

Algemeen Nederlands Woordenboek (ANW)

A Dictionary of Contemporary Dutch



Rob Tempelaars: rob.tempelaars@inl.nl

The ANW

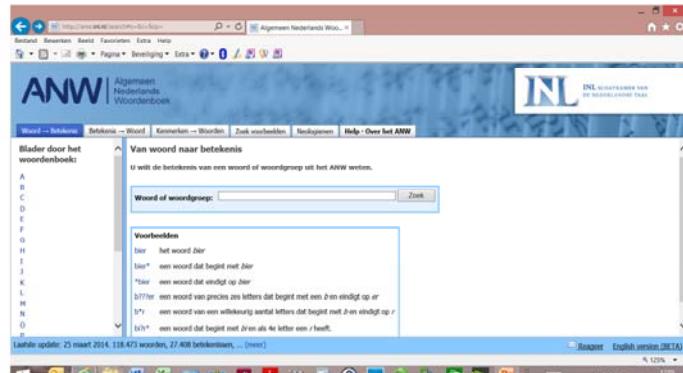
- synchronic scholarly dictionary of contemporary Dutch in Belgium and the Netherlands
- describing words from 1970 onwards
- only digitally available; no printed version
- basic words and neologisms
- semasiological and onomasiological
- many information categories; much more than just word meanings



The *Algemeen Nederlands Woordenboek* (ANW) is a synchronic scholarly dictionary of contemporary Dutch in Belgium and the Netherlands, also paying attention to the Dutch of Surinam and the Netherlands Antilles. It is only digitally available and describes Dutch words that are found from 1970 onwards. The focus of the ANW lies both on the basic words of Dutch and on Dutch neologisms. The dictionary can be used for semasiological as well as for onomasiological search and contains many information categories. The ANW is much more than a dictionary with word meanings only.

Editing ANW articles

- Corpus in Sketch Engine
- Dictionary Writing System
- Online application



For editing the ANW the editors use a corpus of contemporary Dutch which can be accessed through the SketchEngine and a Dictionary Writing System (DWS) which was developed in-house. The users access the ANW in the form of an online application (<http://anw.inl.nl>).

ANW Corpus

- from 1970 onwards
- more than 100.000.000 tokens, regularly updated
- source material from the Netherlands, Belgium (and Surinam)

NRC  HANDELSBLAD literature

knvb.nl

voedingscentrum.nl

dieren.startpagina.nl

MEPPELER  COURANT

ds De
Standaard



The ANW-Corpus consists of more than 100.000.000 tokens of Dutch, collected from 1970 onwards. The main source material comes from Belgium and The Netherlands, but there is also more and more material from Suriname and the Netherlands Antilles. The source material is mainly taken from literature, newspapers and websites in Dutch.

Sketch Engine

The screenshot shows the Sketch Engine interface. At the top, there is a navigation bar with links for 'About', 'Search', and other user information. Below the navigation bar, the main content area is titled 'Corpora'. It displays three tables:

- Corpora**: A table showing corpora by language. It includes columns for 'Language', 'Corpus name', 'Tokens', and 'Words'. Two entries are listed:

Language	Corpus name	Tokens	Words
Dutch	ANW_INL-2.1	115,281,783	100,770,153
Dutch	Dutch web corpus (old WSG)	127,838,255	110,892,646
- My corpora**: A table showing corpora owned by the user. It includes columns for 'Language', 'Corpus name', 'Configuration template', and 'Tokens'. One entry is listed:

Language	Corpus name	Configuration template	Tokens
no corpora			
- Other users' corpora**: A table showing corpora owned by other users. It includes columns for 'Language', 'Corpus name', 'Tokens', and 'Owner'. Four entries are listed:

Language	Corpus name	Tokens	Owner
Dutch	Corpus Spoken Dutch	10,403,293	Carole Tiberius
Dutch	CorpusLiteraryTexts	53,835,940	Carole Tiberius
Dutch	DutchSampleCorpus	292,760,335	Carole Tiberius
Dutch	SohAr-newspapers	167,051,122	Carole Tiberius

The corpus material of the ANW is available through the Sketch Engine, which enables the editors to make use of concordances, word sketches and some special features.

ANW in SketchEngine

- concordances
- word sketches
- special features



Concordance

The screenshot shows the Sketch Engine Concordance interface. At the top, there's a navigation bar with 'Sketch Engine' logo, user information ('user: Tanneke Schoonheim corpus: ANW_INL-2.1'), and links for 'About', 'Search' (with 'mager' typed in), and 'mager'. Below the navigation is a search bar with 'Query mager 2,426 > GDEX > Sort Right 2,426' and page controls ('First | Previous Page 51 of 122 Go Next | Last'). A dropdown menu says 'Concordance is sorted. Jump to: ▾'. The main area displays a list of 2,426 entries, each with a document name and year, followed by a snippet of text containing the word 'mager'. The first few entries are:

Document	Text Snippet
Pluscorpus,NN	vertellen, maar het verdoezaat toch een te mager gegeven. Ondanks de vaart waarmee je de
Neologismen,NN,2002	weten alle acteurs het maximale uit het mager gegeven te halen. Het geruzie tussen de
Kranten,NN,1994	inbewaringstellering aan de orde zijn gekomen. Uit de mager gegevens die voorhanden zijn, mag worden
Domeinen,BN	rundvlees, kalfsvlees, kip, kalkoen, konijn, mager gehakt, kabeljauw, tonijn (uit blik: op
CLT,NN,2000	besloten had tot een zuing half pondje mager gehakt hoorde ze zichzelf opeens bijna
Kranten,NN,1995	kampioenschapswedstrijd in Klazienaveen was van een mager gehalte. Het zal ongetwijfeld met de spanning
Kranten,NN,1995	tweede bedrijfs was er eentje van een erg mager gehalte. Stadskanaal probeerde wel wat,
Kranten,NN,1995	actie van André Steenberg. Het bleef 1-0. Mager gehalte Het hele tweede bedrijfs was er
Kranten,NN,1994	dat leidt tot eenzijdigheid. Het moet geen mager geheel worden. De komende veertig jaar
CLT,NN,2002	. Ze hadden de hele dag hard gewerkt, de mager geiten gehoed en de dorre akker beplant
CLT,BN,1981	stem, die spreekt, de ogen staan dof in het mager gelaat. maar het driftig geluid en de warmte
Kranten,BN,1995	Heinrich Schütz. Geel licht beschijnt zijn mager gelaat en zijn knokige handen, waarin hij
Kranten,BN,1996	renners dreigen met staking vanwege het mager geldgewin. Dat stond niet in verhouding
Kranten,NN,1995	rendement van 10 procent in de regel al als mager geldt, en waar de reserveringen voor de
CLT,BN,1978	plaats zetten, wat eten zoeken voor een mager gehalte hond die onder een woonwagen sliep
Kranten,BN,1995	dat ik me niets te verwijten heb voor het mager gelijkspel op Tongeren. Andere spelers
CLT,NN,2001	geflapt. Elke grap bleek bij nader inzien een mager gemaskeerde beleidings te zijn. Zodra het
Kranten,NN,1994	en sporadisch worden rolmodellen in een mager gemeenschappelijke Amerikaanse cultuur.

Concordance of the head word *mager* sorted on the word on the right. We can see combinations such as *magere gegevens* ‘meager data’, *mager gehakt* ‘low-fat mincemeat’ and *een magere gestalte* ‘a lean figure’.

Word Sketch

Sketch Engine

user: Tanneke Schoonheim corpus: ANW_INL-2.1

Concordance	1,305 7.3	bijw+ADJ	471 12.5	en/of	165 2.6	ADJ+vzgroep	6
<input type="checkbox"/> troost	<u>28</u> 8.82	<input type="checkbox"/> nogal	<u>20</u> 6.92	<input type="checkbox"/> halfvol	<u>8</u> 9.84	<u>ADJ_van-s</u>	<u>3</u> 3.8
<input type="checkbox"/> Hein	<u>22</u> 8.59	<input type="checkbox"/> te	<u>69</u> 6.82	<input type="checkbox"/> spichtig	<u>4</u> 9.13	<u>ADJ_voor-s</u>	<u>3</u> 3.1
<input type="checkbox"/> yoghurt	<u>17</u> 8.42	<input type="checkbox"/> uiterst	<u>15</u> 6.66	<input type="checkbox"/> bleek	<u>11</u> 8.62		
<input type="checkbox"/> melkpoeder	<u>10</u> 7.86	<input type="checkbox"/> erg	<u>50</u> 6.59	<input type="checkbox"/> gespierd	<u>4</u> 7.86		
<input type="checkbox"/> melk	<u>26</u> 7.64	<input type="checkbox"/> extreem	<u>4</u> 6.39	<input type="checkbox"/> blond	<u>4</u> 6.25		
<input type="checkbox"/> vlees	<u>40</u> 7.64	<input type="checkbox"/> wat	<u>24</u> 5.38	<input type="checkbox"/> zwak	<u>3</u> 4.34		
<input type="checkbox"/> scharminkel	<u>8</u> 7.59	<input type="checkbox"/> bijzonder	<u>8</u> 5.32	<input type="checkbox"/> klein	<u>3</u> 1.15		
<input type="checkbox"/> zesje	<u>2</u> 7.44	<input type="checkbox"/> hoe	<u>20</u> 4.91				
<input type="checkbox"/> oogst	<u>14</u> 7.44	<input type="checkbox"/> zo	<u>54</u> 4.87				
<input type="checkbox"/> kwark	<u>6</u> 7.18	<input type="checkbox"/> lief	<u>3</u> 4.84				
<input type="checkbox"/> vleessoort	<u>6</u> 7.17	<input type="checkbox"/> relatief	<u>4</u> 4.61				
<input type="checkbox"/> beestje	<u>8</u> 7.1	<input type="checkbox"/> heel	<u>22</u> 4.54				
<input type="checkbox"/> jongeman	<u>12</u> 7.1	<input type="checkbox"/> eerder	<u>6</u> 4.45				
<input type="checkbox"/> jus	<u>5</u> 6.77	<input type="checkbox"/> vrij	<u>5</u> 4.31				
<input type="checkbox"/> borstkas	<u>5</u> 6.61	<input type="checkbox"/> zeer	<u>19</u> 4.23				
<input type="checkbox"/> kaas	<u>9</u> 6.54	<input type="checkbox"/> steeds	<u>10</u> 3.91				

>>

subst+ADJ	71 4.7
<input type="checkbox"/> Harrold	<u>3</u> 10.32
<input type="checkbox"/> beetje	<u>5</u> 2.58
<input type="checkbox"/> Jan	<u>3</u> 2.1
<input type="checkbox"/> dag	<u>3</u> 0.7

>>

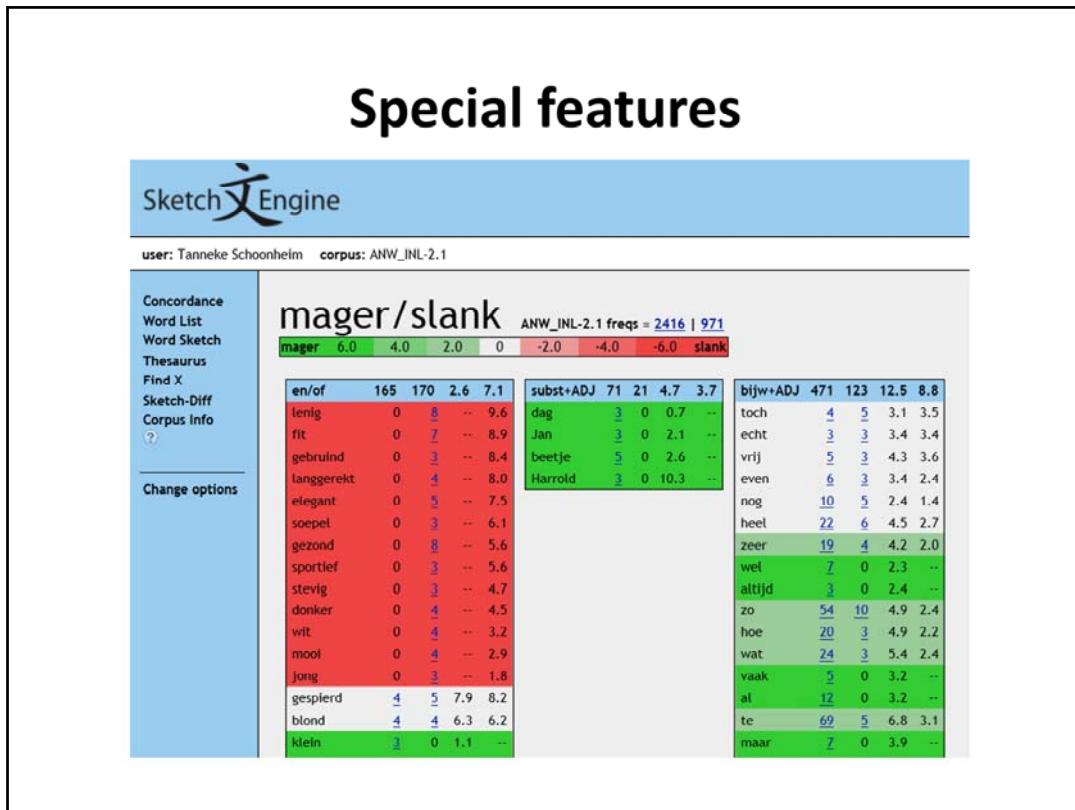
Word sketch of *mager*. We can see combinations with a noun on the right, such as *magere troost* ‘cold comfort’, *magere yoghurt* ‘low-fat yoghurt’, combinations with another adjective, such as *bleek en mager* ‘pale and thin’ and combinations with a noun on the left, such as *een beetje mager* ‘a bit skinny’.

Special features

The screenshot shows the Sketch Engine interface with the title "Sketch Engine" and the user information "user: Tanneke Schoonheim corpus: ANW_INL-2.1". On the left, a sidebar lists navigation options: Concordance, Word List, Word Sketch, Thesaurus, Find X, Sketch-Diff, Corpus Info, and a help icon. The main area is titled "Word Sketch Differences Entry Form" and contains fields for "Lemma" (mager), "Part of speech" (adjective), and "Sketch diff by" (set to lemma). It also includes fields for "Second lemma" (slank), "First subcorpus" (None (whole corpus)), "Second subcorpus" (None (whole corpus)), and "word form" fields for both words. There are "Advanced options" and "Show Diff" buttons at the bottom.

The Sketch Engine gives for instance the similarities and differences in use between the words *mager* 'skinny' and *slank* 'slim'.

Special features



From this overview it becomes apparent that *slank* has more positive associations and *mager* more negative. *Slank* goes with adjectives such as *lenig*, *fit*, *soepel* en *gezond* ‘healthy’, whereas *mager* is combined with *klein* ‘small’.

Dictionary Writing System

All functionalities needed for writing dictionary entries



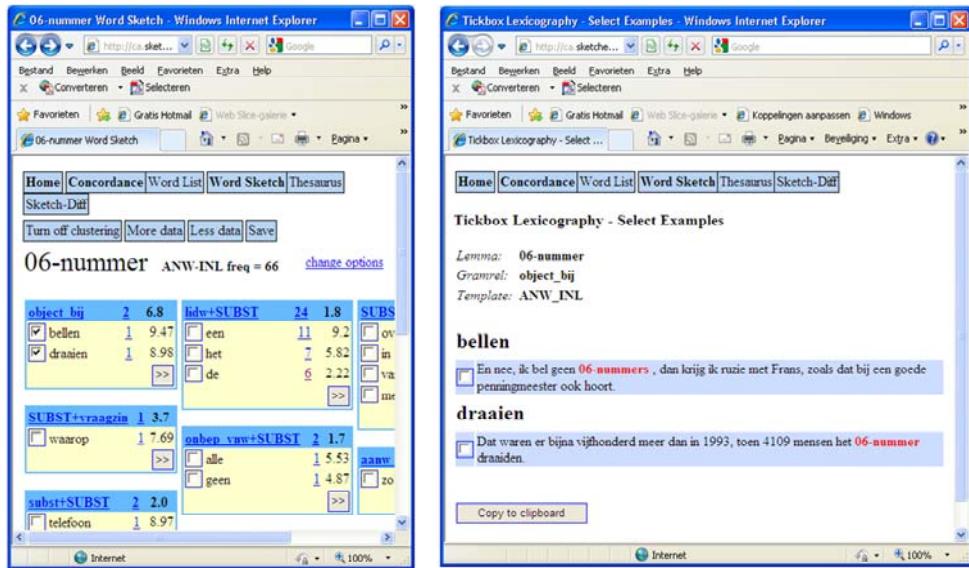
The DWS of the ANW consists of the ANW-corpus in the SketchEngine, an entry list containing all entries that have to be described, editorial guidelines and memos, links to other dictionaries and secondary sources, links to internet sources and the ANW-editor.

Editor

- in-house developed
- written in Java
- entries stored in MySQL database
- entries link to other entries using persistent identifiers
- builds the user interface from a xml-schema



Linked to SketchEngine



We use the Tickbox Lexicography to select the most appropriate combinations and examples for the dictionary. These examples are copied directly from the Sketch Engine into the dictionary, including references such as author, source and date. Combinations are also copied directly from the corpus database into the dictionary entry, complete with the selected examples and their references.

Entries

partly edited entries (September 2014: ca. 16.000)

- edited information on part of speech, spelling, abbreviation, pronunciation and use
- links to concordances in corpus
- links to information in other dictionaries

fully edited entries (September 2014: ca. 16.000)

- edited information on part of speech, spelling, abbreviation, pronunciation and use
- edited information on word meanings, word combinations (collocations) and expressions
- edited example sentences

The ANW contains partly edited entries (September 2014: ca. 16.000) and fully edited entries (September 2014: ca. 16.000).

Partly edited entry

The screenshot shows the ANW website interface. At the top, there is a blue header bar with the ANW logo and the text 'Algemeen Nederlands Woordenboek'. Below the header, a navigation bar includes links for 'Woord → Betekenis', 'Beteenis → Woord', 'Kenmerken → Woorden', 'Zoek voorbeelden', 'Neologismen', 'Help', and 'Over het ANW'. The main content area has a light blue background with a wavy pattern. On the left, a sidebar for the word 'kaak' lists 'Algemene grammaticale gegevens' with options for 'Woordsoort', 'Spelling en flexie', and 'Woordvorming'. It also shows 'Toon: Hele artikel' and 'Zoek 'kaak' ook in:' sections. On the right, the main article page for 'kaak' displays 'Algemene grammaticale gegevens' (noting it contains no betekenisprofile), 'Woordsoort' (substantief), 'Spelling en flexie' (including forms like 'kaak' and 'kaaks'), 'Woordvorming' (including 'kaakje'), and a 'Reageer' button at the bottom.

Partly edited entries contain edited information on part of speech, spelling, abbreviation, pronunciation and use, links to concordances in corpus and links to information in other dictionaries. These entries contain no definitions, no example sentences and no collocations. In the near future they will be provided with corpus material (concordances) to give the user a general idea of the meaning(s) of the entry. We're also looking into the possibility to add a kind of word sketches as well. Partly edited entries can become fully edited entries in the course of time.

Fully edited entry

The screenshot shows the ANW website interface. At the top, there's a navigation bar with links like 'Woord → Beteenis', 'Beteenis → Woord', 'Kenmerken → Woorden', 'Zoek voorbeelden', 'Neologismen', 'Help', and 'Over het ANW'. The logo 'INL INSTITUUT VOOR DE NEDERLANDSE TAAL' is visible on the right. The main content area is titled 'kaakchirurg'. Below the title, there's a section for 'kaakchirurg 1.0:' which defines it as a surgeon specialized in operations of the jaw bones and surrounding tissue. There are also sections for 'Semagram', 'Algemene voorbeelden', and a sidebar with detailed information about the word's type (substantief), gender (mannelijk), and other linguistic properties like spelling and flexie.

Fully edited entries contain edited information on part of speech, spelling, abbreviation, pronunciation and use, edited information on word meanings, word combinations (collocations) and expressions and edited example sentences.

Part of Speech

noun; adjective; verb; adverb; preposition, etc.

noun:

- word type (appellative, proper name)
- word gender (male, female, neuter or a combination)
- article (*de, het* or both)
- number (no singular, no plural, rare in singular, rare in plural)
- word class (personal name; abstractum; collective noun, etc.)

The ANW structure contains elements such as part of speech, spelling, pronunciation, morphology and pragmatics. Most elements contain sub elements. For part of speech for instance, the element *noun* contains sub elements such as word type, word gender, article, number and word class.

Part of Speech

noun; adjective; verb; adverb; preposition, etc.

verb:

- function (auxiliary verb, intransitive verb)
- syntactic class (transitive, intransitive, reflexive, or a combination)
- flexion (weak, strong, irregular, or a combination)
- auxiliary verb (*hebben*, *zijn* or both)

The element *verb* in the same category contains sub elements such as function, syntactic class, flexion and auxiliary verb.

Spelling

official Dutch spelling; also for neologisms

- 1 aprilgrap, e-fiets, facebooken

abbreviation

- 1 april.grap, e-fiets, face.boo.ken

variants

- 1 aprilgrap, eenaprilgrap



Spelling is given according to the official Dutch spelling rules. Besides spelling information on abbreviation and variants are given. This is also done for neologisms.

Pronunciation

amount of syllables

position of the main stress

way of pronunciation

phonetic transcription

cornedbeef

- 3 syllables; stress on 2nd syllable
- Dutch pronunciation [kɔr'nɛtbif]



Morphology

types

- simplex (*hand, huis*)
- derivative (*handig, huisje*)
- compound (*handschoen, huissleutel*)
- ...
- acronym (*aids, NATO*)
- blend (*smirten, twitteratuur*)
- shortening (*appen < whatsappen*)

The category morphology consists of various types such as simplex, derivative, compound, acronym, blend and shortening.

Pragmatics

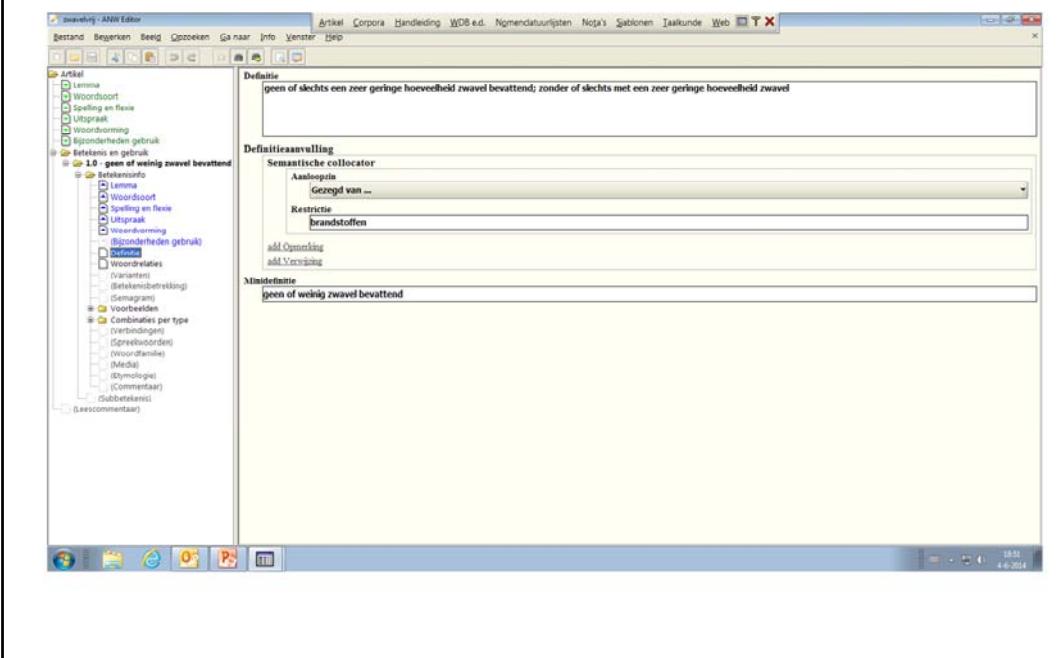
- language variety (Dutch in Belgium, Dutch in the Netherlands, Dutch in Surinam)
- style (formal, informal, vulgar, etc.)
- attitude (ironic, sarcastic, offensive, etc.)
- domain (law, politics, sport, etc.)
- frequency in the ANW-corpus
- time (archaic, neologism, etc.)
- medium (spoken language, written language)

The ANW contains also information on pragmatics, such as language variety, style, attitude, domain, frequency, time and medium. Not every category will be filled in in every entry, but if relevant, the information is given. Some indications appear automatically, as the Sketch Engine for instance indicates language variety.

Definitions in the ANW

- Analytical definitions
- Short definitions
- Semantic collocators
- Remarks

Definitions



The entry *zwavelvrij*

Definition: *geen of slechts een zeer geringe hoeveelheid zwavel bevattend; zonder of slechts met een zeer geringe hoeveelheid zwavel* 'containing no or only a small amount of sulfur; without or with only a small amount of sulfur'

Semantic collocator: *gezegd van* 'said of'

Restriction: *brandstoffen* 'combustibles'

Short definition: *geen of weinig zwavel bevattend* 'containing no or little sulfur'

Lexical relations

- hyperonymy/hyponymy
huis > gebouw; gebouw > huis
- synonymy
fiets – rijwiel
- antonymy
zwart – wit
- andronym/feminym
boer – boerin; Sloveen – Sloveense

Semagram

A semagram is a conceptual structure that describes a lexical concept on the basis of its characteristics

- invented by Fons Moerdijk, former editor-in-chief *ANW*
- presentation of word knowledge in a frame with **slots** and **fillers**

Semagram

- **slots** are conceptual elements naming characteristics and relations of words, e.g. colour, size, place, etc.
- **fillers** are the data in the slots, e.g. is yellow, is big, lives in een nest, etc.
- part of the information can be encyclopaedic
- particularly useful for nouns, but also for verbs and adjectives



Why semagrams?

- There is often more relevant information on words than you can fit in a definition without making it unreadable for the dictionary user
- The definition contains the prototypical lexical semantic information on the words, the semagram contains also other relevant information
- Semagrams are well suited for electronic dictionaries such as the ANW, in which it is easy to search for specific information
- **The semagram helps to formulate the right definition**

Word combinations

Word combinations are well-known, rather conventional **syntactic** combinations of words.

You understand a word combination because you know the meaning of the separate words that are part of it.

- to go to the cinema
- to make a decision
- to drink beer
- to smoke a cigarette

Cigarette as object to a verb

een sigaret **aansteken**; een sigaret **opsteken**; een sigaret **roken**; een sigaret **oproken**; een sigaret **inhaleren**; een sigaret **doven**; een sigaret **uitdoven**; een sigaret **uitdrukken**; een sigaret **uitduwen**; een sigaret **uitmaken**; een sigaret **draaien**; een sigaret **rollen**; een sigaret **aanbieden**; een sigaret **krijgen**; een sigaret **nemen**; een sigaret **presenteren**; een sigaret **bieten**

sigaretten **halen**; sigaretten **kopen**; sigaretten **verkopen**; sigaretten **smokkelen**

Cigarette in other combinations

Combinations with an adjective

een **nieuwe** sigaret; een **verse** sigaret; de **laatste** sigaret; zijn **laatste** sigaret; een **Amerikaanse** sigaret; een **Egyptische** sigaret; een **Engelse** sigaret; een **Franse** sigaret; een **Turkse** sigaret; **Amerikaanse** sigaretten; **Egyptische** sigaretten; **Engelse** sigaretten; **Franse** sigaretten; **Turkse** sigaretten; een **dunne** sigaret; een **losse** sigaret; een **lichte** sigaret; **lichte** sigaretten; de **eeuwige** sigaret; zijn **eeuwige** sigaret; **gewone** sigaretten; een **halve** sigaret

Combinations with a substantive

een **pakje** sigaretten; een **slof** sigaretten; een **paar** sigaretten

Idiomatic word groups

Fixed idiomatic word groups often appear in figurative speech

Sometimes a few words, sometimes formula-like sentences

Meaning is not (easy) to deduct from the separate parts of the word group.

- *klein bier* ‘nothing important; nothing to worry about’
- *eerste viool* ‘leading violinist in an orchestra’
- *gele trui* ‘jersey of the leader in the Tour de France’

Collocation *gele trui*



trui

de gele trui

geel tricot dat in de ronde van Frankrijk (Tour de France) gedragen wordt door de (tussentijdse) leider van het algemeen klassement

De eerste gele trui werd in 1919 uitgereikt aan Eugene Christophe. De kleur geel van de gele trui is zo gekozen, omdat de sponsor van de Tour de krant L'Auto werd gedrukt op geel papier.

- <http://www.wielersportinfo.nl/tourdefrance.htm>

Lance Armstrong blijft uiteraard leider en kreeg een nieuwe gele trui uit de handen van een andere "Terminator", Arnold Schwarzenegger.

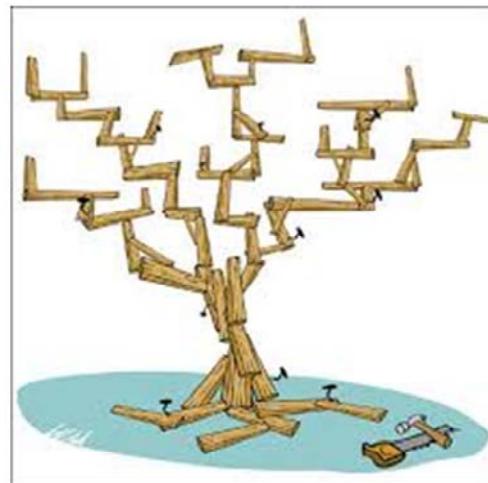
- <http://www.sportupdate.be/wielrennen/tour.htm>, 2003



Word family

Word formations of which the headword is one of the elements.

- Derivatives
- Compounds
- Others



Word family of *azijn* ‘vinegar’



azijn

Woordfamilie

Als deel van een afleiding: [azijnachtig](#); azijnen; azijnig

Als rechterlid in samenstellingen en samenkoppelingen: alcoholazijn; appelazijn; halsamicoazijn; balsemazijn; champagneazijn; ciderazijn; cranberryazijn; dragonazijn; frambozenazijn; honingazijn; ijzerazijn; inmaakazijn; keukenzijn; kruidenazijn; natuurazijn; rijstazijn; schoonmaakazijn; sherryazijn; tafelazijn; tarweazijn; vanilleazijn; [wijnazijn](#); xeresazijn; zeevruchtenazijn

Als linkerlid in samenstellingen en samenkoppelingen: [azijnaltaar](#); azijnanhydride; azijnbacterie; azijnbeker; azijnbekken; azijnboom; azijncompres; azijndoek; azijnessence; azijnether; azijnfabriek; azijnfabrikant; azijnfles; azijnflesje; azijngeur; azijnhydride; azijnkuur; azijnlegger; azijnlucht; azijnontdekking; azijnoplossing; azijnpisser; azijnsaus; azijnsausje; azijnscoort; azijnstel; azijnsyndroom; azijnvleug; azijnwater; azijnwinkel; azijnwolk; azijnzuur

Overige woordfamilieleden: olie-en-azijnstel

Examples

- Example sentences illustrate the meaning of an entry
- Taken from the corpus via the SketchEngine
- If necessary taken from the internet, e.g. neologisms
- Inserted bij the lexicographers
- Corrected by the lexicographical assistants

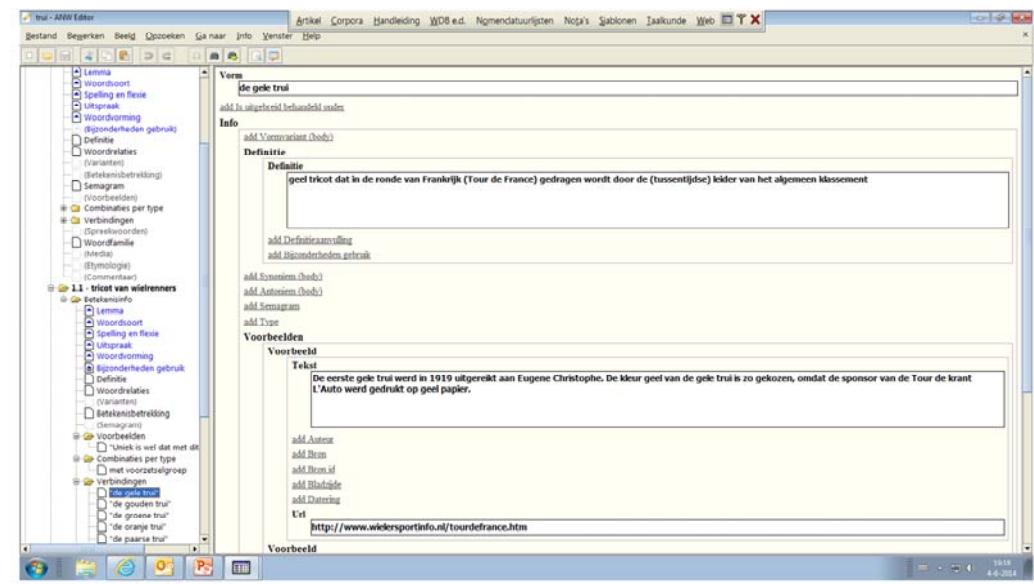
- Multiple examples per entry
- Multiple examples per meaning
- At least one example per collocation

Examples in the SketchEngine

Query trui 1,833 > GDEX > Sort Left 1,833	
First Previous	Page 40 of 92 Go Next Last Concordance is sorted. Jump to: ▾
Kranten,BN,1996	Zorgen dat we er vooral niet in de gele trui heen gaan om de supporters van Indurain
CLT,BN,1979	tour heeft gewonnen maar ook in de gele trui als eerste in het prinsenpark is aangekomen
CLT,NN,1984	bergklassementen, of veertig dagen in de gele trui . dat ziet er redelijk uit, maar ik zei
Kranten,BN,1996	van het Vlaams Parlement krijgt de gele trui in de Ronde van Vlaanderen , een symbolische
Neologismen,NN,2001	de Tour bezig. In Duinkerken lag de gele trui voor het grijpen. " Ik heb drie dagen met
Kranten,NN,1993	1991 toen hij een kleine week lang de gele trui droeg in de Tour de France, nadat hij eerder
Kranten,BN,1993	Greg Lemond, droeg achttien dagen lang de gele trui en werd in het klassement pas kort voor
Kranten,BN,1996	geweest. De terreinkennis levert me de gele trui op. De moeite waard, niet?" Auteur: (
Kranten,BN,1996	van de Champs Elysées beklimt met de gele trui , bén ik er niet ", weerde die af.,, Maar
CLT,NN,2003	Gaat u nu al? reageert de man met de gele trui .' Het wordt net gezellig, mevrouw.' 143
CLT,NN,2002	Fransch en betekend weg) peeste met de gele trui omme bast maar ik had die gele trui niet
Kranten,BN,1996	sterk nummertje onderweg was naar de gele trui . De Belgische kampioen reed van Beveren
CLT,NN,1984	die hij ging winnen, hij had net de gele trui veroverd) verbannen wegens fraude bij een
Kranten,BN,1997	ook na de manche in Brussel nog de gele trui mocht aantrekken.,, Ik ben echter wel van
Neologismen,NN,2000	etappes zal ONCE onderweg beslissen of de gele trui wordt verdedigd. Pas volgende week trekt
Kranten,BN,1996	ploegmaats, al lang eingestiegen om de gele trui van Bjarne Riis te vrijwaren, merkten dat
Kranten,BN,1996	Masson. Geen vlammetende strijd meer om de gele trui in het onherbergzame decor van de Sierra
Kranten,BN,1996	Ploegen die niets met de strijd om de gele trui te maken hadden, sportief geen enkele baat
Kranten,NN,1994	momenteel gericht op de strijd om de gele trui in de Tour de France. In het verre Ecuador
Kranten,BN,1996	de Ronde met 55 seconden.",, Ook de gele trui . Alsof ik die niet mócht dragen. Proloog

Example sentences are indicated in the Sketch Engine and copied directly into the entry in the ANW editor.

Examples in the ANW editor



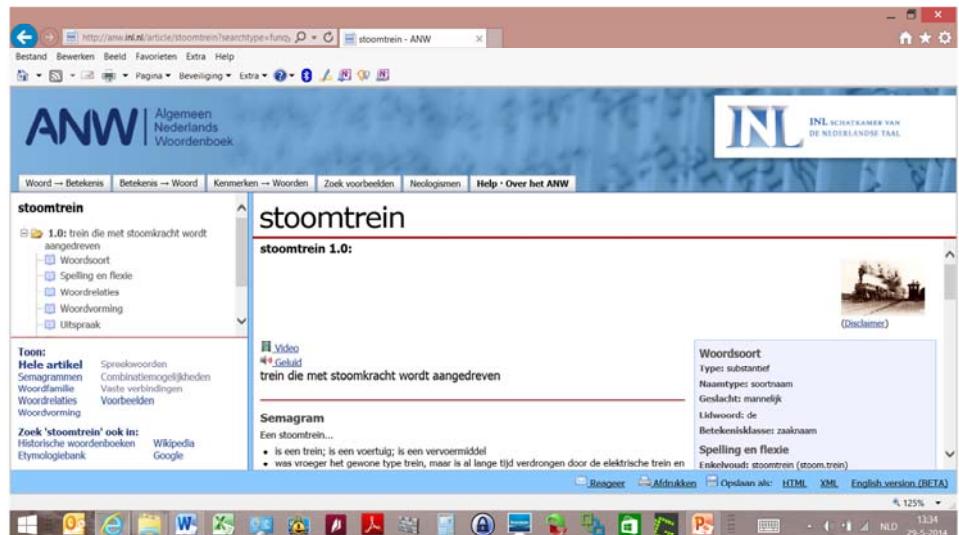
Hypermedia

Sometimes it is difficult to capture a concept in words.
Images and sounds support the definition in the ANW

- pictures
- sounds
- movie clips



Hypermedia



The availability of multimedia is indicated in the alphabetic entry list on the left of the search screen and on the top of the article screen, close by the definition.

Etymology

ANW is a synchronic dictionary.

ANW lexicographers don't write new etymologies.

Etymological information in the ANW:

- link to www.etymologiebank.nl
- neologisms: information on first appearance, reason of introduction, inventor of the word, motive for the word, etc.

Data analysis flow

- Lexicographic Assistants:** check automatically compiled information, grammatical information, word family →→
- Lexicographers:** add word class, definition, word relations, combinations, expressions and examples →→
- Lexicographic Assistants:** check data and examples, add multimedia →→
- Editor in chief/project manager:** proofreading →→→
- Lexicographers:** add corrections →→
- Lexicographic Assistants:** check multimedia →→
- Editor in chief/project manager:** final check →→→ GO ONLINE

The online application

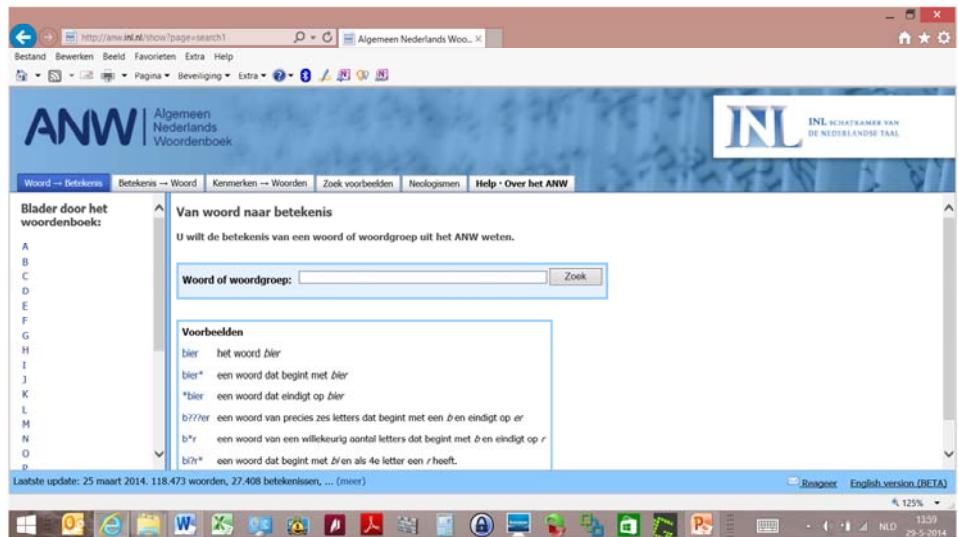


<http://anw.inl.nl>

Technical information

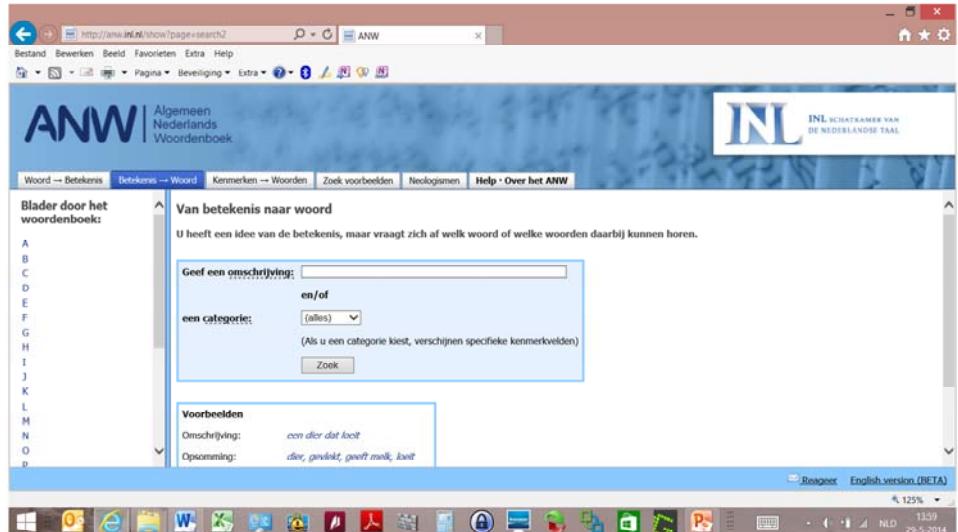
- The application is written in Java. The user interface consists of HTML, CSS en Javascript/ECMAScript.
- The application uses components of the Apache Software Foundation, e.g. Tomcat, Lucene, Xalan, Log4J en Velocity.
- The application uses MySQL as database.
- For the application a querytaal called FunQY was developed.
- The application is tested under Explorer 6-8, Firefox 3 and Safari 4.

Word → Meaning



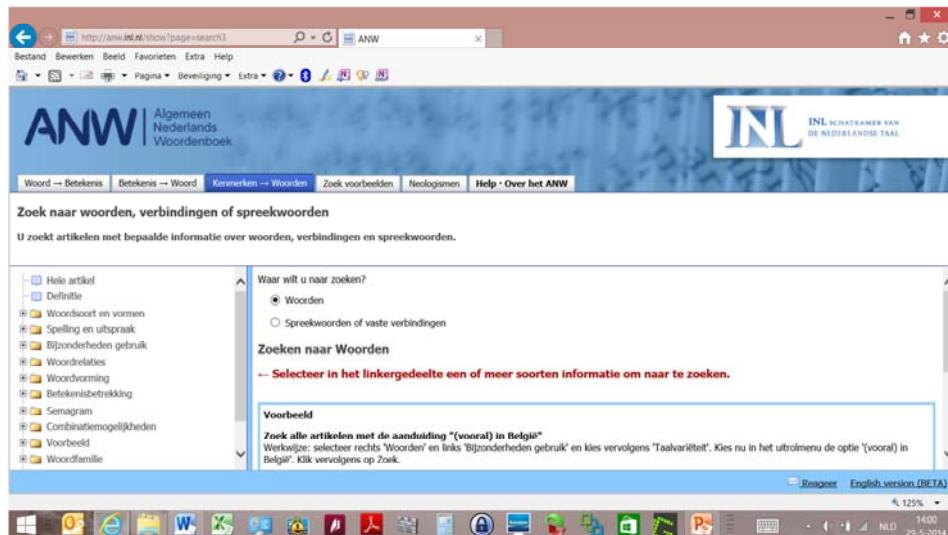
The most common search is from word to meaning. This is used by people who hear or read a word and who want to know what it means. In the ANW you can enter the word in the search frame in the middle of the screen, but you can also search for words using the alphabetical list of available entries on the left.

Meaning → Word



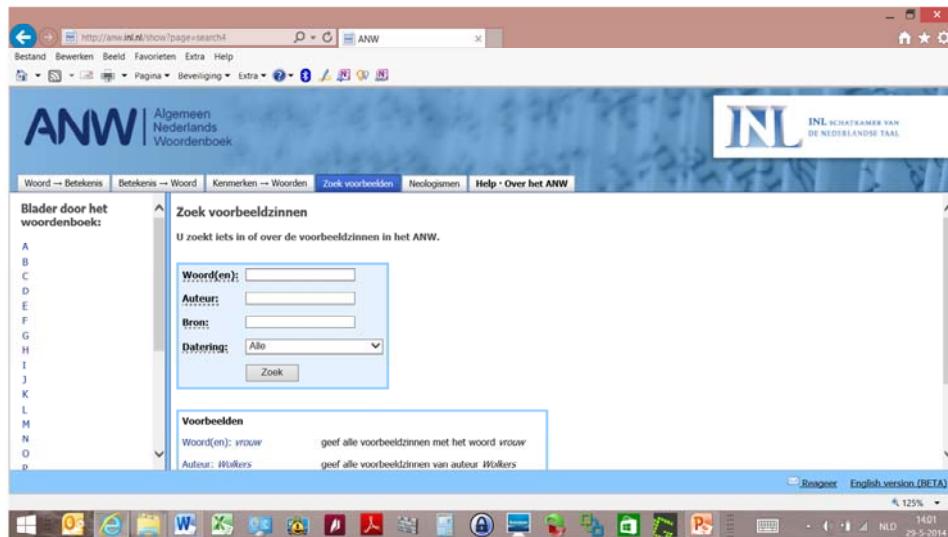
It is also possible to find a word you don't know (although it might be on the tip of your tongue). In this case you can enter characteristics of the word you're looking for in the search frame on the screen and the system will provide you with suggestions on the word you might be looking for. If you enter 'animal; mammal; grey; Afrika' you get five results (*genetkat, olifant, oryx, stokstaartje, teek*), if you also add 'trunk', you only get the elephant.

Features → Word



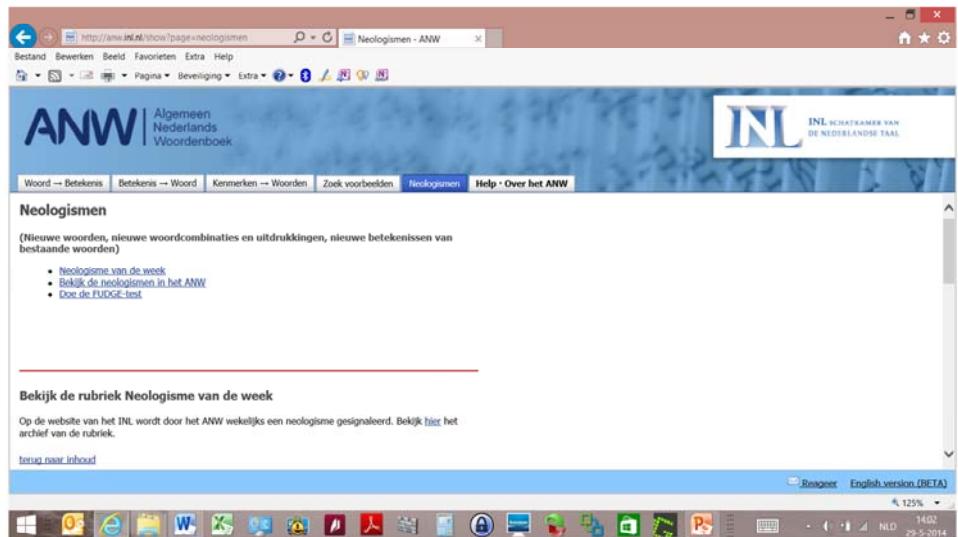
The search from features to word is for the really sophisticated user. Here you can find anything you want to know about the words in the dictionary but it asks a lot of you. All elements and sub elements of the ANW structure are available, but it is sometimes hard to combine them in the way to get the needed information. We're working on it to make this search option better usable by remodelling the page and predefining frequently asked questions.

Find examples



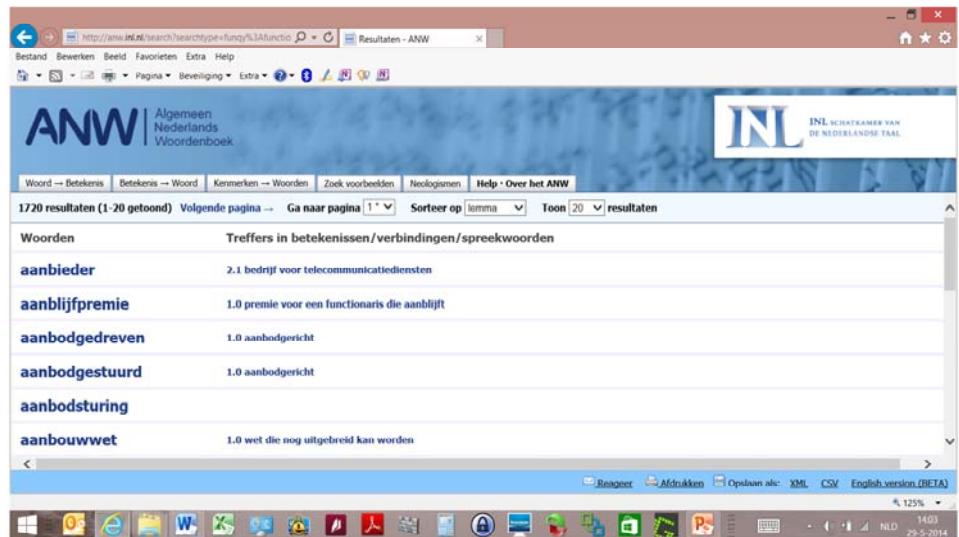
It is also possible to look for words, certain authors, sources or dates in the example sentences that are used in the ANW. This makes it for instance possible to collect the neologisms from certain newspapers.

Neologisms



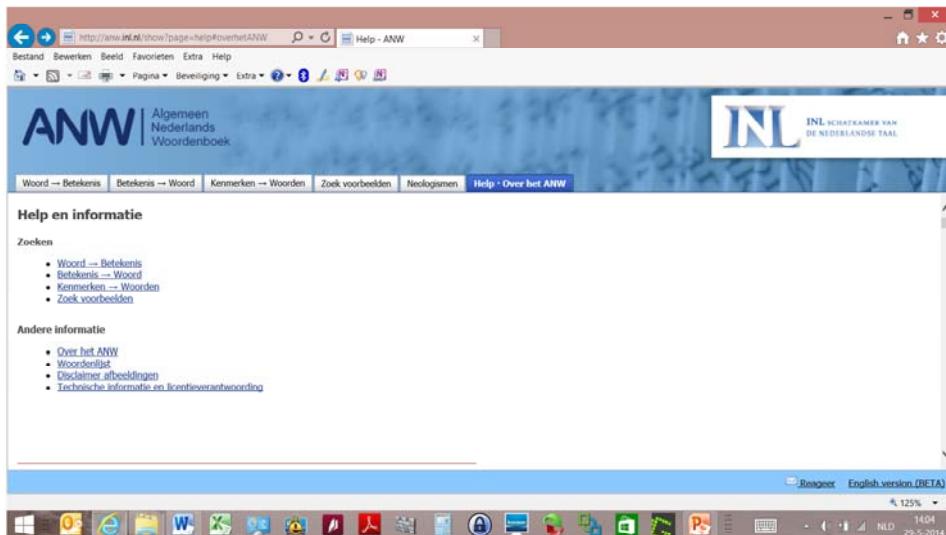
Neologisms are important for the ANW. Not all of them are already included into the dictionary. Some of them are highlighted as 'neologism of the week'. Users can also do the FUDGE test, to determine the chances of a neologism to become a regular part of the vocabulary.

Neologisms



On October 20th 2014, the ANW contained 1864 neologisms, from *aanbieder* 'provider' to *zzp'er* 'someone who works for himself without having employees'.

Help and Information



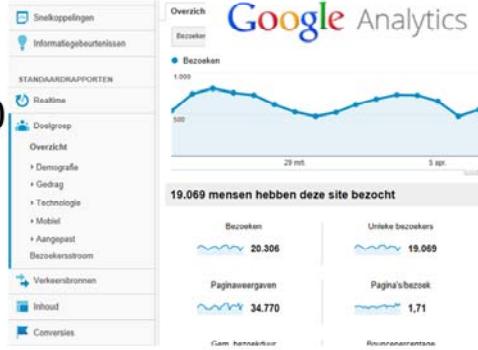
There is also a page with help and information. This page contains information on the ANW and the ANW corpus and explains all search possibilities.

Logfile analysis

➤ Logfiles van 12/2009

```
134.58.253.55 - - [12/Feb/2013:14:43:08 +0100] "GET  
/search?searchtype=descword&category=dier&description=zoogdier%3B+groot+grijs%3B+slurf  
HTTP/1.1" 200 33209 "http://anw.inl.nl/show?page=search2" "Mozilla/5.0 (compatible; MSIE  
9.0; Windows NT 6.1; Trident/5.0; BOIE9;NLNL)"
```

➤ Google Analytics 4/20



Logfile analysis gives a lot of information but doesn't tell you everything. They don't tell you for instance whether the users like the interface. They only contain information about the page views the user generates. A page view occurs whenever you click a link or submit a form. ANW entries have a tree view on the left. Clicking in the tree changes the view of the article for the user, but does not generate another page view. Therefore, we don't know *what* users do while viewing an article, only *that* they viewed the article.

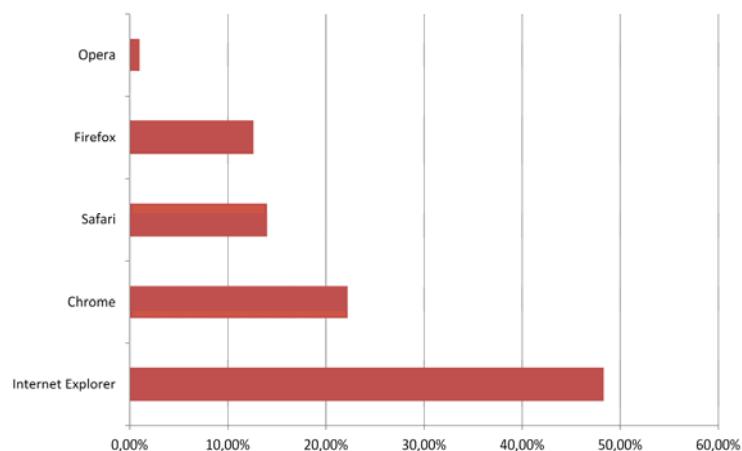
Period: 12/2009 – 3/2013

- Number of pageviews: ± 2,088,000
- Number of searches: ± 236,000
- Number of unique IP-addresses: ± 591,000
- Number of sessions: ± 857,000



About half of requests on the ANW are from robots. The vast majority are from Google. The number 2, Microsoft's Bing, doesn't visit quite as often (a little over half as much). Many automated hack-attempts are also visible in the logs, searching for common vulnerabilities in widely used software. Occasionally, application errors occur in the logs, and users do not always report them to us. Tracking these errors in the log will help us fix errors even if users do not report them.

Used browsers



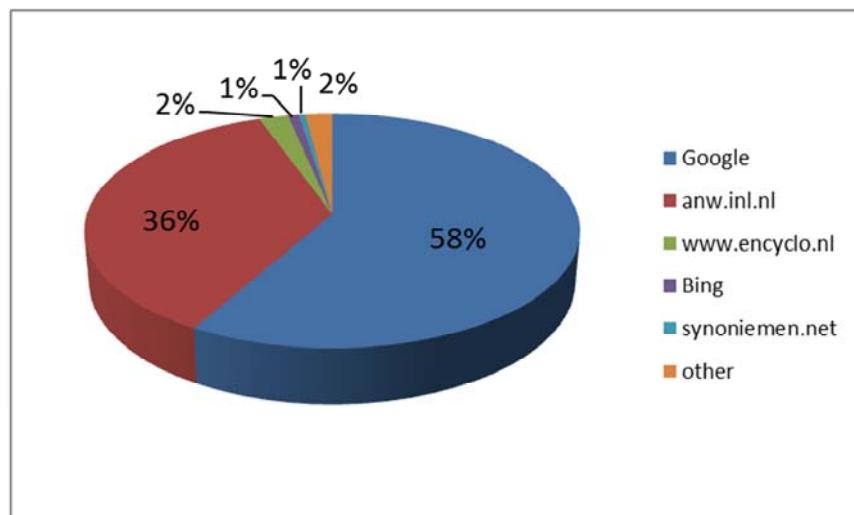
Most users use Internet Explorer, further use is mainly made of Chrome, Safari and Firefox.

Desktop versus mobile



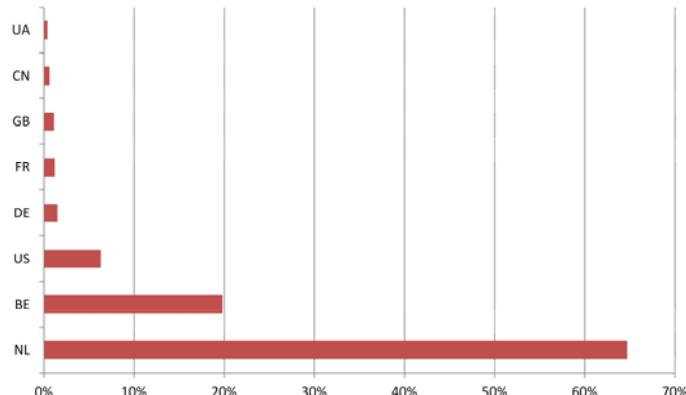
Mobile usage is around 10%, up from 2% a few years ago. We have made the ANW a bit more mobile-friendly, which might be a contributing factor.

Referrers



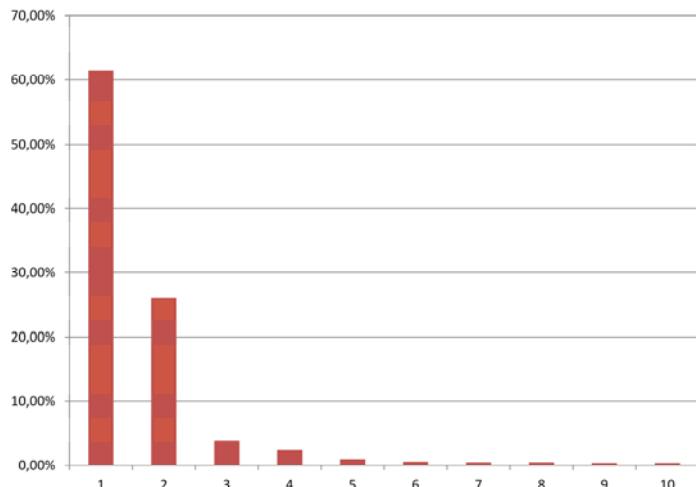
About 60% of users initially arrive at the ANW via a bookmark. Almost 40% arrive via Google. Around 3% arrive via Encyclo, an online encyclopedia that links to the ANW. From the IP addresses, we can conclude that over 2% of users come from the National Library of the Netherlands.

Users by country



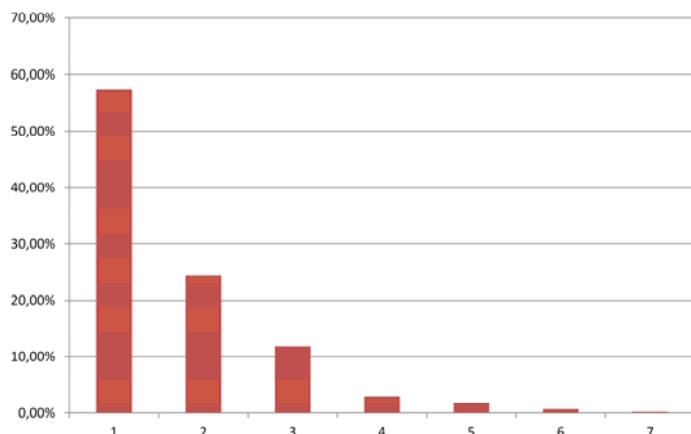
Also from IP addresses, we can see what country people come from. Not surprisingly, 65% comes from the Netherlands, 20% from Belgium. About 6% comes from the US. Germany accounts for about 1,5% of ANW users. Also notable are 0,2% from both Surinam and the Netherland Antilles.

Pageviews per session



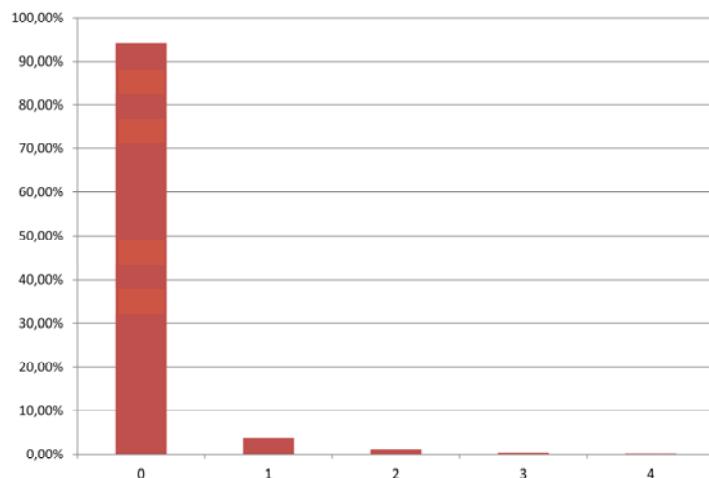
60% of users only view a single page. Of course, we don't know how long they look at this page; if they are viewing an article, it might be several minutes. Most user sessions are short, at most a few minutes, even if they view multiple pages. However, there are people who spent more than an hour on the site.

Entries viewed per session



Around 60% of visitors view at least one article. Over 20% view 2 or more, and a handful of users viewed more than 50 articles.

Searches per session

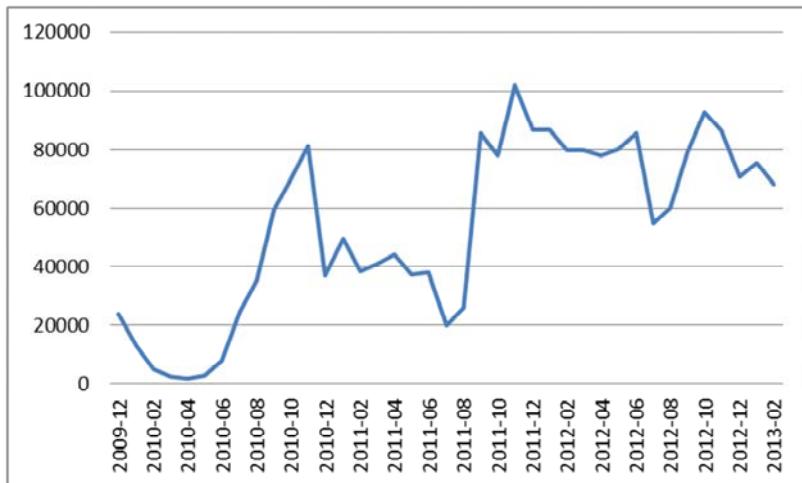


Only 6% of users perform a search at all. This makes sense as most people arriving from Google immediately see an article page.



About 1 in 50 people look at help pages, a number that is actually encouraging, because we thought no-one ever looked at them.

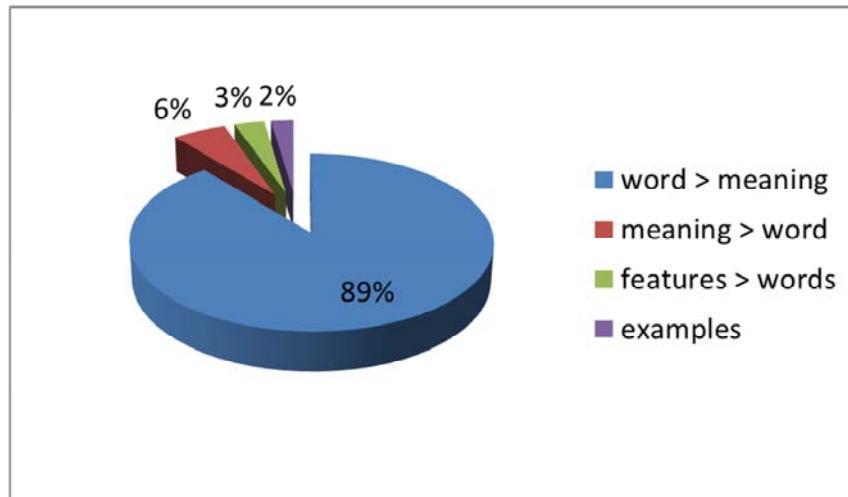
When is the ANW used?



We can see that there are peaks in use at the moment there is a special action. Peak 1: October 2010, the launch of an INL dictionary game; peak 2: October 2011, the election of the neologism of the year; peak 3: October 2012, the election of the nicest and ugliest word of the year. After these actions the use decreases somewhat, but remains higher than it was before the action.

Users appear to be the most active searching the ANW on Tuesdays and Wednesdays and then especially between 13.00 and 16.00 hours.

Searching the ANW



In earlier analyses we observed a clear tendency towards the traditional search from word > meaning. Looking at the more recent data, we see that the 'new' options are increasing.

Word > Meaning top 30 - 2012

- | | | |
|---------------------|------------------------|--------------------------|
| 1. b*r | 12. Qaidastrijder | 23. hybridekameel |
| 2. bi?r* | 13. koe | 24. mogelijkheid |
| 3. googelen | 14. Q | 25. hond |
| 4. sterrenkind | 15. balen | 26. lolbroekerij |
| 5. proactief | 16. q | 27. boek |
| 6. ook niet | 17. opportuun | 28. algemeen |
| 7. grexit | 18. voorproefje | 29. waarheidsgetrouw |
| 8. Stool | 19. yammeraar | 30. aap |
| 9. schermtijd | 20. pandapunten | |
| 10. verstarring | 21. aardhommel | |
| 11. y | 22. huis | |

You see that people look up examples from the help function of the dictionary itself (*b*r*, *bi?r*, *koe*, etc.) and words that were part of games and elections (*googelen*, *proactief*, *grexit*, etc.). But luckily also ‘normal’ words are searched.

Features > Words top searches 2012

- Time: neologism
- Language variety: (mainly) in Belgium
- Pronunciation manner: German
- Origin: loanword

You can also see how many features are combined in the searches, mostly people use only 1 feature, in 15% of the searches 2 features are combined, in 4% 3 features are combined and in less than 1% 4 or more features are combined in a search.

The most used features in search combinations are part of speech, definition, language variaty, style and time.

Conclusions after logfile analysis

- We should try to accommodate both older browsers and modern mobile devices.
- Security matters.
- Search engine optimization and strategic partnerships with popular sites are the most promising way of increasing traffic.
- We only have a short time to ‘hook’ our users. The interface should be self-explanatory and engaging.

Some publications on the ANW

Tanneke Schoonheim and Rob Tempelaars (2010),
'Dutch Lexicography in Progress, The *Algemeen
Nederlands Woordenboek (ANW)*'. In: Anne Dykstra
and Tanneke Schoonheim (eds.), *Proceedings of the XIV
Euralex International Congress. Ljouwert.*
[http://www.euralex.org/elx_proceedings/Euralex2010/
059_Euralex_2010_3_SCHOONHEIM
TEMPELAARS Dutch Lexicography in Progress the
Algemeen Nederlands Woordenboek ANW.pdf](http://www.euralex.org/elx_proceedings/Euralex2010/059_Euralex_2010_3_SCHOONHEIM_TEMPелаars_Dutch_Lexicography_in_Progress_the_Algemeen_Nederlands_Woordenboek_ANW.pdf)

Some publications on the ANW

Jan Niestadt (2009), 'De ANW-artikeleditor: software als strategie', in: E. Beijk, e.a. (red.), *Fons verborum*. Leiden/Amsterdam, pp. 215-222.

www.inl.nl/images/stories/onderzoek_en_onderwijs/publicaties/fonsverborum2009/niestadt.pdf

Carole Tiberius en Adam Kilgarriff (2009), 'The Sketch Engine for Dutch with the ANW corpus', in: E. Beijk e.a. (red.), *Fons verborum*. Leiden/Amsterdam, pp. 237-255.

www.inl.nl/images/stories/onderzoek_en_onderwijs/publicaties/fonsverborum2009/tiberius_kilgarriff.pdf