

The ANW Lexicographical Process

Carole Tiberius & Tanneke Schoonheim

Algemeen Nederlands Woordenboek

- ❖ An online dictionary under construction
- ❖ Ausbauwörterbuch (Schröder 1997)
- ❖ Dynamisches Wörterbuch (Lemberg 2001)



The ANW

- ❖ 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

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

ANW editor

The screenshot displays the ANW editor interface. At the top, a menu bar includes 'Artikel', 'ANW Corpus', 'Handleiding', 'WDB e.d.', 'Nomenclatuurlijsten', 'Nota's', 'Sjablonen', 'Taalkunde', 'Web', and several utility icons. A dropdown menu is open on the left, listing actions such as 'Open artikel', 'Bewerk metadata', 'Word afdrukweergave', 'Browser afdrukweergaves', 'Converteer naar koepelartikel', 'Verander lemmavorm', 'Nieuw artikel', 'Verwijder artikel', and 'Open artikel in Altova'. Below these are 'Recente artikelen' with a list of article titles like '000-ANW-automatisering', 'occasie', 'waarnemer', and 'zwezerik'. The main window, titled '06-nummer - AnwEdit', has a menu bar with 'Bestand', 'Bewerken', 'Beeld', 'Opzoeken', 'Ga naar', and 'Help'. Its toolbar contains icons for file operations and editing. The left pane shows a tree view for 'Artikel' with sub-items: 'Lemma', 'Spelling', 'Uitspraak', 'Woordsoort', 'Woordvorming', 'Bijzonderheden gebruik', 'Betekenis en gebruik' (expanded to show '1.0 - nummer van een mobiele telefo', '(Etymologie)', '(Forum)', and 'Leescommentaar'), and 'Leescommentaar'.

Sketch Engine



INL SCHATKAMER VAN
DE NEDERLANDSE TAAL

Concordance - Windows Internet Explorer

http://ca.sketchengine.co.uk/auth/preloaded_corpus/anw-inl/ske/view?q=aww

Bestand Bewerken Beeld Favorieten Extra Help

Concordance

Home Concordance Word List Word Sketch Thesaurus Sketch-Diff

View options Sample Filter Sort Frequency Collocation Save

Corpus: ANW-INL
Hits: 67
[conc description](#)

First | Previous Page 2 of 4 Go Next | Last

CLT	Bijman het aan. Ze heette Esther. Het was een	06-nummer	. Zelf had hij nog een vaste telefoon. Toen	
Kranten	verkochten. Dat kon alleen maar via een speciaal	06-nummer	. Zoiets doen Turken niet. Die kopen kaarten	
Kranten) kunnen worden gebeld zijn de reguliere	06-nummers	: 06-0417 en 06-0430. . TA. Financiën	
CLT	reclamebureau, maar daar was ze niet. ' Heeft u haar	06-nummer	? ' Dat had ik. Op de achtergrond hoorde	
Kranten	' verslavingsachtigs ' in de sfeer van	06-nummers	achter deze feiten heeft gezeten. Overigens	
CLT	een onbereikbare buitenwereld. Dat je een	06-nummer	belt als je even wilt rondkijken in Hong	
Kranten	Zoals ook haar pogingen om via een speciaal	06-nummer	de bevolking te mobiliseren buitengewoon	
Kranten	belkosten tot 0,75 per minuut. ' Er zijn	06-nummers	die goedkoper zijn, maar dan krijg je een	
Kranten	Hermanus Koning Een Turk koopt geen kaart via	06-nummer	Door onze redacteur HANS KLIPPUS UTRECHT	
Kranten	vijfhonderd meer dan in 1993, toen 4109 mensen het	06-nummer	draaiden. In 1992 waren dat er nog maar	
Kranten	toegevoegd. Klanten kunnen een speciaal	06-nummer	draaien waarmee snel en efficiënt een	
Neologismen	kantoor had gebruikt voor het frequenteren van	06-nummers	en het zenden en printen van vieze en gemene	
Kranten	jaar. Met ingang van het nieuwe jaar is het	06-nummer	evenwel gewijzigd. Moest tot voor kort	
Kranten	Com en Sense in Alphen aan de Rijn, dat de	06-nummers	exploiteert, bevestigt dit. Maar: ,, Er	
CLT	mocht mislopen, kon ze altijd nog voor een	06-nummer	gaan werken. Sarah legde uit dat ze een	
Kranten	alleen op maandag 1 mei via en speciaal	06-nummer	gegeven worden. Uit de goede antwoorden	
Kranten	munten hebben weggenomen en met haar telefoon	06-nummers	hebben gebeld voor een bedrag van 400 gulden	
Kranten	masturbatielij ' nodig heeft, maar wiens moeder alle	06-nummers	heeft laten afsluiten. Een andere hulpverlener	
Neologismen	diefstal van hun mobiele telefoon hun eigen	06-nummer	hielden bleken ook een bombardement van	
Neologismen	van een gestolen telefoon gewoon hun eigen	06-nummer	houden, krijgen zij op hun nieuw gekochte	

First | Previous Page 2 of 4 Go Next | Last

Internet 100%

SkE talks to ANW-editor

06-nummer Word Sketch - Windows Internet Explorer

http://ca.sketch...

Bestand Bewerken Beeld Favorieten Extra Help

Home Concordance Word List **Word Sketch** Thesaurus

Sketch-Diff

Turn off clustering More data Less data Save

06-nummer ANW-INL freq = 66 [change options](#)

object bij	2	6.8	lidw+SUBST	24	1.8	SUBS
<input checked="" type="checkbox"/> bellen	1	9.47	<input type="checkbox"/> een	11	9.2	<input type="checkbox"/> ov
<input checked="" type="checkbox"/> draaien	1	8.98	<input type="checkbox"/> het	7	5.82	<input type="checkbox"/> in
		>>	<input type="checkbox"/> de	6	2.22	<input type="checkbox"/> va
						<input type="checkbox"/> me

SUBST+vraagzin	1	3.7	onbep vnw+SUBST	2	1.7	aanw
<input type="checkbox"/> waarop	1	7.69	<input type="checkbox"/> alle	1	5.53	<input type="checkbox"/> zo
		>>	<input type="checkbox"/> geen	1	4.87	

subst+SUBST	2	2.0
<input type="checkbox"/> telefoon	1	8.97

Tickbox Lexicography - Select Examples - Windows Internet Explorer

http://ca.sketch...

Bestand Bewerken Beeld Favorieten Extra Help

Home Concordance Word List **Word Sketch** Thesaurus **Sketch-Diff**

Tickbox Lexicography - Select Examples

Lemma: 06-nummer
Gramrel: object_bij
Template: ANW_INL

bellen

En nee, ik bel geen **06-nummers**, dan krijg ik ruzie met Frans, zoals dat bij een goede penningmeester ook hoort.

draaien

Dat waren er bijna vijfhonderd meer dan in 1993, toen 4109 mensen het **06-nummer** draaiden.

Copy to clipboard

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
- ❖

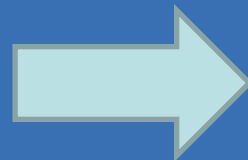
Data analysis flow



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



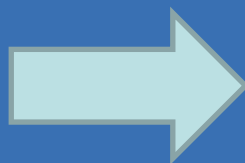
Editor in chief / project manager

Proofreading



Phase of preparation for online release

Editor in chief / project manager



Proofreading

IT

Online release

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

The screenshot shows the ANW (Algemeen Nederlands Woordenboek) website. The header includes the ANW logo and the INL logo. The main content area is titled 'Van woord naar betekenis' and contains a search form with a 'Zoek' button. Below the search form, there is a 'Voorbeelden' section with several examples of words and their meanings. The footer of the page displays the date 'Laatste update: 09 oktober 2012', the number of words '51.561 woorden', and the number of meanings '13.188 betekenissen'. The page is in Dutch and is viewed in a browser window.

Thanks to:

Chief editor: Tanneke Schoonheim

Project manager: Rob Tempelaars

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

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

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