

TERMINOLOGIE ET TRADUCTION  
n° 1/1988

Sommaire

	Page
Présentation	1
Groupe de travail interinstitutionnel "Terminologie - Documentation" (GIIT) La terminologie, pour quoi faire ?	5
Alevantis, Panayotis Creating the greek CELEX database, technical or managerial challenge?	11
Andersen, Merete Databasen CELEX på Dansk	23
Almeida, António Les commissions permanentes du Parlement Européen: dénominations et analyse linguistique	31
Service de Terminologie du Conseil des Communautés européennes Le gouvernement espagnol, terminologie officielle en neuf langues	38
Schumacher, Nestor Metaphern des europäischen Sprachgebrauchs	55
Marchand, Claude Alice Glossaires "Pêche"	81
Fiamozzi, Irmgard Glossar "Fische und andere Meerestiere"	85
Hames-Holtz, Annette Des mots utiles à connaître	109

	Page
Wilson, Barry / Carlisle, Alan GEPRO - an online system for Translation Management	115
Jeanne, Bernard La bataille du français: nouvelle Guerre de Cent ans?	119
Robert, Marie-Christine Soyez "reactics" sur le "field"!	121
Martin-Dupont, Claude Pro and contra	124
Jeanne, Bernard Mastère et magistère: Le français bouge à la japonaise!	127
Lerat, Pierre Entre traduction et normalisation terminologique	129
Berteloot, Pascale / Mulders, Leo / Winterhoff, Karsten Rapport sur le congrès de Londres "Translating and the Computer"	137
Quicheron, Jean-Bernard Rapport sur le congrès international de Trèves "Terminologie et génie de la connaissance"	147
Ozdoba, Joachim Rapport sur le séminaire de Paris "Recherche et développement dans les industries de la langue"	155
Pimpinella, Pietro Idea. VI Colloquio Internazionale del Lessico Intellettuale Europeo	159
Heimbeck, Margret / Kalamvoka, Panayota Le dictionnaire des industries du CILF: un livre utile	163
Annonce du congrès "Terminologie et Industries de la Langue" Paris, 10-20 janvier 1989	165
Ammundsen, Peter Ausonius on Fish	167
Sommaire récapitulatif de l'année 1987	173

## CREATING THE GREEK CELEX DATABASE, TECHNICAL OR MANAGERIAL CHALLENGE?

### 1. INTRODUCTION

CELEX is the interinstitutional computerized documentation system for European Community law. It is a database that contains acts of Community legislation and case law with their full text, as well as bibliographical data on preparatory acts and parliamentary questions. (see ref. 1).

CELEX has been established as a multilingual system because the Treaties provide for nine official working languages. This multilingual aspect of CELEX is not a mere luxury but rather a matter of primary importance for the European citizen: Community law supersedes in many cases national law and that is the reason why CELEX, the main distribution channel for Community law, is considered a pillar of the united Europe of 1993.

CELEX exists already in French, English, German, Dutch and Italian. The Danish and Greek versions are under preparation while the Spanish and the Portuguese versions are to follow.

### 2. THE TECHNICAL CHALLENGE OF MULTILINGUISM

The existing versions of CELEX do not contain special Latin characters (i.e accented French or German letters); the code used for these versions is the basic Latin alphabet (ISO 646). Texts with special characters are transcoded into standard ASCII. In addition texts are further transposed into upper case ("appauvrissement"). Some characters are transliterated as one character and some as two (see Table 1).

The creation of a Greek CELEX database in such a "poor" environment is impossible. One way would be to replace lower case Latin letters with upper case Greek ones or replace the Latin alphabet by the Greek. Serious problems would arise if these solutions were adopted; in fact, legal texts written only in upper case letters are not recognised as binding in Greece. On the other hand, the Community's legal texts in Greek do contain words in Latin ("sui- generis" decisions, "ad hoc" committees, "ESPRIT" programme) that cannot be translated or transliterated into Greek and can in fact constitute useful search terms in the context of a Greek database. Last but not least the accents can in no way be omitted from Greek texts. In

Greek, accents mark the syllable that is stressed and can play an important role in making the meaning of a word clear. Words may be written the same but have different meanings, depending on the syllable which is stressed (see Table 2, ref. 2).

All these considerations led to the conclusion that an extended character set was needed for the Greek base. This set would provide for the basic Latin characters plus the Greek accented and non-accented letters. In fact, what was needed was an 8-bit standard for the Greek-Latin alphabet similar to the one used for Western European languages (ISO 8859/1). The standard would have to be international because that is what the Commission is expected to prefer (see ref. 3).

On the other hand, the standard would have to be implemented in real world products, particularly in:

1. terminals and printers;
2. the Database Management System (DBMS);
3. the programs that "feed" data into the base.

The Commission's Informatics environment made the challenge still greater by stipulating conformity with the then newly established Informatics Architecture (see ref. 4). That was quite logical. It would be unthinkable to introduce a special multilingual terminal (including Greek) into the Commission - especially the Translation divisions - in order to have them interrogate CELEX, without it being possible to connect such a terminal to the word processing system on the departmental computer, to EURODICAUTOM or to other bases outside the Commission. The existing situation where three different terminals are used for the above-mentioned tasks (VT 100s for accessing internal and external databases, ETS 2010s for word-processing and SIEMENS 9751s for the Greek EURODICAUTOM) is not very rational and could not serve as an example.

### 3. MEETING THE CHALLENGE

In January 1986, when an official was given the task of setting up a Greek version of CELEX,

- there were no existing standards;
- there was no question of supporting a Greek environment in the three main products mentioned above; in fact Greek was not required as obligatory in the Commission's procurement specifications for hardware and software products;
- the new Informatics Architecture was still evolving;
- the creation of the Greek base was not considered as a major project by Commission departments although an initial feasibility study was duly completed in July 1985.

It was thought originally that the Greek CELEX base could be created with the same procedures used for the existing language versions. By mid-86, however, it was realised that a special project and additional resources would be needed.

Before the project could advance, standards had to be drawn up. Special contacts were established between ELOT (Hellenic Standards Organization) and the convenor of the corresponding Working Group of ISO (who was working at the time as an expert at the Commission's DG XIII). As a result, an 8-bit Greek standard was established jointly by ELOT and ISO in June 1986 (ELOT 928) and it soon became an ECMA and ISO standard (ISO 8859/7).

Meanwhile, contacts were established with industry (see ref. 5), to provide for terminals incorporating Greek characters according to the standard while conforming with the Informatics Architecture (VT 220 compatibility). An interim solution via DRCS (dynamically redefinable character sets) was also envisaged but was not put into practice at the time because of lack of resources.

Following established procedures (see ref. 5), a new feasibility study was adopted by Commission departments in January 1987. Progress on the project was conditioned by the availability of terminals on the list of approved hardware and software products for use by the Commission and by the availability of a DBMS (MISTRAL V5) supporting a multilingual environment (including Greek).

By autumn 1987, a multilingual terminal was submitted for testing in the Commission's Informatics Workshop (SCRIBEL terminal by CREL; the company was later renamed TIL Technologies and resubmitted the product under the name of ALTAIR). Although the layout of the Greek keyboard was not ergonomically correct, a series of tests with the DBMS proved conclusive; MISTRAL V5 was accepted in February 1988 as capable of supporting a multilingual environment. The acceptance was based on the creation of a test base in Greek containing 5 documents; it was possible to interrogate that base successfully from Greece (using an 8-bit transparent access).

In parallel, following an ever-increasing political interest for the proper introduction of Greek into the Commission's computer systems and looking ahead to the Greek presidency (July 1988), special attention was given to the implementation of a correct version of Q-one, the Commission's standard word-processing package under UNIX. Until then, Q-one had been using an ambiguous coding for Greek and that had caused serious problems, e.g. in file transfer, alphabetical ordering etc. A new table for Greek was implemented, and the appropriate keycaps, printcaps, termcaps and collate files created. It thus became possible by June 1988 to produce and print Greek texts on Q-one, using standard VT 220 terminals (WYSE 65) and laser printers. Of course, a DRCS solution was implemented. Some problems that still remain will be solved through the use of true Greek-Latin

VT 220 terminals and the implementation of special Greek laser printer fonts. A special program (still to be developed) will allow data transfer from the database towards Q-one (downloading) as well as from Q-one to the database (uploading - data input). It is hoped that the upgraded Q-one will go into general use by October 1988. It will be accompanied by an updated version of ILS, the Commission's pivot transcodification system. ILS (INSEM Local Server) permits the transfer of texts between a series of word-processing systems (Olivetti ETS 2010, Philips, etc.) and Q-one.

Other products properly supporting multilinguism have also been submitted for testing in the Informatics Workshop (EURO-PC by SIEMENS) and are expected to figure on the Commission's lists of approved products soon.

#### 4. THE GREEK CELEX ITSELF

CELEX documents consist of two main parts, the analytical part (keywords classifying documents by type or subject matter, relevant dates, names of authors, relations to other documents etc.) and the textual part (title, full text). Data of the analytical part are fed in a coded form into a special file (called ARCHIVE). By the use of special multilingual tables, the ARCHIVE codes are machine-translated into the respective languages before being introduced into the corresponding CELEX bases. Titles in all languages are also introduced into the ARCHIVE in order to offset the fact that texts arrive later.

Therefore, for a Greek CELEX database, three types of data are to be considered:

- translated analytical codes (CELEX tables);
- titles of documents;
- texts.

The CELEX tables can be completed for Greek very easily through the use of the THELEM data entry package.

The titles of CELEX documents are to be introduced via Q-one or by the use of data already existing on magnetic media (e.g. the tapes used to produce the Greek version of the "Directory of Community Legislation in force").

A large proportion of the texts (legislation, case law) is available, mainly from the Office for Official Publications on magnetic media of different formats. Simple transcodification and formatting programs are to be developed in order to feed these data into the base.

Now that the basic building blocks (terminals, DBMS, data entry software) are almost in place and (almost) linked together, the writing of the above mentioned programs that feed the data into the base is to be considered a rather routine task. If the necessary resources (6 programmer months) are allocated as expected in July 1988, the infrastructure should be completed by early 1989; the base will then be loaded with analytical data and texts and will be opened to the public in the course of 1989.

## 6. SOME INTERESTING PROSPECTS

When a general problem is solved, minor specific sub-problems become easier to solve too. The introduction of the Greek script into the Commission's Informatics Architecture - through the Greek CELEX project - means that all the Latin languages can be treated correctly too. The "Latin" versions of CELEX can now contain special characters ("rich" versions) and these versions can be incorporated into the Commission's office automation environment. The special transfer programs developed for Greek merely need adapting. The new modernization plan for CELEX, the adoption of which is pending, provides for the introduction of the full text of Commission proposals into CELEX. When the final text of these proposals is adopted by the Council, translators will need only to download the proposal into their word-processor, introduce amendments as necessary and send off the final text to the relevant authorities (electronically or on paper).

As far as the Greek CELEX itself is concerned, there are further prospects. It will be the first full-text database to be available on-line in Greece. This means that it will serve as a pilot for the opening up of an information market in Greece, in line with the Community programme for the development of a specialized information market in Europe (see ref. 6).

On the other hand, the Greek CELEX can serve as a benchmark for a more advanced morphological study of the Greek language. Last but not least, all language versions of CELEX can serve as testbeds for new machine translation packages. In fact CELEX documents are identified by a single document number - the same in all its language versions. So, if one finds a decision or a regulation in the English base, one can find the German version in seconds by using its document number.

Panayotis Alevantis  
Integrated Information Systems/Terminology  
and Computer Applications Department (TAI)  
Commission of the European Communities  
Brussels

## ACKNOWLEDGEMENTS

Special thanks are due to the system supplier, Mr Jose MARIN-NAVARRO of the Service for the Development of Applications; special assistance was provided through valuable advice by Mr. Marios RAISSIS of the Computing Centre, Ms Georgia EFTHYMIPOULOU of DG XIII and MM. Michael BALTSAVIAS and Nikolas PAPADIMITRIOU of the Translation Directorate. Q-one was adapted by Ms Monique VINCENT of TER/Brussels. Thanks are also due to all persons in and outside the Commission who supported this project in any way (critical, moral or political).

## REFERENCES

1. For information on CELEX please contact:

- Officials of European Community Institutions:  
Commission of the European Communities, Service des Abonnes Internes, resp. Mr MAES, IMCO 6/3, Commission - Brussels, tel. 23-57560;

- External users:  
Commission of the European Communities, Service des Abonnes Externes, resp. Mr BERGER, 200 rue de la Loi, ARL 3/2, B- 1049 Brussels, tel. 235 00 01, 235 00 03; or

Office for Official Publications of the E.C., Sales Department, rue Mercier 2, 1-2985 Luxembourg, tel. 4992 82563.

2. A. Passias, "Problèmes liés à l'élaboration d'instruments linguistiques du projet INNOMOS: Synonymie et polysémie dans le langage juridique hellénique". 2° Convegno Informatica Giuridica al servizio del Paese, Roma giugno 1978.

3. Council Decision of 22 December 1986 on standardization in the field of information technology and telecommunications (87/95/EEC) OJ L 36, 7/2/87, p. 31.

4. Guidelines for an Informatics Architecture (1986 - 1991), November 1986.

5. Guide Informatique des Institutions Europeennes.

6. 84/567/EEC : Council Decision of 27 November 1984 adopting a Community programme for the development of the specialized information market in Europe. OJ L 314, 4/12/84, p. 19.



FIGURE 1: ISO 646 code table (ASCII)

FIGURE 2: ISO 8859/1 code table (covers Danish, Dutch, English, Faeroese, Finnish, French, German, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish and Swedish).

FIGURE 3: ISO 8859/7 code table for Greek

#### ΠΕΡΙΛΗΨΗ

Η δημιουργία της ελληνικής βάσης CELEX είναι μια πρόκληση που συνδέεται με τη γενικότερη εισαγωγή της πολυγλωσσίας στα συστήματα πληροφορικής των Κοινοτήτων. Η θέσπιση προτύπων για το ελληνικό αλφάβητο, η ενσωμάτωσή τους σε χειροπιαστά προϊόντα (τερματικά, εκτυπωτές, πακέτα προγραμμάτων) και η σύνδεση όλων των δομικών λίθων μεταξύ τους προκειμένου να αποτελέσουν ένα ενοποιημένο σύστημα, να τι θα έχει επιτευχθεί μέχρι το 1989, όταν η ελληνική βάση θα ανοίξει στο κοινό.

Οι θετικές επιπτώσεις από το έργο αυτό περιλαμβάνουν τη δυνατότητα εισαγωγής των ειδικών λατινικών χαρακτήρων στο σύστημα, την προώθηση της δημιουργίας μιας αγοράς πληροφοριών στην Ελλάδα και την εξασφάλιση ενός χώρου δοκιμών για μια πρώτη ουσιαστικότερη μελέτη της μορφολογίας της ελληνικής γλώσσας. Υπάρχουν επίσης δυνατότητες χρησιμοποίησης του συστήματος για δοκιμές πρωτότυπων συστημάτων αυτόματης μετάφρασης. Πάνω απ' όλα όμως το ελληνικό CELEX θα συμβάλει στην καλύτερη ενημέρωση των νομικών, των επιχειρηματιών και των στελεχών του δημόσιου και του ιδιωτικού τομέα στο Κοινοτικό Δίκαιο.

TABLE 1: TRANSLITERATION OF SPECIAL LATIN CHARACTERS IN CELEX DOCUMENTS (apauvrissement).

é, è, ê	= e		
á, à, â	= a;	ä = ae;	ǣ = aa
æ	= ae		
ç	= c		
í, ì, î, ĩ	= i		
ó, ò, ô	= o;	ö = oe;	ø = oe; œ = oe
ú, ù, û	= u;	ü = ue	
ß	= ss		

(upper case letters are similiary transliterated as one or two capital letters, or as one capital and one lowercase, if they are in the beginning of a word).

TABLE 2: POLYSEMY IN GREEK WORDS (table adapted from ref. 4)

άλλα	= others,	αλλά	= but
άλλου	= another's,	αλλού	= somewhere else
αξία	= value,	άξια	= worth (fem.)
βάτος	= bramble,	βατός	= passable
γέρος	= old man,	γερός	= solid
διαφορά	= difference,	διάφορα	= different things
δίκη	= trial,	δική(μου)	= mine (fem.)
η	= fem. article,	ή	= or
κεφάλαιο	= capital, chapter,	κεφαλαίο	= uppercase
μόνος	= alone,	μονός	= odd, single
νομός	= county,	νόμος	= law
Ξάνθη	= Xanthi (town)	ξανθή	= blonde
Ξέρα	= bank, reef,	ξερά	= dry (plural, neutral)
όρος	= condition, mountain,	ορός	= serum
πόσο	= how much?	ποσό	= amount
πότε	= when?	ποτέ	= never

				b <sub>7</sub>	0	0	0	0	1	1	1	1
				b <sub>6</sub>	0	0	1	1	0	0	1	1
				b <sub>5</sub>	0	1	0	1	0	1	0	1
					0	1	2	3	4	5	6	7
b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>									
0	0	0	0	0	NUL (DLE)	TC <sub>7</sub>	SP	0		P		P
0	0	0	1	1	TC <sub>1</sub> (SOH)	DC <sub>1</sub>	!	1	A	Q	a	q
0	0	1	0	2	TC <sub>2</sub> (STX)	DC <sub>2</sub>	"	2	B	R	b	r
0	0	1	1	3	TC <sub>3</sub> (ETX)	DC <sub>3</sub>	£(#)	3	C	S	c	s
0	1	0	0	4	TC <sub>4</sub> (EOT)	DC <sub>4</sub>	\$ (¤)	4	D	T	d	t
0	1	0	1	5	TC <sub>5</sub> (ENQ)	TC <sub>8</sub> (NAK)	%	5	E	U	e	u
0	1	1	0	6	TC <sub>6</sub> (ACK)	TC <sub>9</sub> (SYN)	&	6	F	V	f	v
0	1	1	1	7	BEL	TC <sub>10</sub> (ETB)	'	7	G	W	g	w
1	0	0	0	8	FE <sub>0</sub> (BS)	CAN	(	8	H	X	h	x
1	0	0	1	9	FE <sub>1</sub> (HT)	EM	)	9	I	Y	i	y
1	0	1	0	10	FE <sub>2</sub> (LF)	SUB	*	:	J	Z	j	z
1	0	1	1	11	FE <sub>3</sub> (VT)	ESC	+	;	K		k	
1	1	0	0	12	FE <sub>4</sub> (FF)	IS <sub>4</sub> (FS)	,	<	L		l	
1	1	0	1	13	FE <sub>5</sub> (CR)	IS <sub>3</sub> (GS)	-	=	M		m	
1	1	1	0	14	SO	IS <sub>2</sub> (RS)	.	>	N	^	n	—
1	1	1	1	15	SI	IS <sub>1</sub> (US)	/	?	O	—	o	DEL

Code table

					b <sub>0</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>	b <sub>5</sub>	b <sub>6</sub>	b <sub>7</sub>	b <sub>8</sub>	b <sub>9</sub>	b <sub>10</sub>	b <sub>11</sub>	b <sub>12</sub>	b <sub>13</sub>	b <sub>14</sub>	b <sub>15</sub>
					0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
					0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
					0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
					0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
					00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
b <sub>0</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>																
0	0	0	0	00	SP	0	à	P	`	p	NBSP	°	À	Ð	à	ð				
0	0	0	1	01	!	1	A	Q	a	q	ı	±	Á	Ñ	á	ñ				
0	0	1	0	02	"	2	B	R	b	r	¢	²	Â	Ò	â	ò				
0	0	1	1	03	#	3	C	S	c	s	£	³	Ã	Ó	ã	ó				
0	1	0	0	04	\$	4	D	T	d	t	¤	´	Ä	Ô	ä	ô				
0	1	0	1	05	%	5	E	U	e	u	¥	µ	Å	Ö	å	ö				
0	1	1	0	06	&	6	F	V	f	v	¦	¶	Æ	Ø	æ	ø				
0	1	1	1	07	'	7	G	W	g	w	§	·	Ç	×	ç	÷				
1	0	0	0	08	(	8	H	X	h	x	¨	,	È	Ø	è	ø				
1	0	0	1	09	)	9	I	Y	i	y	©	¹	É	Ù	é	ù				
1	0	1	0	10	*	:	J	Z	j	z	ª	º	Ê	Ú	ê	ú				
1	0	1	1	11	+	;	K	L	k	l	«	»	Ë	Û	ë	û				
1	1	0	0	12	,	<	L	\	l		¬	¼	Ì	Ü	ì	ü				
1	1	0	1	13	-	=	M	]	m	}	SHY	½	Í	Ý	í	ý				
1	1	1	0	14	.	>	N	^	n	~	®	¾	Î	ß	î	þ				
1	1	1	1	15	/	?	O	_	o	¯	¿	Ï	ß	ï	ÿ					

