

REPORT

ENeL TRAINING SCHOOL 2015

Standard tools and methods for retro-digitising dictionaries

Faculdade de Ciências Sociais e Humanas
Universidade Nova de Lisboa

6 – 10 July 2015, Lisbon, Portugal

Rute Costa
Training Manager

1. Introduction	3
1.2. Objectives of WG2: RETRO-DIGITIZED DICTIONARIES	3
2. Objectives and Program of the Training School	4
2.1. Objectives of the training school	4
2.2. Schedule	4
2.3. Methods	4
2.4. Class Materials	5
2.5. Software	5
3. Training Team	5
4. Trainees	5
4.1. Selected Candidates	6
4.2. Waiting list	9
4.3. Excluded Candidates	9
4.4. Non COST Countries	10
5. Facilities	10
5.1. Host	10
5.2. Accommodation	10
6. Budget	11
Annex 1: Certificate of participation	12
Annex 2: Attendance List	13

1. Introduction

ENeL, the European Network of e-Lexicography has been established to increase, coordinate and harmonise european research in the field of e-lexicography and to make authoritative information on the languages of Europe easily accessible.

ENeL is subdivided in 4 Working Groups:

- WG1 Integrated interface to European dictionary content;
- WG2 Retro-digitized dictionaries
- WG3 Innovative e-dictionaries
- WG4 Lexicography and lexicology from a pan-European perspective.

One of the main aims of ENeL is to organise a series of three Training Schools designed to underpin the activities of WG 2, WG3 and WG4, whereas WG1 will get most of its scientific input from these three WGs.

The first ENeL Training School was held in Lisbon, from the 6th to the 10th of July, hosted by the Faculdade de Ciências Sociais e Humanas of the Universidade Nova de Lisboa to reinforce the activities of the WG2.

1.2. Objectives of WG2: RETRO-DIGITIZED DICTIONARIES

Working group 2 intends to set up guidelines and standards for turning paper dictionaries into digital format. Work will be carried out on:

- standards for encoding of information and the description of relevant information categories for print dictionaries
- software for the conversion of physical lay-out information to logical information
- the investigation of relevant information categories to be added to the dictionary in order to make the dictionary content more readily accessible and interoperable
- the development of a workplan for digitisation, including parameters necessary for estimating costs
- the investigation of the possible use of dictionary content for computational linguistic applications
- organising a training school on standard tools and methods for retro-digitising dictionaries in 2015.

2. Objectives and Program of the Training School

2.1. Objectives of the training school

By retro-digitisation it's meant the process of converting existing print dictionaries into digital publications. The idea is to introduce the trainees into each step of the process of retro-digitisation by means of theoretical courses and hands-on sessions.

- Introduction (presentation of different successful retro-digitisation projects)
- How to produce a digital copy of dictionaries (digitisation from paper to image to text)?
- How to encode and enrich dictionary data (requirements, standards and tools)?
- How to manage retro-digitisation projects (project design, resources, time schedule, budget)?

2.2. Schedule

SCHEDULE

Time	Monday	Tuesday	Wednesday	Thursday	Friday
9 – 12.30	Introduction (overview, pecha kucha, expectations)	<ul style="list-style-type: none"> - Digitisation instructions and requirements - structure analysis 	<ul style="list-style-type: none"> - Data Modeling - XML representation (XML, TEI) 	Conversion and Enrichment	<ul style="list-style-type: none"> - Retrieval/Dictionary engine - iLex comunico publication platform - Project management
14 – 17	<ul style="list-style-type: none"> - Overview of European dictionaries - presentation of different projects 	<ul style="list-style-type: none"> - Defining requirements and digitisation instructions - OCR practical session 	Practical encoding session I	Practical encoding session II	Presentation of results
19.30		common dinner (university)			

2.3. Methods

Classes combined theoretical instructions on standard tools and methods for retro-digitising dictionaries and practical lessons giving participants the opportunity to show their work on their own material and to learn which steps are necessary to transform a

printed dictionary into a marked-up dictionary as a basis for a database or an online-publication.

On the first day all the participants gave a short presentation about themselves and the dictionaries that they were working on using a special format “Pecha Kucha”. On the last day some of the participants made presentations of their work exemplifying with samples doubts and possible resolutions. This way not only the trainers and the colleagues could participate and suggest other solutions, modifications, but also be aware of possible problems in their future works.

2.4. Class Materials

Class materials (pdf and articles), Pecha Kucha presentations of all participants, trainers’ presentations, instructions of how to use some tools, guidelines, final presentations of student’s works

All the class material is available on a Goggle Drive folder/group created for the training school, but kept online and open to all the participants for future interaction (future collaborations, clarify any doubts, ...)

2.5. Software

During the classes several tools and programs for OCR (Optical Character Recognition), HTR (Handwritten Text Recognition) and Keying were introduced to the participants: ABBY FineReader, Tesseract, Ocrpus, Omnipage, Aletheia, Transkribus, TEI, Oxigen XML, iLex.

3. Training Team (alphabetical order)

The training team were constituted be the following trainers:

Katrien Depuydt	Institute for Dutch Lexicology (INL)	Netherlands
Jesse de Does	Institute for Dutch Lexicology (INL)	Netherlands
Jens Erlandsen	Erlandsen Media Publishing ApS	Denmark
Vera Hildenbrandt	Trier Center for Digital Humanities	Germany
Toma Tasovac	Belgrade Center for Digital Humanities	Serbia

4. Trainees

On the ENeL website¹ the information were given on “who can apply” and “how to apply”. In Lisbon, the 9th of April, the team had a meeting to analyse the candidates applications (brief CV and motivation letter) and we came up with the following results:

Number of candidate applications:	44
Selected trainees:	29 ²
Waiting list:	6
Excluded candidates:	5
Non cost country:	4

4.1. Selected Candidates

List of the selected candidates:

Nr	Trainee	Institution	Country	Gender
1	BARBOSA Silvia <i>silviapaisbarbosa@gmail.com</i>	Faculty of Social Sciences and Humanities, Universidade Nova de Lisbon	Portugal	F
2	BENKO Vladimir <i>vladob@juls.savba.sk</i>	Ľ. Štúr Institute of Linguistics Bratislava	Slovakia	M
3	BILIŃSKA Joanna <i>j.bilinska@uw.edu.pl</i>	Department of Formal Linguistics University of Warsaw	Poland	F
4	BRAČ Ivana <i>ibrac@ihjj.hr</i>	Institute of Croatian Language and Linguistics Zagreb	Croatia	F
5	CHMIELEWSKA Jagoda <i>chmielewska.jagoda@gmail.com</i>	Department of Medieval Latin, Institute of Polish Language Polish Academy of Science	Poland	F
6	FERNANDEZ-PAMPILLON Ana <i>apampi@filol.ucm.es</i>	School of Philology at the Universidad Complutense of Madrid (UCM)	Spain	F
7	HEBAL-JEZIERSKA Milena <i>m.hebal-jeziarska@uw.edu.pl</i>	Institute of Southern and Western Slavonic Studies University of Warsaw	Poland	F
8	KAJANOVÁ Michaela <i>michaelak@juls.savba.sk</i>	Ľ. Štúr Institute of Linguistics Bratislava	Slovakia	F
9	KARASIMOS N. Athanasios <i>akarasimos@gmail.com</i>	Research Center of Scientific Terms and Neologisms, Academy of Athens	Greece	M
10	KOZIARSKI Mirosław <i>miroslaw.kozarski@gmail.com</i>	Adam Mickiewicz University Poznań	Poland	M

¹ <http://www.elexicography.eu/events/training-schools/lisbon-2015/>

² One of the selected student couldn't attend the Summer School. She couldn't be replaced because the organizer were informed too late.

11	KRAMARIĆ Martina <i>mkramar@ihjj.hr</i>	Institute of Croatian Language and Linguistics Zagreb	Croatia	F
12	LINDEMANN David <i>david.lindemann@ehu.eus</i>	UPV-EHU University of the Basque Country	Spain	M
13	MANOJLOVIĆ Sonja <i>sonjaiana.m@gmail.com</i>	Institute for Serbian Language SASA Belgrade	Serbia	F
14	MANOLESSOU Ioanna <i>manolessou@Academyofathens.gr</i>	Research Center for Modern Greek Dialects-ILNE Academy of Athens	Greece	F
15	MOLOCHIEVA Zarina <i>Zarina.Molochieva@sprachlit.uni- regensburg.de</i>	General and Comparative Linguistics University of Regensburg	Germany	F
16	OSTROŠKI ANIĆ Ana <i>ana.ostroski@gmail.com</i>	Institute of Croatian Language and Linguistics Zagreb	Croatia	F
17	PATRAȘCU Mădălin Ionel <i>maddlin.patrascu@gmail.com</i>	Institute of Romanian Philology Romanian Academy, Bucharest	Romania	M
18	PIASECKA Marta <i>piasecka.ma@gmail.com</i>	Institute of Applied Polish Studies University of Warsaw	Poland	F
19	POO Nancy <i>nancyannpoo@googlemail.com</i>	Christian-Albrechts-University Kiel	Germany	F
20	PYTLOWANY Anna <i>a.pytlowany@gmail.com</i>	University of Amsterdam	Ireland	F
21	SARNECKI Mateusz <i>m.sarnecki@uw.edu.pl</i>	Institute of English Studies University of Warsaw	Poland	M
22	SCHULTZE Ursula <i>s2urschu@uni-trier.de</i>	Trier Center for Digital Humanities Trier University	Germany	F
23	ŞİRİN Fahrettin <i>fsirin@gmail.com</i>	Bielefeld University	Germany	M
24	STEPANOVIĆ Željko <i>zhelhko@gmail.com</i>	Serbian Language Institute SASA Belgrade	Serbia	M
25	ŠTRKALJ DESPOT Kristina <i>kdespot@ihjj.hr</i>	Institute of Croatian Language and Linguistics	Croatia	F
26	TEŠIĆ Ana <i>tesicana@gmail.com</i>	Serbian Language Institute of SASA Belgrade	Serbia	F
27	TRESSY, Arts ³		Netherlands	F
28	VACALOPOULOU Anna <i>avacalop@ilsp.athena-innovation.gr</i>	Institute for Language and Speech Processing Athens	Greece	F
29	WITZLACK MAKAREVICH Alena <i>witzlack@googlemail.com</i>	Christian-Albrechts-University University Kiel	Germany	F

In conclusion: 28 trainees attended the Lisbon Summer School:

20 Female / 8 Male

10 different countries: Portugal (1), Slovakia (2), Poland (6), Croatia (4), Spain (2), Greece (3), Serbia (3), Germany (5), Romania (1) and Ireland (1)

25 different institutions, as follows:

³ Arts Tressy informed that she couldn't attend the Summer School



Nr	Institution	Country
1	Institute of Croatian Language and Linguistics Zagreb	Croatia
2	Christian-Albrechts-University Kiel	Germany
3	General and Comparative Linguistics University of Regensburg	Germany
4	Trier Center for Digital Humanities Trier University	Germany
5	Research Center for Modern Greek Dialects-ILNE Academy of Athens	Greece
6	Research Center of Scientific Terms and Neologisms, Academy of Athens	Greece
7	University of Amsterdam	Ireland
8	Adam Mickiewicz University Poznań	Poland
9	Department of Formal Linguistics University of Warsaw	Poland
10	Department of Medieval Latin, Institute of Polish Language Polish Academy of Science	Poland
11	Institute of Applied Polish Studies University of Warsaw	Poland
12	Institute of English Studies University of Warsaw	Poland
13	Institute of Southern and Western Slavonic Studies University of Warsaw	Poland
14	Faculty of Social Sciences and Humanities, Universidade Nova de Lisbon	Portugal
15	Institute of Romanian Philology Romanian Academy, Bucharest	Romania
16	Institute for Serbian Language SASA Belgrade	Serbia
17	Ľ. Štúr Institute of Linguistics Bratislava	Slovakia
18	School of Philology at the Universidad Complutense of Madrid (UCM)	Spain
19	UPV-EHU University of the Basque Country	Spain
20	Bielefeld University	Germany
21	Serbian Language Institute SASA Belgrade	Serbia
22	Institute of Croatian Language and Linguistics	Croatia
23	Serbian Language Institute of SASA Belgrade	Serbia

24	Institute for Language and Speech Processing Athens	Greece
25	Christian-Albrechts-University University Kiel	Germany

4.2. Waiting list

Form the 44 candidates, the CVs and the motivation letters fulfill the prerequisites, but couldn't be accepted, since there were no place left. But we ranked them, in case there would be a quitting, she/he could fill the vacancy. The list is a ranking:

Nr	Candidate	Country	Gender
1	Vasiliki Saloustrou	Greece	M
2	Petra Bago	Croatia	F
3	Marlena Secczek	Poland	F
4	Piero Tassinari	UK	M
5	Johannes Förster	Germany	M
6	Molgorzata Milawska	Poland	F

4.3. Excluded Candidates

Form the 44 candidates, the CVs and the motivation letters of the 5 candidates didn't fulfill the prerequisites, so they had been excluded. The list is in alphabetical order:

Nr	Candidate	Country	Gender
1	Izabela Cedro	Belgium	F
2	Jovana Rupar	Serbia	F
3	Michal Kozak	Poland	M
4	Tijana Markevie	Serbia	F
5	Violeta Stojici	Serbia	F



4.4. Non COST Countries

This group of candidates couldn't be accepted since the countries they come from are not eligible.

The list is in alphabetical order:

Nr	Candidate	Country	Gender
1	Ali Safari	Iran	M
2	Solomiya Buk	Ukraine	F
3	Tatia Mtvarelidze	Georgia	F
4	Valentyna Kotenko	Ukraine	F

5. Facilities

5.1. Host

Funded by COST Action IS 1305 the Faculdade de Ciências Sociais e Humanas of the Universidade Nova de Lisboa hosted the first 'European Network of e-Lexicography' Training School, in Lisbon.

For the whole duration of the Training School, the Faculty provided for free 2 classrooms: 1 classroom with 30 computers, where the IT people installed the requested software's and one classroom to hold the group meetings, when needed.

5.2. Accommodation

The University provided accommodation in a Student Residence:

11 students chose to stay at the residence. The other found other solutions by their own

Accommodation - student residence

Hotels

STUDENT RESIDENCE ALFREDO DE SOUSA

http://sas.unl.pt/alojamento/ras/alfredo-de-sousa-halls-of-residence?set_language=en

Location:

Campus de Campolide
1099-032 Lisboa
T: 0351 213 837 000

Note : 15-20 minutes walking distance

Pre-reservation: 5-11 July 2015 (6 nights)	Beds	Daily rate	Rate for 6 nights*
15 TWIN ROOM with shared bathroom	30	20.00 €	110.00 €
2 SINGLE ROOM with shared bathroom	2	25.00 €	137.50 €

* Staying for 5 days or more – each day as from the 5th day: 25% discount

Bedding and towels supplied.

Breakfast NOT included. The residence comprises shared kitchens and a coffee shop.

Wireless Internet Connection

RESERVATION – DEADLINE: 20 April 2015

Please fill the FORM:

http://sas.unl.pt/alojamento/pedido-de-alojamento-temporario?set_language=en

In the "Comments" area you must reference "ENeL 2015".

After submitting the form a confirmation email will be sent to you.

6. Budget

The budget for the Training School was set at 25.000 euro. This budget was meant for 5 trainers and 30 trainees and for accommodation. Fortunately there was no charge for the training facilities, which saved some money for other events.

Travel and accomodation Trainers	5.989,86 EUR
Grants Trainees	13.800,00 EUR
Locas Organising Support	540,00 EUR
TOTAL	20.329,86 EUR

ANNEX 1 : Certificate of Participation

CERTIFICATE OF PARTICIPATION

Training School

STANDARD TOOLS AND METHODS FOR RETRO-DIGITISING DICTIONARIES

EUROPEAN NETWORK OF e-LEXICOGRAPHY – ENeL

Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa

Lisbon 6–10 July 2015

Name participant:

Institute:

Country:

Rute Costa

Training School Manager

Vera Hildenbrandt

**Chair of Working Group 2
*Retro-digitized dictionaries***

ANNEX 2: Attendance List