

COST WG3- Workflow

Preparing for Bolzano meeting

Carole Tiberius

Printed dictionaries

- Planning 30%
- Writing 50%
- Publishing 20%

(Landau 1984: 227)

Computer-lexicographical process

6 phases:

- **Preparation**
- **Data acquisition**
- **Computerisation**
- **Data processing**
- **Data analysis**
- **Preparation for online release**

(Wiegand 1999; Müller-Spitzer 2003; Klosa 2011; forthcoming)

Phase

Tasks (selection)

Preparation

Dictionary sketch, Organisational plan (Financing, Workflow, Timeplan, Personnel), Pilot study, Dictionary plan (Style Guide, Sample entries, plans for the user interface and access routes, plans for the technical support, plans for user research)

Data acquisition

Acquisition of sources (Corpora, reference dictionaries, etc.) and other data (e.g. images, videos, sounds)

Computerisation

Preparation of corpus (Tagging, Lemmatisation), development/acquisition corpus software, development/acquisition DWS, development of web application, acquisition of hardware and software.

Data processing

Headword and frequency lists, analysis of corpusdata, modelling of data structure

Data analysis

Editing of dictionary article, definition of hyperlinks, linking of dictionary articles and corpus data, integration of multimedia

Preparation for online release

Proofreading, Testing of user interface (hyperlinks, multimedia, search options, etc.) compiling of help texts.

The ANW Lexicographical Process

Carole Tiberius & Tanneke Schoonheim
Trier 2012

The Algemeen Nederlands Woordenboek

- Online scholarly dictionary of contemporary standard Dutch in the Netherlands and Flanders
- General (mainly written) language
- 1970 onwards
- Corpus-based : ANW corpus
- Semasiological and onomasiological
- Users: from laymen to professionals
- First results on internet: 2009
- Project duration: 2001 - 2018
- Dictionary under construction

Phases in the Lexicographical Process

- Preparation
- Data acquisition
- Computerisation
- Data processing
- Data analysis
- Preparation for online release

(Klosa, forthcoming)

Phase of preparation

- Organisational Plan
 - Finance (1998-2018)
 - Personnel (11fte – 10 fte lexicographic staff; 1 fte IT staff)
- Conceptual Plan
 - Content
 - data structure
 - development of initial type templates for the semagrams
 - sample entries (e.g. bier ‘beer’ and koe ‘cow’)
 - lexicographer’s manual
 - corpus design
 - Functional Design Web application (2007-2009)

Phase of data acquisition

- Primary sources:

ANW Corpus (ca. 100.000.000 words and still growing):

- literary texts
- newspapers
- neologisms
- specific domains (e.g. gardening, sports, web, etc.)

- Secondary sources:

- **Dictionaries**, e.g. Van Dale, WNT, OED, eLexiko, Wikipedia, Google, Groene Boekje, ...
- **Linguistics**: Morfologisch Handboek, Algemene Nederlandse Spraakkunst

Phase of computerisation

- Corpus annotation
 - In house tagged and lemmatised
- Installation of software
 - Dictionary Writing System: ANW editor
 - Corpus Query System: Sketch Engine
- Backup and Version Control

Phase of data processing

- Lemma list
- Sketch Grammar
- automatic data acquisition
 - Spelling
 - Regional variation
 - Synonymy
 - Sense distribution

Phase of data analysis

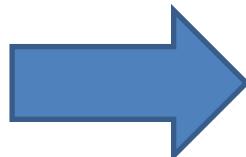
- Lexicographer's job
- Editing per semantic class
-

Data analysis flow



Lexicographic Assistants

- Preparation:
- word family
 - check of automatically compiled information
 - Grammatical information



Lexicographers

Completion of full entry:

- Definition
- Combinations, Idioms
- Examples



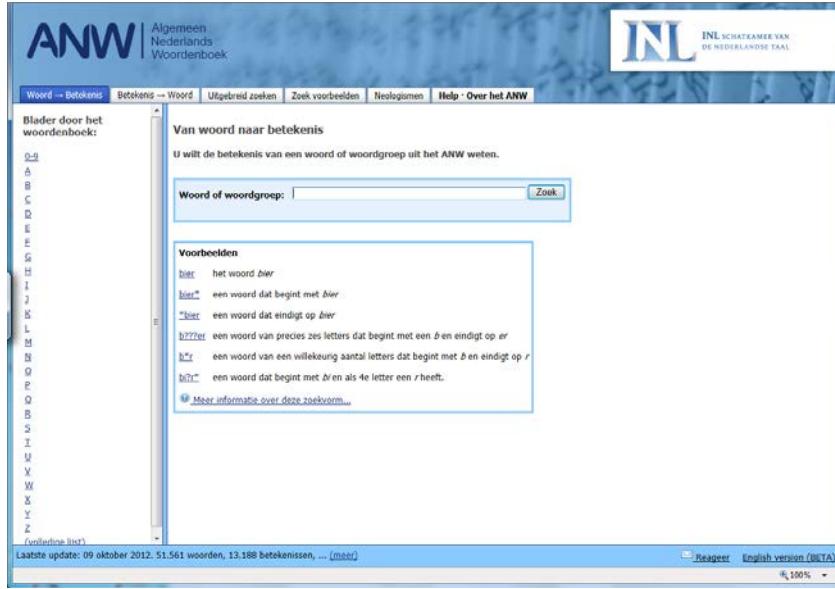
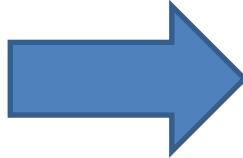
Full check
Check examples
Add multimedia

Editor in chief / project manager



Phase of preparation for online release

Editor in chief / project manager



The screenshot shows the ANW website interface. At the top, there's a navigation bar with links like 'Woord → Betekenis', 'Betekenis → Woord', 'Uitgedreven zoeken', 'Zoek voorbeelden', 'Neologismen', 'Help', and 'Over het ANW'. Below the navigation is a search form titled 'Van woord naar betekenis' with the placeholder 'U wilt de betekenis van een woord of woordgroep uit het ANW weten.' It features a 'Woord of woordgroep:' input field and a 'Zoek' button. A sidebar on the left lists letters from A to Z. The main content area displays search results under 'Voorbeelden' for the query 'bier':

- bier het woord *bier*
- bier* een woord dat begint met *bier*
- *bier een woord dat eindigt op *bier*
- b??ter een woord van precies zes letters dat begint met *b* en eindigt op *er*
- b?r* een woord van een willekeurig aantal letters dat begint met *b* en eindigt op *r*
- b?r* een woord dat begint met *b* en als 4e letter een *r* heeft.

At the bottom of the search results, there's a link 'Meer informatie over deze zoekvorm...'. The footer of the page includes the INL logo, copyright information, and links for 'Reageer' and 'English version (BETA)'.

Online release

- Automatic checks
 - Spelling
 - Broken links
- Release on test environment
- Release

Thanks to:

Chief editor: Tanneke Schoonheim

Project manager: Rob Tempelaars

Lexicographers: Lut Colman, Quin van Dijk, Dirk Kinalble,
Frans Heyvaert, Josefien Sweep, Vivien Waszink

Lexicographic assistants: Arjen Figuee, Pieter Hoekstra,
Marja Schutte, Jasmijn Splinter, Wies van Velsen

IT-people: Jan Niestadt, Carole Tiberius, ...

