Leiden, 19 March 2014

# Subject │ Minutes of the First Meeting of Working Group 3 – innovative e-dictionaries - of COST Action IS1305 “European Network of e-Lexicography (ENeL)”

**at Leiden University, Sterrenwachtlaan, Leiden (Netherlands)**

**on 16 January 2014, from 11h15 till 13h00**

1. **Welcome to participants and adoption of the agenda**

Simon KREK (chair of WG3) opens the WG meeting and welcomes the participants.

The draft agenda of the meeting as found in **Annex 1** was approved without changes or additions by the members of the WG. We start with some practicalities.

After the meeting a mailing list will be set up. New members are still welcome. For financial reasons, it may be necessary to only reimburse participants who are active.

The Chair Simon KREK and the vice-Chair Carole TIBERIUS introduce themselves, followed by a presentation in which an overview is given of the different lexicographical projects that are represented within the WG. The overview can be found in **Annex 2** and was compiled on the basis of information that the participants were asked to provide before the meeting.

Simon KREK then gives an overview of what WG3 is about (see **Annex 3**).

Some thought was given to what we mean by **born-digital**. To set the boundaries of WG3, Simon KREK defines born-digital as those dictionaries which leave out the constraints of the paper format. General characteristics are that dictionaries should be conceptualised as digital, more information can be put in and that they are visualised in the online medium. Dictionaries that are born in paper, but are extended in the digital medium in some way are also included.

It is suggested that the e in e-lexicography should be dropped. This seems difficult as it is part of the name of the network. Within WG3 we can do this. No decision is made.

Simon KREK presents the general background of COST Action IS1305 and WG3. In WG3 we are less concerned with point a and b, but more with c. (see **Annex 3**)

*c) In the past few years, innovative electronic dictionaries have been created that no longer resemble traditional paper dictionaries but try to fully exploit the new possibilities of the digital medium.*

*Though serious attempts have already been made at embedding electronic lexicography into a theoretical framework, a new research paradigm and common standards for electronic lexicography are still lacking.*

*And so are common standards and cooperation for the interlinking of the content of digitized dictionaries and innovative e-dictionaries.*

We would like to get information on what the members of WG3 would like to do outside paper. We would like to get an insight in how different projects do different things. We would like to link information between different dictionaries.

Past few years from the point c) can be interpreted as the past 5 to 8 years.

Concerning the scientific focus, the interface with computational linguistics is important.

*b) mapping current and possible future trends for the creation of born-digital dictionaries, focusing on the latest developments in e-lexicography and the interface between lexicography and computational linguistics*

*(d) exploring the possibilities of extensive linking of dictionary content from different European languages*

The importance of the relationship to corpus linguistics is stressed which is missing in the original proposal.

The issue of the collaboration and interaction with other WGs is raised. This is an important issue and we will deal with it as soon as the work plans of the other WGs are clear. It is suggested to make all resources of all WGs available on a secured part of the website. This will allow us to coordinate results of the different WGs.

1. **Work plan and time table**

Simon KREK proposes a work plan and time table based on the topics that were already mentioned for WG3 in the original proposal of COST Action IS1305.

The idea is to organise the meetings of WG3 around these topics dealing ideally with one theme/topic per meeting. Furthermore, the chair and vice-chair of WG3 would like to identify a key person for each meeting who will take the lead for that particular meeting (thus sharing the work load).

July 2014: **Workflow of corpus-based lexicography; Software to support lexicographical workflow (also backup, version control etc.)**

responsible: Carole TIBERIUS

January 2015: **Software to support lexicographical workflow: DWS and CQS**

responsible: Simon KREK and Andrea ABEL

July 2015: **Automatic acquisition of lexical data and its impact (what works, what doesn’t work – example sentences, collocations, neologisms, definitions, word senses)**

responsible: Adam KILGARRIFF and someone from Leuven

January 2016: **Between Corpora and Dictionaries – analysis of the interface between dictionaries and computational lexica and (syntactically and semantically annotated) corpora**

responsible: Simon KREK

July 2016: **The use of lexicographical data in computational linguistics – investigation of possible use of dictionary content for computational linguistic applications**

responsible: Ilan KERNERMAN and Egon STEMLE

It is noted that feedback from the computational linguistics community is needed here.

The proposed order seems OK and will be followed. The theme of the meeting for July 2016 is a hot topic. Some participants are concerned this may be too late. The idea to do some topics together does not seem favourable if meeting time is limited. As a solution, it is suggested to form temporary subgroups related to the topics who can start working on their topic and preparing the meeting well in advance.

We still have two open slots remaining. Suggestions for topics for those meetings are welcomed. Design issues, linking, interaction with the user and crowd-sourcing are mentioned. No decision is taken.

Dissemination of results is important. It is thus important to think about the possible outputs of the different meetings. For instance, the output of the January 2015 meeting could be a paper for a conference, e.g. eLex2015.

It is suggested to create an internal and public repository for the COST Action IS1305 as a whole. It also suggested to create a list with on the one hand the objectives and on the other the outputs of WG3.

Other issues that are raised:

* How are we going to go about developing standards? Some members (Frieda Steurs, Rute Costa) are part of ISO standardisation efforts which offers perspectives; semantic web standards will be taken into account (OUP experience)
* The relationship between back end and front end. Is a dictionary in a database or in an interface; via API or via web interface?
* The overlap with CLARIN: CLARIN is about language resources infrastructure and has a broader scope.

1. **Tasks for Bolzano**

The work plan for Bolzano is discussed. Carole TIBERIUS introduces the concept for the Bolzano meeting. The topic of the meeting is the workflow of corpus-based lexicography. The idea is to discuss this topic using the six phases distinguished by Klosa (to Appear) as a guideline. Carole TIBERIUS outlines the workflow of the ANW project using these six phases (see **Annex 4**).

The underlying idea is that this approach will allow us to compare the different dictionary projects more accurately which will help us to make the workflow more transparent.

The concept is approved. A 7th phase is suggested, the afterlife (to keep the dictionary alive and going after release).

Most likely we follow the method used in the PARSEME COST Action and we will send out a call for poster or short paper presentations for the Bolzano meeting of WG3.

1. **Suggestions for an application for the Lorentz center**

Participants are asked to think about suggestions for an application for a 5-day meeting at the Lorentz center in Leiden (probably in 2016).

1. **AOB**

None

1. **Closing**

Simon KREK closes the meeting at 13h00.

**List of Annexes**

**Annex 1:** draft of the meeting agenda

**Annex 2:** list of WG participants

**Annex 2:** presentation: overview of different dictionary projects

**Annex 3:** presentation: background of WG3

**Annex 4:** presentation: preparation for Bolzano meeting: workflow ANW