

ESF Exploratory Workshop  
“Constructing bilingual dictionaries with  
special emphasis on lesser used languages”

Thessaloniki, 25th-27th August, 2003

## SCIENTIFIC REPORT

### 1. Abstract

The workshop, which ran from 25th to 27th August 2003 in Thessaloniki, organised by Gaston Goss (LLI, Université Paris 13) and Anna Anastasiadis-Symeonidis (Aristotle University, Thessaloniki), aimed at examining what new information should be built in to electronic dictionaries, in particular with a view to expanding to lesser-used European languages. This is needed since present-day electronic dictionaries are designed to cater for the needs of the wide-spread vehicular languages (English, French and German). Current descriptions stress main grammatical categories (verb, noun, especially compound nouns, adjectives), thereby neglecting the more strictly grammatical elements (lexicalised determiners, prepositions, adverbials, connectors, aspect indicators).

The aim was thus to work out new types of dictionaries giving more accurate and better articulated descriptions with regard to syntactic and semantic problems. This information should be as flexible as possible in order to allow transfer from a vehicular language to less widely used languages with differing internal structures. The outcome is a theoretical tool which generates analyses compatible with widely varying syntactic systems.

### 2. Scientific content

The thirteen papers presented can be grouped around three main themes

#### a) The state of the art

The discussion revolved around the part bilingual dictionaries play in the defence of European language diversity. Emphasis was put on the importance of appropriate bilingual tools for different needs of communication, in particular those of translation. One paper presented the present situation of translation in the various European countries. The existing systems of automatic translation were presented, together with the methodological options on which they are based.

## b) Linguistic aspects

A series of papers pinpointed theoretical and pragmatic improvements to be made on electronic dictionaries. A new approach to bilingual dictionaries was put forward, based on the Framenet system of Fillmore. Other papers presented the role of morphology in making a French-Hungarian dictionary, the interaction of collocations in writing entries and the interest and feasibility of introducing figures into a formalised dictionary. New analyses were put forward to account for Greek compound words. A new view of connectors analyses them as second order predicates whose arguments are the main clause and the subordinate.

## c) Dictionaries and automatic data processing

A close examination was made of how parallel corpora can be used to find translation units and how these resources can contribute to making new dictionaries. One presentation investigated how the Intex system can be applied to Greek and the subsequent handling of flexional languages by the system. The Hub-and-spoke model and its application to electronic dictionaries was also discussed.

To conclude the conference, Gaston Gross and Anna Anastasiadis-Symeonidis summed up the achievements of the three days' work, stressing both theoretical aspects and lexicographical practices which can federate innovative projects in the field of making electronic dictionaries.

## 3. Evaluation of the results

In spite of much work achieved over the last twenty years in the field of automatic language processing, the conference shows that much remains to be done in order to incorporate the necessary information for lesser-used languages. More particularly, it was shown that more accuracy is needed in the description of grammatical categories, as their status varies according to language. Because of these specificities, bilingual dictionaries (and **not** trilingual) should be made, as they must be constructed according to the specificities of the target language.

Since the description must be very detailed, it is not enough to amass even a large nomenclature. The requirements of automatic translation imply the formal expression of lexical rules whenever this is possible. To do this, a dictionary must contain all morphological, syntactic and semantic regularities. Thus, the opposition between dictionaries and syntax becomes obsolete.

Other categories of information must be included systematically. German and Romance language dictionaries tend to lack adequate description of aspect, which impedes application to the Slavic languages, which incorporate aspect morphologically. For German and Romance languages, which do not express aspect morphologically, preliminary work must be done on which lexical units convey this information.

#### **4. Future prospects**

This conference shows that there is a strong consensus on approaches to making electronic dictionaries. It was demonstrated that the descriptions are based on semantic frameworks (Fillmore's works). This perspective has been developed in various European centres of computational linguistics. It turns out that other semantic approaches, especially the object class approach developed at Paris 13 University and adopted in several centres (Munich (Germany), Seoul (Korea), Barcelona (Spain), Thessaloniki (Greece)), are quite close from a theoretical viewpoint. One possible project is to confront these approaches systematically. A systematic description could also be made of connectors using semantic classes, a project for which several participants at the conference expressed great interest, so that a ESF network could be formed. It is also planned to organise a summer school on automatic language processing.

## **5. Final Programme**

### **Monday 25<sup>th</sup> August : The State of the Art**

09.30-10.00 : Bernard Quemada :  
The role of bilingual dictionaries in defence of linguistic pluralism

10.00-10.30 : André Clas :  
The world of translation today

11.00-11.30 : Zoé Gavriilidou and Xavier Blanco :  
Translation systems today

11.30 –12;00 : Sue Atkins :  
A new approach to bilingual dictionaries

### **Tuesday 26th August : Linguistic aspects**

09.00- -9.30 : Ferenc Kiefer :  
The role of morphology in the French-Hungarian dictionary

09.30-10.00 : Eleni Galiotou :  
Processing Greek compounds

10.00-10.30 : Ulrich Heid :  
Interaction between collocations and Syntax in a bilingual dictionary

11.00-11.30 : Michele Prandi :  
Figures of speech in a computerised dictionary

11.00-12.00 : Gaston Gross :  
The treatment of connectors

## **Wednesday 27th August**

- 9.30-10.00 : Wolfgang Teubert :  
Parallel corpora, translation units and their equivalents : the  
future of bilingual lexicography
- 10.00-10.30 : Anna Anastasiadis-Symeonidis :  
The modern Greek INTEX System
- 11.00-11.30 : Willy Martin :  
Computerised Bilingual Dictionaries and the Hub-and-Spoke  
Model
- 11.30-12.00 : Gaston Gross and Anna Anastasiadis-Symeonidis : Working out  
a common theoretical framework to produce new electronic  
dictionaries.

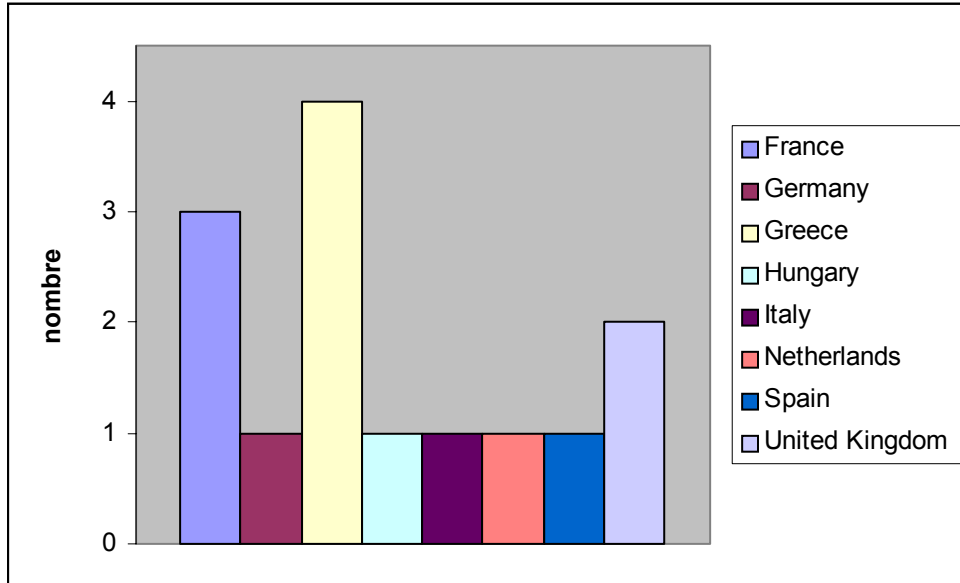
## 6. Final List of participants

Name		Affiliation	Telephone	Fax	E-mail
GROSS Gaston <i>[Convenor]</i>	FR	Laboratoire de Linguistique Informatique Université Paris 13 99, avenue J.B. Clément F-93430 Villetaneuse	01 49 40 35 44 ou 38 56	01 49 40 40 99	<a href="mailto:gross@lli.univ-paris13.fr">gross@lli.univ-paris13.fr</a>
ANASTASSIADIS Anna <i>[Co-Convenor]</i>	GR	Univ. de Thessalonique Fac des Lettres Dép. de linguistique G – 540-06 Thessalonique	B. 00302310991670 D. 00302310911756	00302310997015	<a href="mailto:ansym@lit.auth.gr">ansym@lit.auth.gr</a>
ATKINS Sue	UK	Coach House South Malling Deanery, Church Lane Lewes Sussex BN7 2JA UK	0044(0)12734762 30	0044(0)127348351 3	<a href="mailto:b.t.s.atkins@itri.brighton.ac.uk">b.t.s.atkins@itri.brighton.ac.uk</a>
BLANCO Xavier	SP	Universitat Autònoma de Barcelona Departament de Filologia Francesa i Romanica Edifici B – Campus de la UAB 08193 Bellaterra (Cerdanyola del Vallès) Barcelona Spain	0034 93581 14 10	0034 935812001	<a href="mailto:xblanco@seneca.uab.es">xblanco@seneca.uab.es</a>
CLAS André	CA	7405 Maynard H3R 3B3 Montréal Canada	00 15147331295 ou 0015143437047	0015143432284	<a href="mailto:clasand@ere.umontreal.ca">clasand@ere.umontreal.ca</a>
GALIOTOU Eleni	GR	TEI d'Athènes Département d'informatique Ag. Spyridona G-122 10 Athènes	00 30 210 7275230	0030 210 72 75214	<a href="mailto:egali@di.uoa.gr">egali@di.uoa.gr</a>
GAVRIILIDOU Zoé	GR	Universite Democrite de Thrace Département de Sciences d'éducation et enseignement prescolaire Nea Hilli G-69100 Alexandroupolis	00 30 2510227766	00 30 2510620915	<a href="mailto:zoe@her.forthnet.gr">zoe@her.forthnet.gr</a>

HEID Ulrich	DE	Universität Stuttgart Institut für Maschinelle Sprachverarbeitung, Computerlinguistik Azenbergstrasse 12 7000 Stuttgart-1 Allemagne	0049(711)1211373	0049(711)1211366	<a href="mailto:uli@adler.philosophie.uni-stuttgart.de">uli@adler.philosophie.uni-stuttgart.de</a>
KIEFER Ferenc	HU	Research Institute for Linguistics Hungarian Academy of Sciences Benczur utca 33 H-1068 Budapest	003613510413	003613229297	<a href="mailto:kiefer@nytud.hu">kiefer@nytud.hu</a>
MARTIN Willy	NL	Universit�e Libre d'Amsterdam Vrije Universiteit Amsterdam Depart :Lexicology De Boelelaan 1105 NL-1081 HV Amsterdam	0003120 4446451 ou 66	0003120 661 30 54	<a href="mailto:martin_w@let.vu.nl">martin_w@let.vu.nl</a>
PRANDI Michele	IT	Universit�e de Pavie D�epartement de linguistique 65 strada nuova I-27100 Pavie	0039 0382 504484	0039 0382 504487	<a href="mailto:mprandi@pop.unipv.it">mprandi@pop.unipv.it</a>
QUEMADA Bernard	FR	4, rue Beaurepaire F-75010 Paris	01 42 08 13 55		
TEUBERT Wolfgang	UK	Centre for Corpus Linguistics Department of English Birmingham B15 2TT UK	0044(0)12141457 00	0044(0)121414605 3	<a href="mailto:teubertw@hhs.bham.ac.uk">teubertw@hhs.bham.ac.uk</a>

## 7. Statistical information on participants

Countries of origin :



Age Structure

