languages)2, which will permit us to apply contrastive analysis to original genres and transgenres. The working languages chosen were Spanish, Catalan, English and German. From 2007, French was also included.

The GENTT corpus is organized on the basis of a structure aimed at facilitating the textual *'acculturation'* of the specialized writer or translator by means of a large number of comparable texts (the same genres in different languages, not the same texts translated) that the user can retrieve very simply. The smallest unit is the textual genre framed in a conceptual and functional taxonomy, as we shall see below, when we describe the classification system. We could describe it a **sample corpus** or **example corpus** since its aim is less the quantity than the quality of the items it includes. That is, the objective is not to construct a corpus with the greatest possible number of words, but a corpus with the greatest possible number of textual genres (with at least five sample text of each genre), organized in a previously defined taxonomy and in several languages.

The user interface is fast and easy to use, and allows the documents to be retrieved in text format, image format or HTML format. This interface also allows new samples of genres to be introduced on-line, so that GENTT collaborators from all over the world can add new items to the corpus; these

---

[2] For more information on the differences between comparable corpora and parallel corpora and the difficulties that arise in compiling parallel corpora (whether naturally or artificially generated), see Borja, 2007a.

are uploaded together with the header information described in section 5 (*Annotation of corpus*).The corpus is supplemented by a system developed by Steve Jennings (2003) for managing relational data bases, which enables extremely functional and efficient multiple searches to be made. More information can be found at http://www.gentt.uji.es.
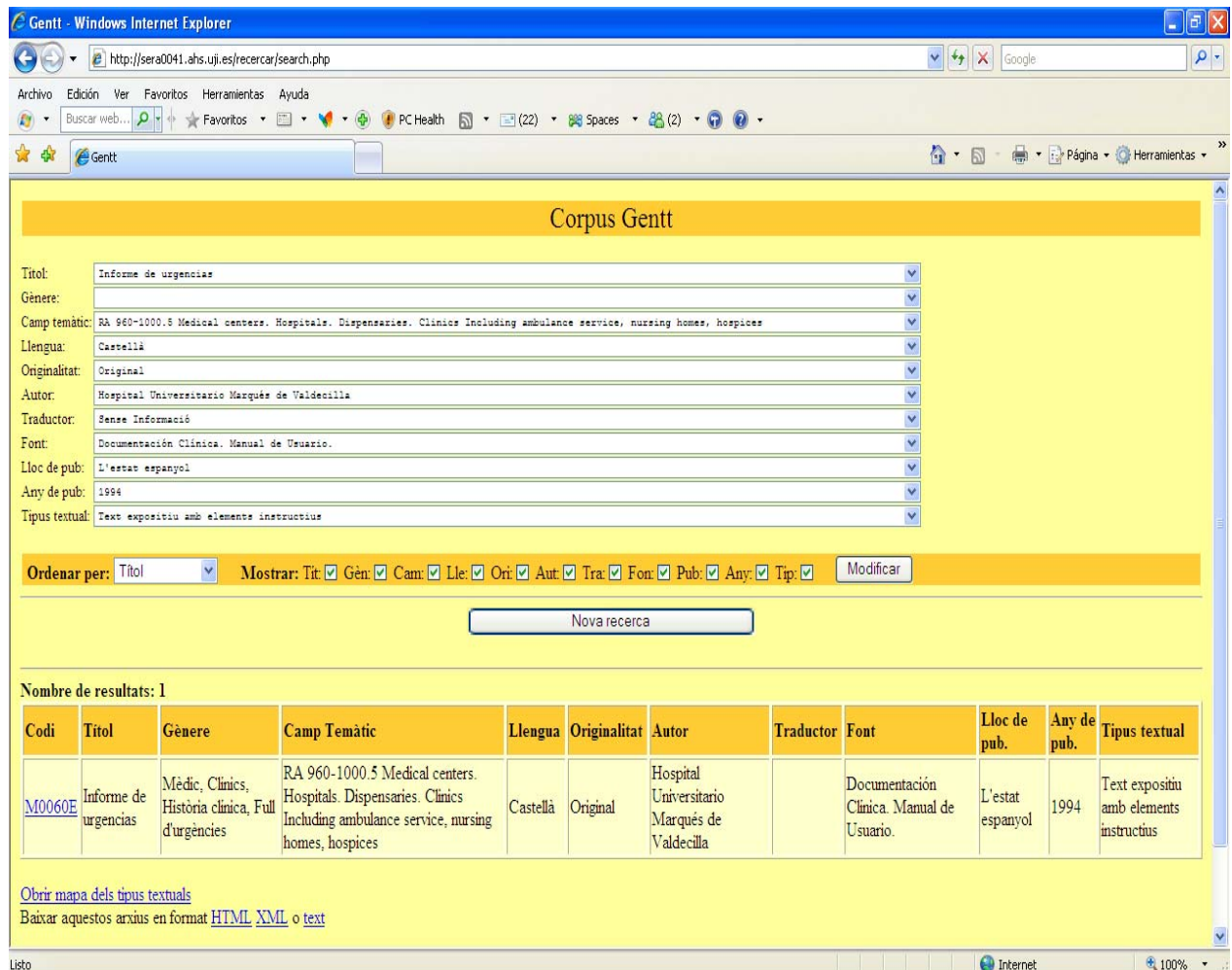
Figure 3. The GENTT search interface

Due to copyright restrictions it cannot be accessed freely but researchers can apply for a special authorization. However, the legal subcorpus (formerly called CDJ-GITRAD) corpus of legal documents (Spanish-Catalan-English) can be searched on-line at http://www.cdj.uji.es. The documents can be downloaded as full text for research purposes after applying for a password through the same web address. We expect to have a search interface ready for the medical and technical subcorpora very soon.

So we can conclude this brief description of the GENTT corpus by saying that it is:

·On-line for consultation and incorporation of new items (presently available only to GENTT collaborators and researchers with special authorization due to copyright restrictions)
·Multilingual (Spanish, English, German, French, Catalan)
·Specialized (three fields: legal, medical and technical texts)
·Textual (full text documents. It also includes an image of the original document to permit research of paratextual features)
·Example corpus of specialized genres intended to be representative of the genologic or textual mapping of three sublanguages
·Annotated (Header + Body markup + Genre information)
·Open and dynamic, in the sense that it is not considered complete at any given time, but is always open to the incorporation of new elements (genre characterisations and genre samples).
·Synchronic
·Comparable because it includes samples of "equivalent" genres in different languages
·In the future it will also be a parallel corpus because we have started the compilation of translated texts (*transgenres)*
·Intended for both professional and research purposes
·User-friendly. The GENTT Interface facilitates the compilation process, the incorporation of metadata in the Header and its consultation online. The research team is presently working to allow the online application of several corpus linguistic tools.

4. Compilation of corpus and classification of corpus units (genres)

In fact, taking as the point of departure the characteristics of dynamism, hybridation, genre systems and transgenre (the latter for translation) a parallel and simultaneous process was initiated to:

1. Define the trees of genre prototypes in the three main fields under study (legal, medical and technical) in as precise a manner as possible, given the changing, hybrid nature of the concept we endorse, and;
2. Collect sample documents for each genre identified including header information.

Each genre identified is included in the tree of genres with details of its distinctive characteristics at various levels, as we shall explain below. Usually, when a genre is identified and described, at least one sample of it is uploaded to the corpus, accompanied, in some cases by further examples of that genre. The **definition of the trees of genres** (or *textual mapping*) of the three specialized fields is done through constant observation of the socio-

professional world and takes into account the opinions of expert users, while we undertake an empirical analysis of the corpus texts compiled. The two processes are dynamic and complementary: sometimes when we identify a new genre we find we have to modify the genealogical tree; and viceversa, when identifying a new branch of the tree, we have to look for genealogical examples to illustrate it. We go from the hypothesis in search of confirmation that genres exist, and from proof that genres exist to their classification; that is, we combine induction and deduction.

We started from the idea that in order to be a good linguistic mediator in the field of discourse of Law, Medicine or Technology, the first step should be to become conversant with the enormous diversity of texts included in each of these socio-professional fields. The specialized writer/translator faces a complex world of cultural elements and new concepts which materialize in genres. For someone who is translating Software licenses for the first time it would be very useful to have a tree of the *Intellectual Property- Computer Contracts* field subspecialty in order to develop his or her textual and communicative competence. On the one hand, this process needs a system for collecting and retrieving knowledge based on a general cognitive schema of the law, and in particular *Intellectual Property Law*, in which the most specialized concepts can be identified in a relational way through a hierarchized semantic network of concepts similar to those used in the terminology.

In order to demonstrate the process, we can take a look at the subcorpus of legal genres. In this field we initially took Borja's proposed classification for legal texts (1998), which was obtained from grouping legal texts into categories, taking into account their discursive situation, the participants in the act of communication, the tone it uses, and its purpose. This classification schema was the result of studying the various existing Law Thesauri and numerous consultations with professors of Law, practicing lawyers, law librarians and legal translators. In this way an initial list of Spanish and English genres was obtained. Subsequently, the research undertaken by Monzó and Borja (Monzó and Borja, 2000) identified the need to work with different trees of genres for each type of legal system and at present, after several years of research in the GENTT group, three legal trees of genres are being used: continental or civil Law, Anglo-American or Common Law and EU Law.

The present organization of this corpus for the legal section, for instance (Legal system $\Rightarrow$ Macrogenre $\Rightarrow$ Genre $\Rightarrow$ Subgenre) generates a classification that is extremely useful for the specialized writer/translator, who can easily place the text on which they are working in the tree of genres

and compare it with the equivalent genre in the legal system of the target language.

ANGLOAMERICAN LAW SUBCORPUS MACROGENRE LABELS	ANGLOAMERICAN LAW JUDICIAL GENRES
1. Rules and regulations	**2. Judicial**
2. Judicial	Acknowledgement of service
3. Reference	Appeal
4. Doctrine	Appeal to the Family Division
5. Administrative	Appeal to the House of Lord
6. Notarial	Appeal to the Queen's Bench Division
7. Multilateral agreements	Bill
8. Unilateral statements	Certificate of Decree Nisi having become absolute
9. Expert reports	Claim form
10. Case law	Decree Nisi
11. Popular legal texts	Information
	Injuction
	Judgment
	Letter
	Minutes of Proceedings
	Motion
	Notice
	Order
	Order for committal
	Petition
	Pleading
	Request
	Rogatory Letters
	Service
	Statement of case
	Stipulation
	Subpoena
	Summons
	Writ
	Writ of summons

Figure 4. Judicial genres of the Anglo American legal system

This methodology has been applied to the other two sections of the corpus: the medical and the technical sections. Moreover, this classification is supplemented by a system of crossed searches that combines, amongst other data, the original language, the status of the text (original or translation), date of creation, the source, field, title, text, genre… . More information can be found at http://www.gentt.uji.es.

The **on-line incorporation of samples of genres** into the corpus is made through the user interface we have described above and, as it was already mentioned, it allows the documents to be retrieved in text format, image format or HTML format. New items are uploaded together with the header information described in the next section (Annotation of corpus).
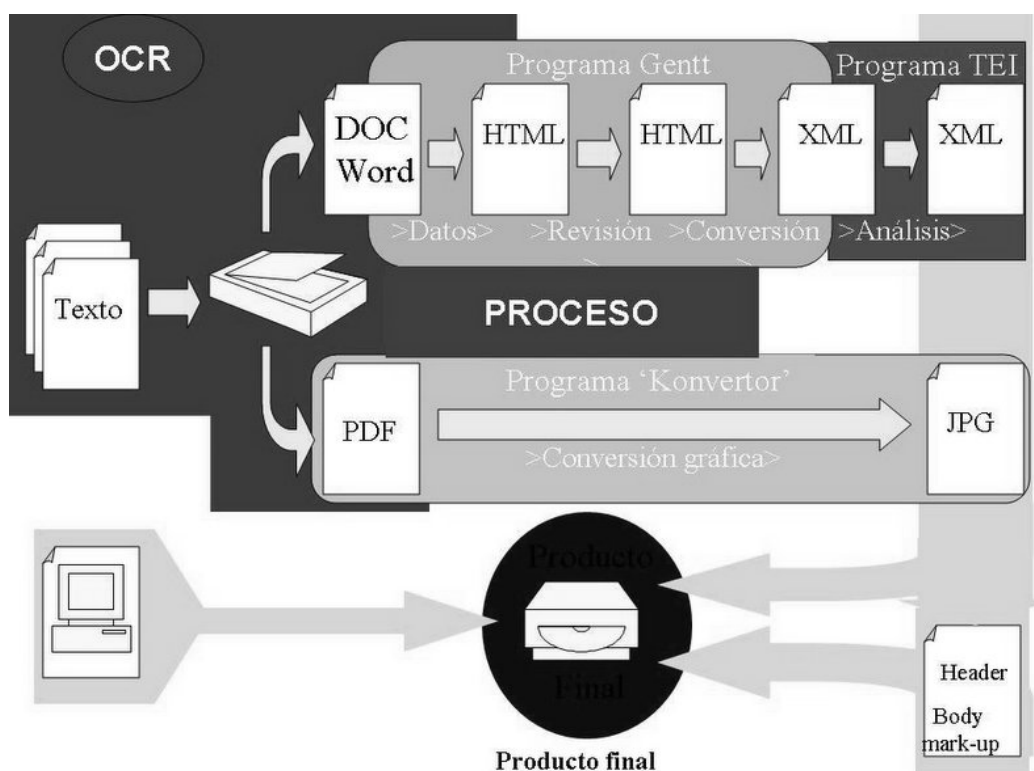

Figure 5. The GENTT corpus compilation process (Jennings, 2003)

5. Annotation

After revising previous and ongoing experiences and projects on text corpora, the GENTT team came to the conclusion that in order to annotate a corpus it is no longer necessary to use expensive software programs that are difficult to use and designed to carry out very specific tasks. Today very valuable textual

information can be obtained by applying very simple search systems. The ever-increasing diversity of annotation possibilities is also observed in modern corpora, depending on the purpose for which they are designed, and the annotation of corpora is becoming increasingly automatic and standardised. So far the annotation process has been carried out in parallel with the compilation of the corpus, which has led us to concentrate first on annotating only the Header.

5.1 Header

The content of the corpus resources is summarized into small descriptions, called metadata, which are introduced manually using the management tool already described. Metadata plays a key role in organizing the ways in which a language corpus can be meaningfully processed. Following the guidelines of the Text Encoding Initiative (TEI) which recommend the definition of a specific metadata component known as the Header, we included in the GENTT corpus header a file description, encoding description and a profile description supplying additional descriptive material about the file not covered elsewhere, such as its situational parameters, topic keywords, function, status of the text (original, translation, translation with the status of original), inclusion in a system of genres...

This profile description has been modified in several occasions to account for the new developments and approaches of the working team. During the compilation process, many samples of genres are gathered and each taken out of the context in which it originally occurred, like a laboratory specimen. Metadata restores and specifies that context, thus enabling us to relate the specimen to its original habitat. In a corpus such as GENTT the social context within which each of the genre samples was produced is arguably at least as significant as any of its intrinsic linguistic properties,

For management of the corpus, the GENTT group uses a Word complement, designed by Jennings (2003) in the first stage of the research, which incorporates a set of macro-instructions to file the document in the appropriate folder in xml or html format; it also allows the researcher to enter the document metadata using a wizard that incorporates, among other things, the system of classification (the genre trees) designed by the group. Once the classification form has been activated, a series of fields are filled in that will automatically establish the *meta* information of the html or xml document.

Firstly, the researcher processing the text has to state which language it is written in and, in the case of Catalan, Spanish and English, also define aspects concerning dialect. Next the status of the document has to be specified and, in this section we find several possibilities: original, translation

52

and translation with the status of original, although up to now, as we have explained, we have been working with original and parallel texts. Thirdly, we establish links between the document we are dealing with and others already in the corpus or that we plan to include, that is, we establish links we consider relevant with parallel texts and equivalents within the same corpus (and, in the future, with translations) as well as with genres belonging to a system of genres.

The next step is to include the title of the text we are working on, the source from which it has been extracted, the place of publication, the year, the mode (restricted to a list of six), the thematic area, which coincides with the group's division of interests and uses the *Library of Congress* classification, and the genres and subgenres defined by the group.
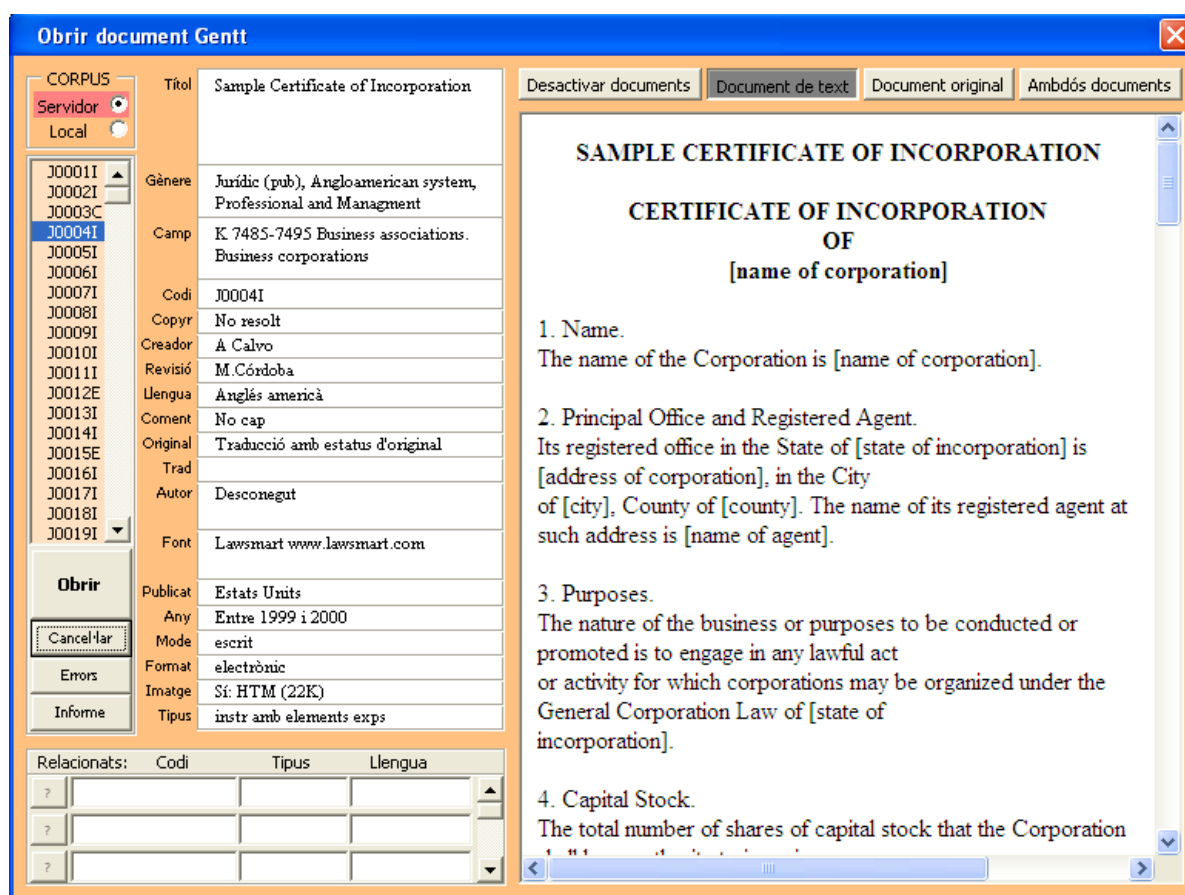

Figure 6. The GENTT corpus compilation interface

Once this specification form has been completed, the document is saved in xml with all the metadocumentary information we have entered and is sent to the server, in the Intranet of Universitat Jaume I.

5.2 Body mark-up

As we have said above, the first annotation carried out was of the Header, almost simultaneously with compilation of the corpus. Once this has been done, the team considered whether it was necessary to mark up the body of the texts with tags which, given the textual character of the corpus, would certainly have involved a major input of time and resources. After evaluating the various possibilities, and consulting experts on corpus linguistics, it was decided that a basic annotation of the corpus would be undertaken, concentrating on lemmatization of its units. In addition, a series of utilities that are today offered by morphological analyzers such as Freeling, TreeTagger or Wordsmith would be implemented in the corpus management program, the user interface, so that the formal analysis and the analysis of the macrostructure of the corpus could be undertaken. At present, then, work is being done on extending the program's utilities that, as well as permitting the selection of texts on the basis of the Header's characteristics – genre, thematic field, language, etc – will also allow us to carry out specific searches of frequencies, concordances, phraseology, regular expressions, etc. in these subcorpora.

The quantitative research using the data we can obtain directly from the corpus through our management program is thus combined with the qualitative information obtained from the same program (data from tagging the Header) and the qualitative research carried out in conjunction with experts in the fields involved. In our opinion, this combination provides us with sufficient data to carry out an exhaustive characterization of the genres with which we are working, without having to undertake more complex annotation of the corpus.

5.3 GENTT genre characterisation

To end this section, and after briefly explaining how we compile and annotate the corpus, we shall analyze the template proposed by the team for carrying out an exhaustive characterization of the genres investigated.

Our starting point is a genre characterisation system based on the template shown below. For each genre we fill in 7 blocks of data. The first section, **genre**[3], refers only to the denomination of the genre in the different working languages (*agreement*/*contrato*/*contrat*, etc.). The second section, **subgenre**,

---

[3] It should be remembered here that the starting point of classification (trees) is, in practice, the concept of **macrogenre**, a label that does not correspond to any "reality" so it is only valid for classificatory or taxonomic purposes. We should also emphasise that the trees that we are proposing are not hard and fast classifications but rather a system of classifying information that has so far proved to be valid.

is a sub-classification of the first (for instance, genre: agreement; subgenre 1: sales agreement; subgenre 2: license agreement; and so forth). At this point it should be noted that we often identify sub-subgenres, while in others there are no subgenres (Bathia, 1999). The **communicative aspects** section gathers data about the register: socio-professional field, mode, level of formality; participants: sender(s), receiver(s); and function. In Gamero's words (2001: 57):

> In relation with translation, we have to take into account that sociocultural contexts: (a) determine conventions typical of the genres in each culture; (b) fulfill a role of paramount importance when creating new genres since these reflect the specific communicative needs that may or may not be common to different cultures, and (c) establish the way genres evolve, since these adapt to the changes that occur in the culture in which they are used..

The **formal aspects** that we examine in each genre we have worked on refer to grammatical cohesion (connectors, metadiscursive elements, collocations, deixis, ellipsis, etc.) and lexical cohesion (terminology, phraseology, semantic fields, etc.), and also include contrastive aspects. **Macrostructure**, **relation to other genres** (systems of genres, Bazeman, 1994) and **additional comments** help us to complete the description of the genre of the texts we are working with.

1. GENRE	Denomination in the different working languages
2. SUBGENRE	If applicable
3. COMMUNICATIVE SITUATION	Register: socio-professional field, mode, level of formality; participants: sender(s), receiver(s); and function
4. FORMAL ASPECTS	Grammatical cohesion (connectors, metadiscursive elements, collocations, deixis, ellipsis, etc.) and Lexical cohesion (terminology, phraseology, semantic fields, etc.). Includes contrastive aspects
5. MACROSTRUCTURE	Identification of the fundamental parts of the text, of the *moves*
6. RELATION TO OTHER GENRES	*Systems of genre*, Bazerman
7. COMMENTS	Bibliographical references, interesting websites, etc. Observations of professional uses…

Figure 7. GENTT Genre characterisation template

This basic characterisation of genres is then used by the GENTT team to conduct an in-depth analysis of the three classic dimensions of genre analysis, i.e. the **formal dimension**, the **communicative dimension** and the **cognitive dimension**. We also suggest mixed approaches that combine two or even all three dimensions. How these studies develop will depend on the aspect that predominates in each case and the quantitative or qualitative method of analysis (or a mixture of both) that is chosen in each case.

This proposal makes it possible to work on a number of different lines of research that are of interest in LSP research and Translation Studies, both from the descriptive and the applied points of view. It also allows the researcher to focus on the aspect(s) of the research that have been chosen in each case.

6. Exploitation

The GENTT corpus can be exploited for various purposes and with different approaches:

- Specialized translators training
- LSP training
- Specialized writers training (in a first or second language)
- Research into specialized translation and LSP
- Automation of specialized translation and specialized writing
- Encyclopedia of specialized genres
- Genre analysis research on the three subcorpora

Here we shall describe briefly two of these possible applications, the exploitation of the corpus for specialized translators training and for LSP teaching. Additional applications and approaches have been analyzed in Borja 2005, Borja 2007b, García Izquierdo forthcoming, and Montalt 2007.

6.1 Specialized translators training

One of the most direct applications of the corpus developed is its potential for teaching purposes. In fact, from the early days of the project, it has been used extensively for training Translators (Borja and Monzó, 2001; Monzó 2005). In fact, we have been using the GENTT corpus in the translation courses at Universitat Jaume I since 2002 with excellent results, and there are ongoing research projects on translation teaching methodology using systems of genre conventions. In the translation class we approach specialized translation with a number of activities involving corpora:

- Identification of existing corpora
- Compilation of ad-hoc mini-corpora
- Extraction of specialized terminology, phraseology and discourse markers
- Search for meanings of specialized terms in corpora as an alternative to traditional dictionary searches
- Drafting documents in mother tongue using genres in the same language as reference
- Translation of documents using parallel genres in the target language as reference

Examining an equivalent text in the target language of the text to be translated allows the translation student to observe the structure, terminology and phraseology typical of the genre. In this way he or she will make informed and "inspired" decisions when translating. This does not mean that the structure of the parallel text should be reproduced, since this would result in an adaptation, not a translation. Rather it consists of familiarizing oneself with the conventions of the genre in both cultures, to help decide the type of translation required, which may be closer to the original or to the target language, but always based on a knowledge of genre conventions in both languages.

Finally, it is evident that corpora of translated texts, as well as fulfilling the functions described above, have important implications for the future development of software for translators. Textual typologies within a specialist language facilitate the systematization of translation by analyzing corpora of standard documents. The results obtained from an analysis of this kind will make it possible to create specialized terminological databases, professional tools for consultation and documentation, databases of standard documents and assisted-translation programs based on translation memories.

6.2 LSP teaching

The possibilities of applying the work on genre-based corpora to LSP teaching are evident. According to Bhatia (1997), work with genres pertaining to the student's specialty awakens in learners an explicit desire for conscious participation in the professional community and a feeling of 'shared ownership' of their communicative resources rather than learning structures and terms mechanically.

Bhatia, who has taught and researched legal language through its genres (see Bhatia, 1997a; Bhatia and alii, 2008), believes that learners of a specialized language such as legal language need to develop four basic skills in order to

overcome the fear of their lack of knowledge of the field of discourse of Law. Thus, as well as mastering the general language, they need: (1) to understand the specialist's communication code; (2) to familiarize themselves with rhetorical resources and those contained in specialized genres; (3) to understand the various contexts in which specialized communication takes place; and (4) to be capable of using specialized genres to respond to new and unexpected situations.

All these four skills can be enhanced using the GENTT corpus textual *mapping* and genres characterisations. Another advantage of this method is that the language is learnt in its true context and learning schedules can be designed with very specific needs in mind. It is possible, for example, to design a plan for teaching-learning restricted to the language of *franchise agreements*, treating it as a single genre or as a system of genres that includes all the documents that accompany franchise agreements (franchise dossier, preliminary agreement, agreement, commercial documentation, operating manual), and even the oral genres related to them, such as negotiation meetings. The same applies to medical or technical genres.

7. Conclusions and future work

The intercultural approach to specialized writing and translation adopted by the GENTT research group assumes that the specialized writer/translator needs information of three kinds: conceptual, textual and linguistic. When they have this information, specialized writers/translators can improve their knowledge, both linguistic and extralinguistic, using a self-taught process. With the GENTT corpus we try to provide writers and translators with this information, which is in fact information about genre conventions: formal aspects, socioprofessional functions and use. As we have tried to demonstrate, the application of the theory of genres has already generated many initiatives in the field of language teaching and training translators. It is foreseeable that linguistic and translation research on electronic corpora of genres will, in the medium term, facilitate the automation of writing, indexing and management of specialized texts. However, some of these lines of research are still at an early stage and many applications remain to be developed.

Our research group has now expanded considerably with the collaboration of groups and individual researchers that have become interested in the project and are participating in it at various levels. We are still immersed in the process of compiling texts, since, as we have said, it is a dynamic process, which can be constantly updated thanks to the system's digital architecture. Some of our collaborators are working on perfecting the trees of fields for each subcorpus, adding characterisations of new genres and classifying them,

while others are looking for examples of genres identified in theory but for which we do not yet have any examples. The digital and interactive nature of the corpus and on-line access to it enables people from all over the world to contribute to it with the identification of new genres or with new examples of genres.

The GENTT team is also currently involved in the preparation of an on-line ´*Encyclopedia of specialized genres*´. It will be an online resource that will provide the user with textual information (he will be able to visualize the text in its original format), formal and linguistic information (genre features compiled applying the GENTT genre characterisation template) as well as conceptual information about the context field and the core concepts appearing in the text (García Izquierdo and Monzó, 2003).

A fundamental aspect for the future of the GENTT project is the development of its corpus management and query system. Some members of the team are currently working on the design of a user-friendly web interface which will contain a set of computational linguistics tools for managing and querying large text corpora with linguistic annotations. Of course, we are taking advantage of the experience and tools of other corpus and computational linguistics research groups but are trying to incorporate the specific data required by a genre-based corpus such as ours, particularly data concerning genre identification and description. We hope to have our own corpus query and management interface ready very soon to facilitate the work of people interested in using it as a research tool. We believe our corpus management software can provide new projects based on textual genres with an easy-to-use framework on which to base their research. Ongoing collaboration with other projects will enhance the system as we identify new data types and modules to include in the framework.

Our objective for the future is to try to link theoretical work in linguistics and Conceptual Information Systems (formal linguistics, lexical semantics, ontological modeling) with empirical data (intensive approaches based on corpus statistics) as well as with socioprofessional qualitative data obtained from personal interviews with users of specialized languages. Conceptual Information Systems provide a multi-dimensional conceptually structured view on data stored in relational databases. On restricting the expressiveness of the retrieval language, they allow the visualization of sets of related queries in conceptual hierarchies.

To end, we list some of the lines of work that GENTT has proposed for the future and in which it invites anyone interested to participate.

- Increasing the electronic corpus by adding more samples of genres already existing in the trees of genres, including header information and genealogical data
- Incorporation of new genre descriptions which are added to the trees of genres
- Incorporation of parallel genres in different languages (genres "equivalent" to those identified and existing in the corpus)
- Incorporation of translated texts
- Tagging particular subcorpora for specific research work
- Exploring new corpus applications for training translators and ways in which corpora can be used as a reference tool for professional translators and writers of specialized material
- Initiating quantitative and qualitative research on transgenres
- Collaborating with computational linguistics researchers to define strategies for the automation of writing, indexing and management of specialized texts.

References

BHATIA, V. K. (1993): *Analysing Genre. Language Use in Professional Settings*, Essex, Longman.

BHATIA, VIJAY V.K. (1997): "Translating legal Genres". In Anna Trosborg (ed.)*Text Typology and Translation*, Amsterdam, John Benjamins.

BHATIA, V. K. (1997a): "Applied genre analysis and ESP", in T. Miller (ed.), *Functional approaches to written text: Classroom applications*, 134-149. English Language Programs: United States Information Agency. http://exchanges.state.gov/education/engteaching/pubs/BR/functionalsec4_10.htm. Date of last access: 14 June 2008.

BHATIA, V. K. (1999): "Integrating products, processes, purposes and participants in professional writing", Candlin, C. N. & K. Hyland (eds.): *Writing: Texts, Processes and Practices*, London, Longman, pp. 21-40.

BHATIA, V. K. (2004): *Worlds of Written Discourse: A Genre-Based View*, London & New York, Continuum International Publishing Group.

BHATIA, V. K.; CANDLIN, C. N. & J. ENGBERG, (eds.)(2008) *Legal Discourse across Cultures and Systems*, Hong Kong University Press.

BAZERMAN, C. (1988): *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison, University of Wisconsin Press.

BAZERMAN, C. (1994): "Systems of genres and the enactment of social intentions", in A. Freedman & P. Medway (eds.): *Genre and the New Rethoric*, London, Taylor & Francis Ltd, pp. 79-101.

BAZERMAN, C. (1994a). "Where is the classroom?", in A. Freedman & P. Medway (eds.). *Learning and teaching genre*, pages 25-36, Portsmouth, NH, Heinemann.

BAZERMAN, CH. (1999): "Introduction: Changing Regularities of Genre", *IEEE Transactions on Professional Communication, vol. 42, nº 1*, pp. 1-3.

BERKENKOTTER, C. &   HUCKIN, T.N. (1995): *Genre knowledge in disciplinary communication: Cognition / Culture / Power.* Hillsdale, New Jersey, Lawrence Erlbaum Associates.

BORJA ALBI, A. (2007a): "Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project", in Gunilla Anderman & Margaret Rogers (eds.) *Incorporating Corpora - The Linguist and the Translator*, Series Translating Europe, pp. 243-265.

BORJA ALBI, A. (2007b): "Los géneros jurídicos", in E. Alcaraz Varó, J. Mateo & F. Yus (eds.): *Las lenguas profesionales y académicas*, Barcelona, Ariel, pp. 128-68.

BORJA ALBI, A. (2005): "Organización del conocimiento para la traducción jurídica a través de sistemas expertos basados en el concepto de género textual", in I. García Izquierdo (ed.): El género textual y la traducción, Bern, Peter Lang, pp. 37-68.

BORJA ALBI, A. (2000): *El texto jurídico inglés y su traducción al español.* Barcelona, Ariel.

BORJA ALBI, A. (1998): *Estudio descriptivo de la traducción jurídica. Un enfoque discursivo.* Tesis doctoral, Castellón, Universitat Jaume I.

BORJA, A. &   E. MONZÓ NEBOT (2001): "Aplicación de los métodos de aprendizaje cooperativo a la enseñanza de la traducción jurídica: cuaderno de bitácora" in F. Martínez Sánchez (ed.) *EDUTEC '01. Congreso Internacional de Tecnología, Educación y Desarrollo Sostenible*, Múrcia, Edutec, CAM.

BORJA ALBI, A., GARCÍA IZQUIERDO &   V. MONTALT (forthcoming): "Research methodology in specialised genres for translation purposes", in I. Mason (ed.): *The Interpreter and Translator Trainer,* vol. 2.

DAM, H.V.; ENGBERG, J & H. GERZYMISCH-ARBOGAST (eds.) (2005): *Knowledge Systems and Translation*, Berlin, Walter de Gruyter Inc.

ENGBERG, J.; GROVE DITLEVSEN, M.;KASTBERG, P. & M. STEGU (eds.) (2007): *New Directions in LSP Teaching*, Ben, Peter Lang.

GAMERO, S. (2001): *La traducción de textos técnicos*, Barcelona, Ariel.

GARCÍA IZQUIERDO, I. (2000): *Análisis textual aplicado a la traducción*, Valencia, Tirant lo Blanch.

GARCÍA IZQUIERDO, I. (2002): "El género: plataforma de confluencia de nociones fundamentales en didáctica de la traducción", *Revista Discursos. Série Estudos de Tradução, nº 2*, Lisboa, Universidade Aberta.

GARCÍA IZQUIERDO, I. (ed.) (2005): *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas.* Bern, Peter Lang.

GARCÍA IZQUIERDO, I. (2005a): "Estudios recientes en traductología basados en corpus: el corpus electrónico GENTT", *VII Seminário sobre traduçao científica e técnica em Língua Portuguesa*, Lisboa, Unión Latina.

GARCÍA IZQUIERDO, I. (2007): "Los géneros y las lenguas de especialidad", in E. Alcaraz Varó, J. Mateo & F. Yus (eds.): *Las lenguas profesionales y académicas*, Barcelona-Alicante, Ariel-IULMA, pp. 119-127.

GARCÍA IZQUIERDO, I. (forthcoming): *Divulgación médica y traducción: el género Infomración para pacientes/Fact Sheet for Patients*, Berna, Peter Lang.

GARCÍA IZQUIERDO, I. & V. MONTALT (2002): "Translating into Textual Genres", *Linguistica Antverpiensia, vol. 1*, Bruselas, pp. 135-143.

GARCÍA IZQUIERDO, I. & E. MONZÓ (2003): "Corpus de géneros GENTT. Una enciclopedia para traductores", E. Sánchez Trigo y O. Díaz Fouces (eds.): *Traducción & Comunicación, vol. 4*, Servicio de Publicacións Universidade de Vigo, pp. 31-55.

HALLIDAY, M. A. K. & HASAN, R. (1985): *Language Context and Text. Aspects of Language in a social-semiotic perspective*, Victoria, Deakin University Press.

HATIM, B. & E I. MASON (1990/1995): *Teoría de la traducción. Una aproximación al discurso*, Barcelona, Ariel. Traducción de Salvador Peña.

JENNINGS, S. (2003): *The Development of Software Tools for Corpus-Based Genre Research in Translation Studies: A Practical Application in the Context of the Gentt Project* [Research project], Castelló de la Plana, Departament de Traducció i Comunicació, Universitat Jaume I.

MONTALT, V. (2005): "El género como espacio de socialización del estudiante de traducción científico-técnica", in García Izquierdo, I. (ed.) *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*, Bern, Peter Lang, pp. 19-36.

MONTALT, V. & M. GONZÁLEZ DAVIES (2007): *Medical Translation Step by Step,* Manchester, St Jerome Publishing.

MONTALT, V. E I. GARCÍA IZQUIERDO (2002): "Multilingual corpus-based research of medical genres for translation purposes: the medical corpus of the GENTT project", *Actes del Congrés Traduir la Ciència*, Barcelona, Universitat Pompeu Fabra.

MONTALT, V. GARCÍA IZQUIERDO, I. & EZPELETA, P. (Forthcoming): "The acquisition of Textual Competence through Genre", (e.p.) *Translation Journal, 12 (3)*.

MONZÓ NEBOT, E. (2002): *La professió del traductor juridic i jurat: descripció sociològica del profesional i anàlisi discursiva del transgènere*. Tesis doctoral, Castellón, Universitat Jaume I.

MONZÓ NEBOT, E. (2005): "Reeducación y desculturación a través de géneros en traducción jurídica, económica y administrativa", in *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*, pp. 69-92, I. García Izquierdo (ed.). Bern, Peter Lang.

MONZÓ NEBOT, E. & A. BORJA ALBI (2000): "Organització de corpus. l'estructura d'una base de dades documental aplicada a la traducció jurídica. *Revista de llengua i dret*. Vol . 34 (9:21). Diciembre, 2000.

ORLINOVSKI, W.& J. YATES (2002): "Genre Systems: Structuring interaction through Communicative Norms", *Journal of Business Communication, vol. 39, nº 1*, pp. 13-35.

SHÄFFNER, C. (2002) (ed.): The role of Discourse Analysis for Translation and in Translation Training, Clevendon, Multilingual Matters Ltd.

SHÄFFNER, C. (2004) (ed.): Translation Research and Interpreting Research. Traditions, Gaps and Synergies, Clevendon, Multilingual Matters.

STRAUSS, A.L. & J. CORBIN (1998): *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, London: Sage.

SWALES, J. (1990): *Genre Analysis: English in Academic and Research Settings*, Cambridge, Cambridge University Press.

SWALES, J.& C. FEAK (2000): *English in Today's Research World: A Writing Guide*, Ann Arbor, MI, The University of Michigan Press.

TROSBORG, A. (ed.) (1997): *Text Typology and Translation,* Amsterdam, John Benjamins.

VENUTI, L. (1995): *The Translator's Invisibility. A History of Translation*, London, Routledge.

WEISSBERG, R. & S. BUKER (1990): *Writing Up Research. Experimental Research Report Writing for Students of English.* Englewood Cliffs, New Jersey, Prentice Hall.

WILSON, T. D. (2000): "Recent Trends in user studies: action research and qualitative methods", *Information Research, 5 (3),* available in http://informationr.net/ir/5-3/paper76.html

***

ABSTRACT

A Multidisciplinary Approach to Specialized Writing and Translation Using a Genre Based Multilingual Corpus of Specialized Texts

Isabel García Izquierdo & Anabel Borja Albi
GENTT Research Group, Universitat Jaume I
Spain

GENTT's main focus is the multilingual study of genres in professional legal, medical and technical contexts. Its aim is to identify genre conventions by means of a hypothetical deductive approach using corpus technologies. The project has focused on compiling a multilingual (Catalan, English, German, Spanish and French) annotated corpus of specialized discourse texts that could prove useful to translators and writers of professional texts, providing them with text models and patterns to be used as textual, conceptual, linguistic and terminological reference. The corpus design is also intended to create a knowledge management system, similar to terminological knowledge representation systems, structured around the notion of genre, for both translators and writers of professional texts (Borja, 2005). The interdisciplinary approach adopted by the GENTT research group assumes that specialized writers and translators need information of three kinds: conceptual, textual and linguistic (García and Monzó, 2004). When in possession of this information, they can progressively improve their professional competence, both linguistic and extralinguistic, using a self-taught process.

This research applies an interdisciplinary approach: functional linguistics (García Izquierdo, 2000), genre analysis theory applied to specialized discourse (Borja, 2000) and sociology of professions (Monzó, 2002). The GENTT group understands the concept of genre as a product, i.e. as a successful achievement of a specific communicative purpose which is characterized by its conventionality, restrictions and typicality (Bazerman, 1999). Following Bekenkotter and Huckin (1994), we also understand genres as devices which we use to package our speech and make it a recognizable response to the demands of a given situation. But we also see genre as a process, a social construct that gradually emerges, evolves and disappears according to communicative needs (Bhatia, 2004).

***