

“Workflow of Corpus-based Lexicography”

COST ENeL WG3 meeting

Bolzano, 19 July 2014

In this meeting we will focus on the **workflow of corpus-based lexicography**. Up to now, the lexicographical workflow/process has mainly been described for printed dictionaries (Dubois 1990; Landau 1984; Zgusta 1971). Landau (1984:227) distinguishes three phases: planning, writing and producing, which for printed dictionaries follow nicely one after the other. This is different for online dictionaries (and in particular for those under construction – where the process is continuously running until the end of the project). Klosa (2013) distinguishes a finer-grained process for online dictionaries under construction, identifying six phases (Figure 1).

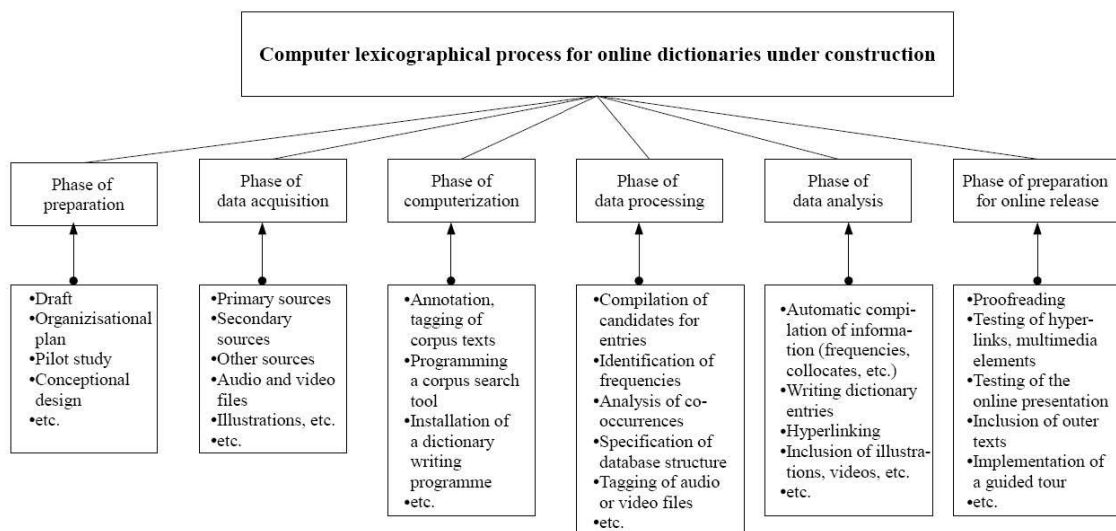


Figure 1 Computer lexicographical process for online dictionaries under construction (Klosa 2013).

Moreover, for online dictionaries, the different phases do not nicely follow one after the other anymore, but merge (e.g. online dictionaries can be published step-by-step: producing an online dictionary may thus begin before the phase of writing is finished) and it is sometimes difficult to identify exact borders between the different phases. (Klosa 2013, Tiberius and Schoonheim, To Appear)

For a completed dictionary, a 7th phase can be identified, i.e. the afterlife which entails maintenance and preservation of the dictionary after the end of the project.

The changes in the construction of dictionaries raise a number of research questions, which we aim to address during the Bolzano meeting:

- Can different online dictionaries (under construction) fit their lexicographical process into these six phases?
- How to deal with version control and archiving of different versions of the dictionary?

- Print dictionaries mostly follow the alphabetic order of the entries, this is not necessarily so for online dictionaries. What guidelines are used in the different projects for determining which entries should be described and in which order they should be described?
- How can visualisation of lexicographical information influence the lexicographical process?
- How do the changes in the lexicographical process affect the user?
- How can users be involved in the lexicographical process and during which phase should this be done?
- How does automatic extraction of lexicographical data from corpora affect the lexicographical process?

For the Bolzano meeting we ask **contributions (2-4 pages)** to be presented at the meeting. Attached is a sample template that you can use for your contribution. You are, however, free to use a different format. The contributions should describe **the lexicographical workflow** of a project using the phases distinguished above (i.e. specifying the tasks that take place in each phase in the project) plus an indication of the time span for each phase. Contributions addressing one of **the other research questions above** can also be submitted.

An example illustration of the time span of the different phases is given in Table 1 for the *Algemeen Nederlands Woordenboek (ANW)*¹ and the *elexiko*² project.

	1	5	10	15
Preparation	ANW			
	elexiko			
Data acquisition				
Computerisation				
Data processing				
Data analysis				
Preparation for online release				

Table 1: Process phases of ANW and elexiko and their time span (Klosa & Tiberius, To Appear)³

A detailed description of the lexicographical process of the *Algemeen Nederlands Woordenboek* can be found in Tiberius and Schoonheim (To Appear).

Important Dates:

- 12 June 2014** Submission of contributions as Word document or PDF (to carole.tiberius@inl.nl)
20 June 2014 Notification of acceptance
19 July 2014 COST WG3-Meeting in Bolzano, in the afternoon

¹ <http://anw.inl.nl>

² <http://www.owid.de/wb/elexiko/start.html>

³ Note that if a block is highlighted, this does not necessarily mean a full-time occupation, but a continuous effort is meant.

References:

Dubois, C. (1990): Considérations générales sur l'organisation du travail lexicographique. In: Hausmann, F. J./Reichmann, O./Wiegand, H. E./Zgusta, L. (eds.): Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie. (= HSK 5.2), Berlin/New York, 1574-1588.

Klosa, Annette (2013) The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus H./Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herberst Ernst (Hgg.): Dictionaries. An international Encyclopedia of Lexicography. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin, Boston: de Gruyter, S. 517-524. (Handbücher zur Sprach- und Kommunikationswissenschaft; 5.4). **PDF enclosed.**

Landau, S. L. (1984): Dictionaries. The Art and Craft of Lexicography. New York.

Tiberius, C. and T. Schoonheim (To Appear) The *Algemeen Nederlands Woordenboek* (ANW) and its Lexicographical Process In: Vera Hildenbrandt (Ed.): Der lexikografische Prozess bei Internetwörterbüchern. 4. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim: Institut für Deutsche Sprache. (OPAL – Online publizierte Arbeiten zur Linguistik X/XXXX). **Preprint enclosed.**

Zgusta, L. (1971): Manual of Lexicography. The Hague.

In the context of the German Netzwerk Internetlexikografie a meeting was organised on the lexicographical process. Presentations from this meeting can be accessed at:

<http://multimedia.ids-mannheim.de/mediawiki/web/index.php/4>. **Arbeitstreffen**