EURAMIS, the Platform of the EC Translator

Dimitri Theologitis

Abstract

Linguistic technology brings new tools to the desktop of the translator: full-text retrieval systems, terminological systems, translation memories and machine translation. These systems are now being integrated into a single, seamless workflow in a large organisation, the Translation Service of the European Commission. SdTVista is used for full-text search of reference documents; Euramis powerful aligner creates translation memories which are stored in a central Linguistic Resources Database; not-found sentences are automatically sent to EC-Systran machine translation; pertinent terminology is retrieved in batch mode from Eurodicautom. The resulting resources are brought together on the translator's workbench. Screen-shots and a demonstration complete the presentation.

Dimitri Theologitis

Born in Athens, civil engineer, specialised in integrated transportation systems and computers. Opted for a major change of career in 1984 when he joined the Translation Service of the European Commission. Responsible for the Rationalisation of Working Methods from 1990. In 1994 became head of unit *Development of Multilingual Computer Aids*, a multilingual team active in the technological modernisation of the Translation Service.

The Translation Service of the European Commission

Situated in Brussels and Luxembourg, the SdT (for Service de traduction) houses today some 1 200 translators, 100 linguistic support staff, approximately 100 management staff and close to 500 secretaries and assistants. The world's largest translation service, it produces well over one million pages per year, in a combination of in house, free lance and machine production.

Currently undergoing major technological modernisation. The main translation aids in use, being installed, or overhauled, include: the Eurodicautom termbase; the Euramis Linguistic Resources Database and search engines combined with translator's workbenches; Systran machine translation; and a document server with full text search and retrieval possibilities.

Dimitri Theologitis Translation Service, Development of Multilingual Computer Aids CCE JMO B2/21 L-2920 Luxembourg Tel: +352 4301 33632, fax: +352 4301 34069 E-mail:dimitrios.theologitis@sdt.cec.b



Euramis, the Platform of the EC Translator

Dimitri Theologitis Head of Unit "Development of Multilingual Computer Aids" European Commission Translation Service

Copenhagen, May 1997

SdT

Overview

- The SdT (Service de Traduction)
- Why Euramis
 - ... and not only Euramis
- Tools in the SdT
 - Descriptions
 - Demonstrations
- Workflow considerations
- Conclusions

SdT





EAMT Workshop, Copenhagen, May 1997







Basic Software in SdT

- Operating system:
 - Windows 3.1 (Windows NT as of 1998)
- Word-processing
 - Before: WordPerfect 5.2+ Win
 - Now: Winword 6.0
- Electronic Mail
 - All European Institutions on X.400 mail with Internet gateways

SdT

Major Systems

• Administrative

- Poetry, requesting department interface
- Suivi, internal translation follow-up
- Document Management

• Linguistic

- Euramis, Institution-wide Translation Memory
- Eurodicautom, Multiterm, Terminological
- Systran, Machine Translation

SdT

Why have Euramis at All?

What are the translator's needs?

- Reference documents in electronic form
- Pertinent terminology
- Already translated sentences = Translation memory

Model Based on Repetition





EAMT Workshop, Copenhagen, May 1997





Authoring Tools

- From spelling-checkers
- Through grammar and style checkers
 Important for Machine Translation
- To standardised multilingual document generators
- And, ultimately, to multilingual document drafting aids

SdT

Administrative Systems

- Systems that take a document from the author to the desk (the PC) of the translator
- Transmit all administrative information
 - Author
 - Department
 - Deadlines

	_					Poetry	r				_		
	<u>Profil A</u>	lde											
a 0 % c			I)E				3, 2, 1						
	D. TOTAL			STAND	0.000 m		**************************************	1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -					
	0 6		1996	1111		00		~		A 🚳) Lize	Links, des	
ղ Դեստերեն։ Աղլ				57. 77. 77. 77. 77. 77. 77. 77. 77. 77.					20-077				i di circi di di circi All'ordina di circi
	hgihigh							o a more and a				-	
C America	: 02			- 19 C				aladi Turk				. · · ·	ा रहे इ.स.
									dure orale	, réunida	Commissi	on, le	
a se se de la company				() yeshirdini (an ind constanting a	2		12 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	same in the second				n), <u>2.466</u>
tery Com			¥		000-20 CT 213								
	DELSART	· ·	5		Sec. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.			SANT	ER Jacqu	**	× 93900	BREY 1	
	engles (and a straight and a The straight and a straight an	2000 - C. C.	i san ja						50 9 92
heteler (1995) Mitoria	÷			- M			wind w	<u>~ * * </u>	Protect	÷.:			
	FR 🗮				277) 2007 - 19								
	L.		i-to-	ž	and the second sec			386 - P. (Sel)		Ì.			
	5631_2 			≈ 3.05 m \odot	200760 R.1522				anten in				
	nin Ministria Ministria Ministria					Enter In				R UR			
		· .	1947 - C., M	ж., : <u>-</u> с.юй	71387	mean State			L'ARRAN	<u></u>			0 98 57
		م به آسینیم در این تونند اساسی		×	xxxxxx	درچ زید در در							
Langues	AU	DA	FR	IT	DE	EL	EN	ES	FI	NL	PT	SV	
	***	ww	ww	ww	ww	ww	ww	ww	ww	ww	ww	ww	
651	12-11-96	14-12-97	13-12-96	12-11-96	12-11-96	12-11-96	12-11-96	12-11-96	12-11-96	12-11-96	12-11-96	12-11-96	
	å <u> </u>	23	24	<u></u>	······································	· ·	1				•••••••••••	-	1
ACC Ref.	DELSAR			-	ter of the second	State of the second second		}		DELSAR	MARTIG		*-1%% <_0000000
	GL.		۲۰۰۹: همچنه ا	1 Concerne		Ban Albin Gan Albin S		87 - 1					
Stanformal State			· · · · · · · · · · · · · · · · · · ·	\$	an man	Con Strain	A				1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1		463.9

The Document Server

- Electronic document handling
- (More than 2/3 of the originals arrive in electronic format)
- Full-text indexing (3 years on line)
 Using "SearchTools" by Fultext (Fulcrum)
- Can display on screen two parallel language versions of same document

		View	Help					
					ss of d			
							<u>al Hara</u> na (al Angele) Na Sensa (al Angele)	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	OF.		Yeer:	Doc 1	Care Care Care Care Care Care Care Care	32. 🕂 🖌	12.45	
	SP.		20	****				
E E				1.2.2%			2-2-2-2-4658	
Addition	nal attribu	÷						
CA CALL COMMON AND		1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 - 1999 -	radit in mirzyńska	d:		A Crewin Y A chart of a com		
			1497-0-10-10-10-10-10-10-10-10-10-10-10-10-1	######################################	1 Mar Block Control of the State of the Stat	aur mille Merken werden die stelle die stelle auf		
S - A	ta L	reisen e						
	ບ່ານຈານບໍ່ມີມີມີມີບໍ່ບໍ່ມີມີບໍ່ບໍ່ມີມີບໍ່ມີບໍ່	woioiùoñôiiic		يسفيك شفشي				
			MA 142.2	2	and the second	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1		286 21
[<u></u>	
Comm	mication (ommis si	on sut	les suites données aux	avis el résolutions	adoptés	
Comm	mication (ommis si	on sut		avis et résolutions	adoptés	
Comm par le	nication Parloment	europ	ommis si	on sur do la	les suites données aux		adoptés	
Comm par le	mication (europ	ommis si	on sur do la	les suites données aux session de juin I 1995]		adoptés	
Commipar le	nication Parlement	europi	ommissi éen lors Disso	on sus de la	les suites données aux session de juin 1995 session de juin 1995			
Commipar le	nication (Parloment	europi	ommissi ćen lors lizecij lizecij lizecij Vie	on sus de la	les suites données aux session de juin 1995] Session de juin 1995] Session de juin 1995] Session Session 1995]			
Comme par le	nication Parloment	europi	ommissi éen lors loss Vé loss Vé lise (j)	on sur de la A (201)	les suites données aux session de juin I 1995 session de juin de juin I 1995 session de juin I 1995 session de juin I 1995 session de juin de	National States	Second Second	
Commipar le	nication Parlement IAUS DE TRASE TRASE	europi 4 1310 1335 (ommissi ćen lors lizecij lizecij lizecij Vie	on sus do la	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001		5 Size 5 105 100 pt 217688	
Comme par le - Drostan Note Face	Mication Parlement MANNE TRASP TRASP	europi 4 3260 (1996 (1996 (in lors fen lors lors loss Ve loss Ve loss Ve	on sus do la Cal Da Da O1 Cal	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001	that the sen of hits for Ida de el en es fifrit n	5ino 5ino 217688 pt 172436	
Comme par le - Drostan Note Face	Mication Parlement MANNE TRASP TRASP	europi 4 3260 (1996 (1996 (tommissi éen lors köst Ve tost Ve 2027 00 292 00	on sus do la Cal Da Da O1 Cal	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001	l (gl. in close a clos	5ino 5ino 217688 pt 172436	
Comme par le - Drostan Note Face	Mication Parlement MANNE TRASP TRASP	europi 4 3260 (1996 (1996 (tommissi éen lors köst Ve tost Ve 2027 00 292 00	on sus do la Cal Da Da O1 Cal	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001	l (gl. in close a clos	5ino 5ino 217688 pt 172436	
Comme par le - Drostan Note Face	Mication Parlement MANNE TRASP TRASP	europi 4 3260 (1996 (1996 (tommissi éen lors köst Ve tost Ve 2027 00 292 00	on sus do la Cal Da Da O1 Cal	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001	l (gl. in close a clos	5ino 5ino 217688 pt 172436	
Comme par le - Drostan Note Face	Mication Parlement MANNE TRASP TRASP	europi 4 3260 (1996 (1996 (tommissi éen lors köst Ve tost Ve 2027 00 292 00	on sus do la Cal Da Da O1 Cal	les suites données aux session de juin I 1995 Suite Suites Inter Senà Van FR/SP/96/02770001	l (gl. in close a clos	5ino 5ino 217688 pt 172436	



Document Server Spin-offs

- Powerful terminological tool
 - Terminology "On-demand"
 - Especially with parallel scrolling
- Provides the material for all linguistic systems
 - Added-value with alignment
 - At term level (terminology)
 - At sentence level (translation memory)

SdT

Document Comparison

- Docucomp (by Mastersoft)
- Simple (\$ 99) software
- Ideal to compare two versions of the same document
- Marks graphically in great detail
 - Added text
 - Deleted text
 - Moved text

SdT

The Linguistic Systems

- Terminological
 - Eurodicautom
 - MultiTerm
- Translation Memories
 - Euramis
- Machine Translation
 - Systran

SdT

Eurodicautom

- Central terminological database
 - 800.000 multilingual entries
 - 300.000 abbreviations
- Available now via a WWW (World-Wide-Web) interface
- For the moment, free connection via ECHO, (European Community Host Organisation) http://www2.echo.lu/edic



EC-Systran

- In the Commission since 1976
- Usage explosion after introduction of simple e-mail access
- Currently available in 16 language pairs
 - Of which 8 give satisfactory results
- 220.000 pages in 1996
 - Of which 80.000 requested by SdT

_		_				lail Sd1						2
<u> </u>	dit <u>M</u> ail (Options	Window	<u>и Н</u> е	sip							
Sak					Digional Contract	<u>s</u>	. EI 800	:e				
			Adıl	ress [3ook - S	SHARED.A	вк	·····	······································			
П.	SIDE SYSTRA	NENGII	SH -> POF	31116U	n se ×	363 <u>78</u> JMt	1 <u>_A27</u>	4 <u>1</u> H	02			1
	GEORGE MAN		Detaile.	100	97. 4	M Continu						
						• 60	. 194. 					
t I.S.	SYSTRAN		·/////////////////////////////////////	-> F	RENC		X	363		JMO	A	帯
AB	9				ERMA		x	363		JMO	A	
	SYSTRAN	I ENGI	LISH -	-> G	REEK		x	363	78	JMO	A	
	SYSTRA	I ENGI	LISH -	-> I	TALI	AN	x	363	78	JMO	A	15.62
GH	SYSTEAD	ENG	LISH -	> 1	ortu	GUESE	Х	363	78	JMO	A. and	
LUK.	SYSTRA	I ENGI	LISH -	-> s	PANI	SH	X	363	78	JMO	A	
8 · · · · · · · · · · · · · · · · · · ·	SYSTRAN	I FREI	VCH ->	עם ≺	TCH		х	363	78	JMO	A	
14	SYSTRAN	I FREI	VCH ->	► EN	GLIS	н	х	363	78	JMO	A	
	SYSTRAN	I FREI	исн ->	≻ GE	RMAN	r	х	363	78	JMO	A2	2
	SYSTRAN		VCH ->	> IT	ALIA	N	x	363	78	JMO	A:	
F Ad	an Bost	· · · · · · · · · · · · · · · · · · ·										
15.						Carrier 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990 - 1990				¥,		2



EC-Systran language couples

• English into

- Dutch
- French
- German
- Greek
- Italian
- Portuguese
- Spanish
- Spanish into
 - English
 - French

- French into
 - Dutch
 - English
 - German
 - Italian
 - Spanish
- German into
 - English
 - French

SdT

European Advances Multilingual Information System



DG XIII - Translation Service