


/instituut voor
de Nederlandse
taal/

Corpus Pattern Analysis for Dutch in Woordcombinaties

Carole Tiberius and Lut Colman
carole.tiberius@ivdnt.org /
c.p.a.tiberius@hum.leidenuniv.nl
lut.colman@ivdnt.org

PhrasaLex III
11/7/2024



 Voorbeeldzinnen

 Combinatiemogelijkheden


 Patronen



- Introduction to the project
- Pattern editing in *Woordcombinaties*
- Challenges:
 - Idioms, conversational routines, metonymy, syntactic variants
 - Automating the process



a unique point of access for anyone who wants to learn more about Dutch phraseology

 Voorbeeldzinnen


 Combinatiemogelijkheden

 Patronen



- <https://woordcombinaties.ivdnt.org>
- Reference work for using and understanding words in context
- Resource for teaching and testing materials
- Advanced learners of Dutch; native speakers
- Verbs (1), nouns (2)



 Voorbeeldzinnen

 Combinatiemogelijkheden

 Patronen




Combination: any meaningful and statistically relevant combination of words with spaces, excluding compounds and combinations not formed in Dutch:

- Collocation: **offer** + support, advice, service, opportunity
- Multiword lexical unit: yellow **spot**
- Idiom: **hand** something to someone on a silver platter
- Proverb: all good things **come** to an end
- Formula: **have** a nice day!
- Semantically motivated pattern:
something **takes place** vs someone **takes his place**

Not considered combinations

- Compounds without spaces
 - *appeltaart* ('apple pie')
 - *kruidje-roer-me-niet* ('touch-me-not', 'plant/touchy person')
- Loan combinations
 - *ad hoc*
 - *ad fundum*
 - *out of the blue*



 Voorbeeldzinnen

Combinatiemogelijkheden

Patronen



- Examples
- Sometimes good examples suffice
 - to work out the meaning
 - to get a bird's-eye view of the usage
 - patterns
 - collocates

(to base)

baseren werkwoord woordvormen

(vaak in het passief)

1. Ik **baseer** me op informatie.
2. Hij **baseert** zijn conclusies dus op een analyse van het aanbod.
3. Deze mythe is zelden of nooit **gebaseerd** op echte feiten.
4. Een solidariteit die **gebaseerd** is op objectieve gegevens.
5. Ze **baseerde** zich op feiten die ik haar zelf had verstrekt.
6. Hun kennis is louter **gebaseerd** op theoretische kennis.
7. Hij **baseert** zich op informatie van de politie.
8. Ons besluit was **gebaseerd** op de volgende overwegingen.
9. Onze hedendaagse vormgeving is hoofdzakelijk **gebaseerd** op deze formele benadering.
10. Dit realistische scenario is uiteraard **gebaseerd** op hedendaagse economische inzichten.
11. Het dividend was **gebaseerd** op de verwachte opbrengst van de verkopen.
12. Het optimisme is louter **gebaseerd** op de hoop op verandering.
13. Wat als mensen conclusies trekken die op niets zijn **gebaseerd**?
14. Veel scènes zijn **gebaseerd** op de ervaringen van mensen.
15. Wetenschap is nooit **gebaseerd** op maar één bron.

SKELL: Sketch Engine for Language Learning



Voorbeeldzinnen

Combinatiemogelijkheden

Patronen



(to base)

baseren

werkwoord

woordvormen

(vaak in het passief)

1. Ik **baseer** me op informatie.
2. Hij **baseert** zijn conclusies dus op een analyse van het aanbod.
3. Deze mythe is zelden of nooit **gebaseerd** op echte feiten.
4. Een solidariteit die **gebaseerd** is op objectieve gegevens.
5. Ze **baseerde** zich op feiten die ik haar zelf had verstrekt.
6. Hun kennis is louter **gebaseerd** op theoretische kennis.
7. Hij **baseert** zich op informatie van de politie.
8. Ons besluit was **gebaseerd** op de volgende overwegingen.
9. Onze hedendaagse vormgeving is hoofdzakelijk **gebaseerd** op deze formele benadering.
10. Dit realistische scenario is uiteraard **gebaseerd** op hedendaagse economische inzichten.
11. Het dividend was **gebaseerd** op de verwachte opbrengst van de verkopen.
12. Het optimisme is louter **gebaseerd** op de hoop op verandering.
13. Wat als mensen conclusies trekken die op niets zijn **gebaseerd**?
14. Veel scènes zijn **gebaseerd** op de ervaringen van mensen.
15. Wetenschap is nooit **gebaseerd** op maar één bron.

Woordvormen (word forms)

(niet samengesteld, optioneel-reflexief)

Niet-reflexief gebruik

presens (tegenwoordige tijd)

ik baseer

je/jij baseert

baseer je/jij?

u baseert (beleefd; singularis en pluralis)

hij/ze/zij/het baseert

we/wij baseren

jullie baseren

ze/zij baseren

imperfectum (verleden tijd)

ik baseerde

je/jij baseerde

hij/ze/zij/het baseerde

we/wij baseerden

jullie baseerden

ze/zij baseerden

verbum auxiliare (hulpwerkwoord) + participium perfectum (voltooid deelwoord)

heeft gebaseerd

participium presens (tegenwoordig deelwoord)

baserend

Linked to INT Central Database

Reflexief gebruik met reflexieve pronomina

presens (tegenwoordige tijd)

ik baseer me/mij

je/jij baseert je

baseer je/jij je?

u baseert zich/u (beleefd; singularis en pluralis)

hij/ze/zij/het baseert zich

we/wij baseren ons

jullie baseren je

ze/zij baseren zich

imperfectum (verleden tijd)

ik baseerde me/mij

je/jij baseerde je

hij/ze/zij/het baseerde zich

we/wij baseerden ons

jullie baseerden je

ze/zij baseerden zich

verbum auxiliare (hulpwerkwoord) + participium perfectum (voltooid deelwoord)

heeft zich gebaseerd

participium presens (tegenwoordig deelwoord)

zich baserend

Woordcombinaties

Voorbeeldzinnen

Combinatiemogelijkheden

Patronen



Post-edited Word Sketch

(to whistle)

fluiten werkwoord woordvormen = idiomatisch

subject ① - *Wie of wat fluit?*

substantief ①

arbiter kogel publiek ref scheidsrechter vogel wind
(arbiter) (bullet) (public) (ref) (referee) (bird) (wind)

object ① - *Wie of wat fluit men of wordt gefloten?*

substantief ①

deuntje duel finale fout liedje match melodie partij penalty strafschop (2 meer)

bepaling ① - *Waar, wanneer, hoe, enz. fluit men?*

adverbium ①

goed hard vrolijk

prepositiegroep ①

in: nadeel voordeel

naar: centen geld meisje

om: [de, je] oer

op: vingers

voor: overtreding rust fout strafschop

subject ① - *Wie of wat fluit?*

substantief ①

arbiter kogel publiek ref scheidsrechter vogel wind

1. Er staan bomen om mij heen en er **fluiten** vogels.
(birds are singing)

prepositiegroep ①

in: nadeel voordeel

naar: centen geld meisje

naar je [centen, geld] kunnen fluiten: niet krijgen waar je recht op hebt of waar je op hoopt

1. Sinds de schaduwbank vorige week failliet werd verklaard, kunnen ze naar hun geld **fluiten**.
(they can whistle for their money)



(to demonstrate)
aantonen werkwoord Toon:

1 (someone or something demonstrates something like this)
iemand of iets toont iets (zo) aan
 iemand of iets laat zien of bewijst dat iets aanwezig is, het geval is, of waar is
De test toonde een duidelijk verband aan tussen welvaart en financiële kennis.
[\(meer voorbeelden\)](#)

aantonen werkwoord Toon:

1 **iemand of iets toont iets (zo) aan**
 iemand of iets laat zien of bewijst dat iets aanwezig is, het geval is, of waar is
De test toonde een duidelijk verband aan tussen welvaart en financiële kennis.
[\(meer voorbeelden\)](#)

1.a **iemand of iets toont (zo) aan dat-zin of vraagwoordzin of of-zin**
 iemand of iets laat zien of bewijst dat, of of, of hoe e.d. iets het geval is
Dit document toont aan dat alle belastingen zijn voldaan.
[\(meer voorbeelden\)](#)

1.b **iets of of-zin is of blijkt of valt (zo) aan te tonen**
Of dit waar is, valt lastig aan te tonen.
[\(meer voorbeelden\)](#)

Patterns

(to leave behind) only idioms, proverbs and formulas
achterblijven werkwoord Toon:

4.a uitdrukking (someone is left empty-handed)
iemand blijft met lege handen achter
 iemand krijgt niets, of heeft zijn doel niet bereikt
Ondanks hun verwoede poging bleven ze met lege handen achter.
[\(meer voorbeelden\)](#)

(to think) only idioms, proverbs and formulas
denken werkwoord Toon:

1.e formule **iemand denkt het (niet of wel)**
 iemand vermoedt of is van mening van niet of van wel
Wie weet – maar ik denk het niet.
[\(meer voorbeelden\)](#)

1.f formule **ik dacht het niet!**
 geen sprake van; dat gaat niet gebeuren
"Ik dacht het niet", verklaart hij strijdbaar.
[\(meer voorbeelden\)](#)

 Voorbeeldzinnen

 Combinatiemogelijkheden

 Patronen



aantonen

werkwoord

Toon: hoofdpatronen

object / Eventuality

1

iemand of iets **toont** **iets** **(zo)** **aan** (someone or something demonstrates something like this)

onschuld (innocence)

Later verklaarde hij het onderzoek juist toe te juichen om zo zijn onschuld te kunnen aantonen .

relatie (relation)

Er is een relatie aangetoond tussen voeding en gezondheid.

tegendeel (contrary)

Wij moesten de bewijslast leveren om het tegendeel aan te tonen .

verband (link)

Het verband is ook niet aangetoond in internationale studies.

verschil (difference)

Zo kan je ook het verschil aantonen tijdens de spitsuren en dalmomenten.

1

iemand of iets **toont** **iets** **(zo)** **aan** (someone or something demonstrates something like this)

afdoend (sufficient)

Het verband tussen alcohol en geweld thuis en/of in uitgaansgebieden is afdoende aangetoond .

al langer (for some time)

De endorfinen die bij zo'n loopje vrijkomen, helpen een mens om zich beter te voelen, dat is al langer aangetoond .

daadwerkelijk (actually)

Bij nog eens 200.000 euro kon de verzekeraar daadwerkelijk fraude aantonen .

definitief (definitively)

Als we dat met een verklaring van een arts definitief aantonen , kan volgens mij de operatie beginnen."

duidelijk (clearly)

Het wordt steeds duidelijker door fysiologen aangetoond , steeds weer opnieuw.

echt (really)

Planeten bij andere sterren zijn nog nooit echt aangetoond .

Inspiration

- **SkE_{LL}**: Sketch Engine for Language Learning (Lexical Computing, Brno)
- **PDEV**: Pattern Dictionary of English Verbs (Patrick Hanks, Wolverhampton)
- **T-PAS**: Typed Predicate Argument Structures (E. Ježek, Pavia)
- **E-VALBU**: Das elektronische Valenzwörterbuch deutscher Verben (IDS, Mannheim)
- **CVVD**: Contrastive Verb Valency Dictionary (Contragram, Gent)
- **SweCcn**: Swedish Constructicon (Språkbanken, Gothenburg)

Tools & methodology: the corpus

Woordcombinaties is corpus-based:

- Approx. 200 million tokens
- Mainly newspaper and web
- Material from the Netherlands and Belgium
- parsed with Alpino parser
- Uploaded to Sketch Engine

```
<s sentid="1">
<g/>
1   '   LET()   '   let 5   --
2   Een LID(onbep,stan,agr) een-t   lid 4   det
3   echte   ADJ(prenom,basis,met-e,stan)   echt-a   adj 4   mod
4   staatsman   N(soort,ev,basis,zijd,stan)   staatsman-n n   5   su   7   su
5   zou WW(pv,verl,ev)   zullen-v   ww 0   root
6   nu BW()   nu-r   bw 7   mod
7   opstappen   WW(inf,vrij,zonder)   op_stappen-v   ww 5   vc
<g/>
8   '   LET()   '   let 5   --
</s>
```



Tools & methodology

- GDEX
 - Good Dictionary Examples
- TBL
 - TickBox Lexicography
- Word sketches

SUBJECT substantief		
<input type="checkbox"/> concern Het concern boekte	24	7.7 ...
<input type="checkbox"/> beurs Europese beurzen boekten	23	7.0 ...
<input type="checkbox"/> verzekeraar verzekeraar boekte	18	6.7 ...
<input type="checkbox"/> theater theaters boekten	9	6.6 ...
<input type="checkbox"/> toerist Toeristen boekten	11	6.3 ...
<input type="checkbox"/> passagier passagier kan voor een redelijke prijs een zetel boeken	9	6.2 ...
<input checked="" type="checkbox"/> bank bank boekte	55	6.2 ...

CONCORDANCE CWParsedAlpino6

lemma **adverteren** • 878
4.22 per million tokens • 0.00042%

Sort GDEX (GDEXWoordcombinaties)

Details

1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Grote bedrijven adverteren steeds met vaker met online video's.
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Bedrijven adverteren liever online dan op tv tegenwoordig.
3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Beide winnaars mogen in het najaar nog driemaal paginagroot en gratis adverteren .
4	<input type="checkbox"/>	<input type="checkbox"/>	De bedrijven sturen de gegevens naar Facebook zodat ze gericht kunnen adverteren .
5	<input type="checkbox"/>	<input type="checkbox"/>	Bedrijven adverteren via de online videomakers en betalen hen om hun producten aan te prijzen.
6	<input type="checkbox"/>	<input type="checkbox"/>	Welke Nederlandse bedrijven adverteren via Facebook?
7	<input type="checkbox"/>	<input type="checkbox"/>	Dat maakt sport bovendien zeer aantrekkelijk voor bedrijven die willen adverteren .
8	<input type="checkbox"/>	<input type="checkbox"/>	Deze diensten geven bedrijven de mogelijkheid om gericht te adverteren .
9	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Ze verhuren ramen aan bedrijven, zodat die kunnen adverteren .

Select on page: ALL NONE show only: SELECTED (4) NOT SELECTED COPY (4)

Tools & methodology: DWS

- INT-DWS: Example sentences and Combinations
- SKEMA: Patterns

The screenshot displays the COMBI Editor interface for the article 'aanbieden'. The left pane shows a tree view of linguistic combinations, with 'subject > substantief (20)' selected. The right pane provides a detailed view of this combination, including its role ('subject > substantief'), various subtypes, and example sentences. The example sentences are:

- Realisatie (body): bank
- Realisatie (body): bedrijf
- Realisatie (body): gemeente

Tools & methodology: Patterns

- **TNE:** Theory of Norms and Exploitations (Hanks 2013)
 - Words have meaning potentials rather than meanings
 - Meanings are evoked by context
 - Norms are ‘normal’ phraseological patterns as main carriers of meaning
 - Exploitations are creative uses of normal patterns
- **CPA:** Corpus Pattern Analysis
 - Corpus-driven technique for mapping meaning onto words in context
 - Sketch Engine supports annotation of usage patterns in corpus samples

Patterns in *Woordcombinaties*

- Corpus-driven approach > corpus-based in *Woordcombinaties*
 - Random sample > GDEX sorted sample of 250 lines

Wie alles goed vindt, **boek** 2 geen vooruitgang meer.

Dat de hele ploeg door hem geen vooruitgang **boek** 2 .

Door dat wereldwijde engagement is er ook daadwerkelijk resultaat **geboekt** 2 .

In onze voorbereidingen valt niet veel winst meer te **boeken** 2 .

Er werd te weinig vooruitgang **geboekt** 2 , zei hij.

Tegelijk konden de beurzen op Wall Street geen winst **boeken** 1 .

Ze staan te kijken van de vooruitgang die we hebben **geboekt** 2 .

Het **boek** 1 een operationele winst van 12 miljoen euro.

Desondanks **boek** 1 het aandeel gisteren een winst van 5 procent.

KV Mechelen kon thuis alweer geen overwinning **boeken** 2 .

Hoe lang zal het duren eer we succes **boeken** 2 ?

Pattern editing

- SKEMA-editor
 - Sketch Engine Manual Annotation (V. Baisa et al. 2020)
- a **corpus pattern editor system** which supports:
 - the manual **annotation of concordance lines** with user-defined pattern labels
 - the **editing of patterns** (in terms of slots, attributes, examples...)
- Each **annotation label** is **linked** to its **corresponding pattern structure** (stored in JSON format), which can be easily customised to individual projects

Pattern editing: SKEMA

ANNOTATION MANAGER

CWParserAlpino6



boeken



CONCORDANCE SEARCH DOWNLOAD

Annotation	Label count	Frequency	Status	Editor
boeken-v	4	9997	FINISHED	inl02



boeken-v

FINISHED

RENAME LABELS

1	iemand boekt iets iemand schrijft iets in de boekhouding <i>Het bedrijf boekte een omzet van 53,78 miljard euro .</i>	27.33%	
2	iemand of iets boekt iets iemand of iets behaalt iets <i>De jongste jaren heb ik veel vooruitgang geboekt .</i>	61%	
3	iemand boekt iets of iemand iemand reserveert iets of iemand <i>Hoe eerder je een ticket boekt, hoe goedkoper dat is.</i>	11%	
4 <small>idiom</small>	iemand of iets staat geboekt als iemand of iets iemand of iets staat bekend als iemand of iets <i>De N-VA-voorzitter staat niet geboekt als een flapuit.</i>	0.67%	

Pattern editing: SKEMA

ANNOTATION MANAGER

CWParserAlpino6



boeken



CONCORDANCE SEARCH DOWNLOAD

Annotation	Label count	Frequency	Status	Editor
boeken-v	4	9997	FINISHED	inl02



boeken-v

FINISHED

RENAME LABELS

1	iemand boekt iets iemand schrijft iets in de boekhouding <i>Het bedrijf boekte een omzet van 53,78 miljard euro .</i>	27.33%	
2	iemand of iets boekt iets iemand of iets behaalt iets <i>De jongste jaren heb ik veel vooruitgang geboekt .</i>	61%	
3	iemand boekt iets of iemand iemand reserveert iets of iemand <i>Hoe eerder je een ticket boekt, hoe goedkoper dat is.</i>	11%	
4	iemand of iets staat geboekt als iemand of iets iemand of iets staat bekend als iemand of iets <i>De N-VA-voorzitter staat niet geboekt als een flapuit.</i>	0.67%	

Slots are represented by **dummies** (indefinite pronouns) in the pattern: *iemand* 'someone', *iets* 'something', *zo* 'like this'
≠ semantic types as in PDEV, etc.

Pattern editing: SKEMA

ANNOTATION MANAGER

CWParserAlpino6



boeken



CONCORDANCE SEARCH DOWNLOAD

Annotation	Label count	Frequency	Status	Editor	
boeken-v	4	9997	FINISHED	inl02	

boeken-v

FINISHED

RENAME LABELS


1		iemand boekt iets iemand schrijft iets in de boekhouding <i>Het bedrijf boekte een omzet van 53,78 miljard euro .</i>	27.33%	
2		iemand of iets boekt iets iemand of iets behaalt iets <i>De jongste jaren heb ik veel vooruitgang geboekt .</i>	61%	
3		iemand boekt iets of iemand iemand reserveert iets of iemand <i>Hoe eerder je een ticket boekt, hoe goedkoper dat is.</i>	11%	
4	idiom	iemand of iets staat geboekt als iemand of iets iemand of iets staat bekend als iemand of iets <i>De N-VA-voorzitter staat niet geboekt als een flapuit.</i>	0.67%	



Pattern editing: SKEMA

boeken-v

1 **Label type** normal Attitude Style Domain Variant

1 **Function** *iemand boekt iets*
iemand schrijft iets in 
Het bedrijf boekte een omzet van 53,78 miljard euro.

2 **Semantic type** *iemand of iets boekt iets*
iemand of iets behaalt iets
De jongste jaren heb ik veel boeken gelezen.

3 **Fixed element** *iemand boekt iets of iemand reserveert iets of iemand staat geboekt*
Hoe eerder je een ticket boekt, hoe goedkoper is het.

4 **Dummy** *iemand of iets staat geboekt*
De N-VA-voorzitter staat geboekt.

Lexical set

Attributes

Implicature
iemand schrijft iets in de boekhouding

Synonym

Meaning
in de boekhouding schrijven

Example	Lexical item	Slot	PRAGMATICS
Example 1 an.linked Het bedrijf boekte een omzet van 53,78 miljard euro.	Lexical item 1		+ PRAGMATICS
Example 2 an.linked De bank boekte een winst van 1,4 miljard euro.	Lexical item 2 bank	subject (1)	+ PRAGMATICS
Example 3 an.linked Het bedrijf boekte een omzet van 53,78 miljard euro.	Lexical item 3 bedrijf	subject (1)	+ PRAGMATICS

Pattern editing: SKEMA

boeken-v

1 iemand boekt iets
iemand schrijft iets in
Het bedrijf boekte een omzet van 53,78 miljard euro.

2 iemand of iets boekt iets
iemand of iets behaalt iets
De jongste jaren heb ik veel boeken gelezen.

3 iemand boekt iets of iemand reserveert iets of iemand staat geboekt
Hoe eerder je een ticket boekt, hoe goedkoper het is.

4 idiom iemand of iets staat geboekt
iemand of iets staat bekeerd
De N-VA-voorzitter staat bekeerd.

1	Attitude	Style	Domain	Variant
Label type	normal			
Function				
Semantic type	Human Institution	Head	Object Money Value	
Fixed element				
Dummy				
Lexical set				
Attributes				
Implicature				
Example 1 an.linked				
Example 2 an.linked				
Example 3 an.linked				

Ontology

- ANYTHING
ST to use as a last resort when [Eventuality], [Entity] and [Property] are equally likely
- ENTITY
[Anything] that exists independently of other things and has a distinct identity. [Anything] which is not an [Eventuality] nor a [Property]
forza, materiale, ambiente, economia, creatura, edificio
- PHYSICAL ENTITY
A tangible [Entity]
ponte, faccia, tavolo, auto, fiore, uccello birra, merci, pietra, bambino, vulcano
- + INANIMATE
A non-living [Physical Entity] which is not a [Location], such as [Stuff] or [Artifact]s
foglia, pipa, bottone, vestito, aria, gas, vaso, quadro, cibo
- ANIMATE
All living [Physical Entity]s except [Plant]s
animale, persona, bambino, ospite, leone, capra, elefante, cane, gatto
- HUMAN**
An [Animate] [Physical Entity]: either man, woman, or child
persona, bambino, vecchio, signora, capo, maestra, dottore, figlio, zia, mamma

Lexical item 3

bedrijf

Slot

subject (1)

PRAGMATICS

SHOW ONTOLOGY
ANNOTATE "boeken-v"

Pattern editing: SKEMA

boeken-v

1 iemand boekt iets
iemand schrijft iets in
Het bedrijf boekte een om.

2 iemand of iets boekt iets
iemand of iets behaalt iets
De jongste jaren heb ik ve

3 iemand boekt iets of iema
iemand reserveert iets of i
Hoe eerder je een ticket b

4 idiom iemand of iets staat gebo
iemand of iets staat beker
De N-VA-voorzitter staat n

1	2	3
Label type normal	Attitude	Style
Function	Domain	Variant
Semantic type	Subject	Head
Fixed element	Object	
Dummy	Human	Money Value
Lexical set	Institution	
Attributes	iemand	iets
	boek	
	<input type="checkbox"/> opt <input type="checkbox"/> or	<input type="checkbox"/> opt
Implicature	Meaning	
iemand schrijft iets in de boekhouding	in de boekhouding schrijven	
Synonym		
Example 1 an.linked	Lexical item 1	Slot
Het bedrijf boekte een omzet van 53,78 miljard euro .		
Example 2 an.linked	Lexical item 2	Slot
De bank boekte een winst van 1,4 miljard euro.	bank	subject (1)
Example 3 an.linked	Lexical item 3	Slot
Het bedrijf boekte een omzet van 53,78 miljard euro.	bedrijf	subject (1)

Pattern editing: SKEMA



boeken-v

1 *iemand boekt iets*
iemand schrijft iets in de
Het bedrijf boekte een omzet van 53,78 miljard euro.

2 *iemand of iets boekt iets*
iemand of iets behaalt iets
De jongste jaren heb ik van de bank boekte een winst van 1,4 miljard euro.

3 *iemand boekt iets of iem*
iemand reserveert iets of
Hoe eerder je een ticket boekt, hoe goedkoper is het.

4 *iemand of iets staat gel*
iemand of iets staat bekend
De N-VA-voorzitter staat bekend als een succesvol politicus. idiom

1 Label type: normal Attitude: Style: Domain: Variant:

Function: Subject Head Object

Semantic type: Human Institution Money Value

Fixed element: iemand boekt iets

Dummy: Lexical set: Attributes: opt or

Implicature: iemand schrijft iets in de boekhouding

Synonym: in de boekhouding schrijven

Example 1 [an.linked](#) Lexical item 1: Slot: PRAGMATICS

Example 2 [an.linked](#) Lexical item 2: bank Slot: subject (1) PRAGMATICS

Example 3 [an.linked](#) Lexical item 3: bedrijf Slot: subject (1) PRAGMATICS

SHOW ONTOLOGY ANNOTATE "boeken-v"

annotate in word sketch

SUBJECT substantief		
<input type="checkbox"/>	concern _	24 7.7 ...
	Het concern boekte	
<input type="checkbox"/>	beurs _	23 7.0 ...
	Europese beurzen boekten	
<input type="checkbox"/>	verzekeraar _	18 6.7 ...
	verzekeraar boekte	
<input type="checkbox"/>	theater _	9 6.6 ...
	theaters boeken	
<input type="checkbox"/>	toerist _	11 6.3 ...
	Toeristen boekten	
<input type="checkbox"/>	passagier _	9 6.2 ...
	passagier kan voor een redelijke prijs een zetel boeken	
<input checked="" type="checkbox"/>	bank _	55 6.2 ...
	bank boekte	
<input type="checkbox"/>	multinational _	7 6.2 ...
	het derde kwartaal boekte de Nederlands-Britse multinational een	
<input checked="" type="checkbox"/>	bedrijf _	126 6.1 ...
	bedrijf boekte	
<input type="checkbox"/>	luchtvaartmaatschappij _	6 6.1 ...
	luchtvaartmaatschappij van Egypte één miljard dollar verlies geboekt	
<input type="checkbox"/>	voortgang _	7 6.1 ...
	voortgang boekt	
<input type="checkbox"/>	keten _	6 6.0 ...
	boekte de keten	
select all		
deselect all		



Tickbox lexicography - select examples (lemma: boeken)

TBL template ANW_INL

GDEX configuration GDEXWoordcombinaties

([select all](#) | [deselect all](#))

↔

SUBJECT substantief

- concern** Het concern boekte
- beurs** Europese beurzen boekten
- verzekeraar** verzekeraar boekte
- theater** theaters boeken
- toerist** Toeristen boekten
- passagier** passagier kan voor een redelijke prijs een zetel boeken
- bank** bank boekte
- multinational** het derde kwartaal boekte de Nederlands-Britse multinational een
- bedrijf** bedrijf boekte
- luchtvaartmaatschappij** luchtvaartmaatschappij van Egypte één miljard dollar verlies geboekt
- vooruitgang** vooruitgang boekt
- keten** boekte de keten

[select all](#) ▼

[deselect all](#)

SUBJECT substantief

• bank

- De **bank boekte** een winst van 1,4 miljard euro.
- De **banken** in de VS **boeken** opnieuw evenveel winst als voor 2008.
- De **bank boekte** in 2015 een winst van ruim 4,5 miljard euro.
- De **bank** met hoofdkantoor in Utrecht **boek** op vrijwel alle onderdelen vooruitgang.
- De **bank boekte** vorig kwartaal 383 miljoen euro winst.
- De meeste **banken boekten** die als onmiddellijke winst.

• bedrijf

- Vorig jaar **boekte** het **bedrijf** een verlies van vijf miljoen euro.
- In 2015 **boekte** het **bedrijf** een verlies van 180 miljoen euro.
- Het **bedrijf boekte** vorig jaar een winst van 4,6 miljard euro.
- Vorig jaar **boekte** het **bedrijf** een winst van bijna 600 miljoen euro.
- Het **bedrijf boekte** een omzet van 53,78 miljard euro.
- Het **bedrijf boekte** een winst van 792 miljoen dollar in het afgelopen kwartaal.



Pattern editing: SKEMA

1
Label type: normal

Attitude: _____ Style: _____ Domain: _____ Variant: _____

	1	2	3
Function	Subject	Head	Object
Semantic type	Human Institution		Money Value
Fixed element			
Dummy	iemand	boekt	iets
Lexical set			
Attributes	<input type="checkbox"/> opt <input type="checkbox"/> or		<input type="checkbox"/> opt
Implicature	iemand schrijft iets in de boekhouding		Meaning: in de boekhouding schrijven
Synonym	_____		
Example 1	Lexical item 1	Slot	PRAGMATICS
Het bedrijf boekte een omzet van 53,78 miljard euro .			
Example 2	Lexical item 2	Slot	PRAGMATICS
De bank boekte een winst van 1,4 miljard euro.	bank	subject (1)	
Example 3	Lexical item 3	Slot	PRAGMATICS
Het bedrijf boekte een omzet van 53,78 miljard euro.	bedrijf	subject (1)	

Challenges

- Idioms and conversational routines
- Metonymy
- Syntactic variants

Patterns and Syntactic Variants

- *iemand strooit iemand zand in de ogen*
 - *iemand strooit zand in de ogen van iemand*
 - *iemand strooit zand in **iemands** ogen*
- = throw dust in someone's eyes / throw dust in the eyes of someone

LABEL
Label type: normal

Attitude: _____ Style: _____ Domain: _____ Variant: _____ Auxiliary: _____

	1	2	3	4	5
Function	Subject	Head	Indirect object	Object	Separated verbal part
Semantic type	Human Institution				
Fixed element					
Dummy	iemand	strooit	iemand	zand	in de ogen
Lexical set					
Attributes	<input type="checkbox"/> opt <input type="checkbox"/> or		<input type="checkbox"/> opt <input type="checkbox"/> or	<input type="checkbox"/> opt <input type="checkbox"/> or	<input type="checkbox"/> opt

Patterns and Metonymy

drinken-v

FINISHED

- 1** *iemand of een dier drinkt*
iemand of een dier neemt vloeistof tot zich door de mond
Je moet eten en drinken, en warm en veilig blijven.
- 2** *iemand of een dier drinkt iets of zo veel*
iemand of een dier nuttigt iets of zo veel van iets als drank of vloeistof, of heeft de gewoonte een bepaalde drank of vloeistof te nuttigen
We drinken een glas witte wijn.

2
Label type: normal
Attitude: _____
Style: _____
Domain: _____
Variant: _____

	1	2	3
Function	Subject	Head	Object
Semantic type	Human Animal		Beverage Liquid Quantity
Fixed element			
Dummy	iemand	drinkt	iets

Lexic Example 37 [an.linked](#) Lexical item 37 **Source: Container**
Attrit Er zijn weinig jongeren die later pas hun eerste glas alcohol drinken . glas **Target: Quantity** Slot object (3)

Example 38 [an.linked](#) Lexical item 38
Ook al blijkt uit studies dat geheelonthouders korter leven dan mensen die een kleine hoeveelheid hoeveelheid Slot object (3)

Patterns and Idioms and conversational routines: 'old' status

Included at the microstructural level

Combinations

fluiten werkwoord woordvormen = idiomatisch idiomatic

subject ⓘ - Wie of wat fluit?

substantief ⓘ

arbitrèr kogel publiek ref scheidsrechter vogel wind

object ⓘ - Wie of wat fluit men of wordt gefloten?

substantief ⓘ

deuntje duel finale fout liedje match melodie partij penalty strafschop (2 meer)

bepaling ⓘ - Waar, wanneer, hoe, enz. fluit men?

adverbium ⓘ

goed hard vrolijk

prepositiegroep ⓘ

in: nadeel voordeel

naar: centen geld meisje

om: [de, je] oren

op: vingers

voor: overtreding rust fout strafschop

In-house editor

Patterns

Only idioms, proverbs and formulas

achterblijven werkwoord Toon: alleen uitdrukkingen, spreekwoorden en formules

4.a uitdrukking **iemand blijft met lege handen achter**

iemand krijgt niets, of heeft zijn doel niet bereikt

Ondanks hun verwoede poging bleven ze met lege handen achter.

(meer voorbeelden)

denken werkwoord Toon: alleen uitdrukkingen, spreekwoorden en formules

1.e formule **iemand denkt het (niet of wel)**

iemand vermoedt of is van mening van niet of van wel

Wie weet – maar ik denk het niet.

(meer voorbeelden)

1.f formule **ik dacht het niet!**

geen sprake van; dat gaat niet gebeuren

"Ik dacht het niet", verklaart hij strijdbaar.

(meer voorbeelden)

SKEMA editor

Idioms and conversational routines: new

- Treatment at macrostructural level
- Record occurrences of idioms and conversational routines in a systematic way in *Woordcombinaties*; Full treatment of idioms and conversational routines separate module
- New search options
 - search for idioms based on image categories (e.g. body parts, food)
een vinger in de pap hebben (‘have a **finger** in the **pie**’)
 - Search for conversational routines based on speech act (e.g. greeting)
goedemorgen (‘good morning’)



Automating the process

 **Woordcombinaties**

Voorbeeldzinnen Combinatiemogelijkheden Patronen

boeken werkwoord Toon: hoofdpatronen

- iemand** **boek** **iets**
iemand schrijft iets in de boekhouding
Het bedrijf boekte een omzet van 53,78 miljard euro.
[\(meer voorbeelden\)](#)
- iemand** **boek** **iets**
iemand behaalt iets
De jongste jaren heb ik veel vooruitgang geboekt.
[\(meer voorbeelden\)](#)
- iemand** **boek** **iets of iemand**
iemand reserveert iets of iemand
Hoe eerder je een ticket boekt, hoe goedkoper dat is.
[\(meer voorbeelden\)](#)
- uitdrukking** **iemand of iets** **staat geboekt** als **iemand of iets**
iemand of iets staat bekend als iemand of iets
De N-VA-voorzitter staat niet geboekt als een flapuit.
[\(meer voorbeelden\)](#)

Pilot study:

Can we use ChatGPT to classify concordance lines under the correct pattern/implicature?

ChatGPT-4o and CPA

You are working on Corpus Pattern Analysis as developed by Patrick Hanks.

You know that the Dutch verb 'boeken' has 4 patterns defined by these implicatures:

1. iemand schrijft iets in de boekhouding
2. iemand of iets behaalt iets
3. iemand reserveert iets of iemand
4. iemand of iets staat bekend als iemand of iets

You have to disambiguate a list of Dutch sentences according to these implicatures, with the output in a json object with a single property "concordances" which is an array containing for each sentence, the sentence and an implicature number.

For each sentence, you return the following information:

Sentence: the sentence itself

Implicature number: the number of the implicature according to the following json format for each sentence:

```
{"text": <sentence>, "implicature": <implicature_number>}
```

Here is an example for the disambiguated output:

```
{  
  "concordances": [  
    [  
      {"text": "Het bedrijf boekte een omzet van 53,78 miljard euro", "implicature": "1"},  
      {"text": "Ze boeken vlucht en hotel apart, zoeken zelf wel een Airbnb", "implicature": "3"},  
      {"text": "Griekenland heeft op een aantal terreinen onvoldoende voortgang geboekt",  
       "implicature": "2"},  
      {"text": "De N-VA-voorzitter staat niet geboekt als een flapuit.", "implicature": "4"},  
    ]  
  ]  
}
```

Now process the following sentences:

We hoeven dus geen hotel meer te boeken als we naar zee willen.

Met de nettowinst van 67 miljoen euro boekt Adecco 40 procent minder winst dan 2019.

In het eerste kwartaal boekte Volvo nog de laagste omzet in vijf jaar tijd.

Er is gestaag vooruitgang geboekt, valt te horen in beide kampen.

In de helft van de gevallen bleken ze resultaat te boeken.

Bierbrouwer Carlsberg boekte minder winst en omzet.

Het ticket kostte maar 250 euro, het hotel moeten we nog boeken.

Niet omdat er rechtstreeks gevaar was, maar die duels stonden geboekt als een risicomatch.

De meeste vooruitgang boekten we op de verdiepingen boven niveau twee.

ChatGPT-4o and CPA

You are working on Corpus Pattern Analysis as developed by Patrick Hanks.

You know that the Dutch verb 'boeken' has 4 patterns defined by these implicatures:

1. iemand schrijft iets in de boekhouding
2. iemand of iets behaalt iets
3. iemand reserveert iets of iemand
4. iemand of iets staat bekend als iemand of iets

You have to disambiguate a list of Dutch sentences according to these implicatures, with the output in a json object with a single property "concordances" which is an array containing for each sentence, the sentence and an implicature number.

For each sentence, you return the following information:

Sentence: the sentence itself

Implicature number: the number of the implicature according to the following json format for

Now process the following sentences:

We hoeven dus geen hotel meer te boeken als we naar zee willen.

Met de nettowinst van 67 miljoen euro boekt Adecco 40 procent minder winst dan 2012.

In het eerste kwartaal boekte Volvo nog de laagste omzet in vijf jaar tijd.

Er is gestaag vooruitgang geboekt, valt te horen in beide kampen.

In de helft van de gevallen bleken ze resultaat te boeken.

Bierbrouwer Carlsberg boekte minder winst en omzet.

Het ticket kostte maar 250 euro, het hotel moeten we nog boeken.

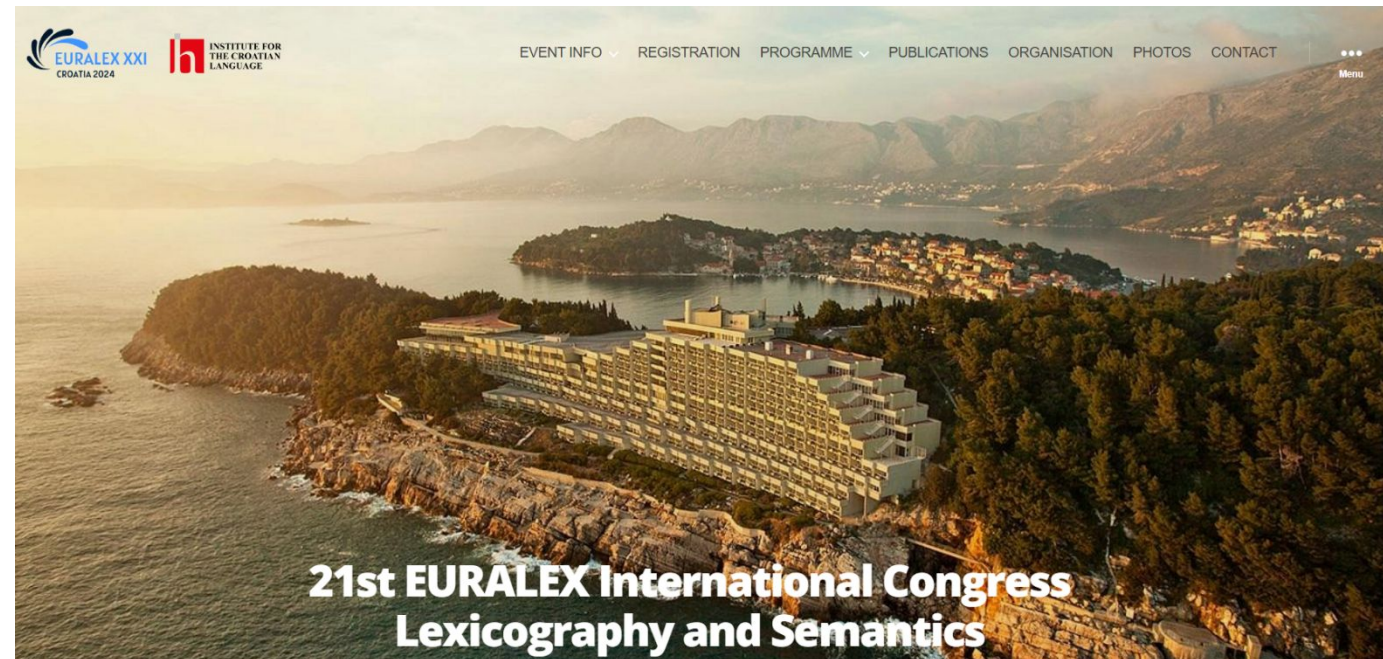
Niet omdat er rechtstreeks gevaar was, maar die duels stonden geboekt als een risicomatch.

De meeste vooruitgang boekten we op de verdiepingen boven niveau twee.

```
"concordances": [  
  {"text": "We hoeven dus geen hotel meer te boeken als we naar zee willen.", "implicature": "3"},  
  {"text": "Met de nettowinst van 67 miljoen euro boekt Adecco 40 procent minder winst dan 2012.", "implicature": "2"},  
  {"text": "In het eerste kwartaal boekte Volvo nog de laagste omzet in vijf jaar tijd.", "implicature": "1"},  
  {"text": "Er is gestaag vooruitgang geboekt, valt te horen in beide kampen.", "implicature": "2"},  
  {"text": "In de helft van de gevallen bleken ze resultaat te boeken.", "implicature": "2"},  
  {"text": "Bierbrouwer Carlsberg boekte minder winst en omzet.", "implicature": "1"},  
  {"text": "Het ticket kostte maar 250 euro, het hotel moeten we nog boeken.", "implicature": "3"},  
  {"text": "Niet omdat er rechtstreeks gevaar was, maar die duels stonden geboekt als een risicomatch.", "implicature": "4"},  
  {"text": "De meeste vooruitgang boekten we op de verdiepingen boven niveau twee.", "implicature": "2"}  
]
```

ChatGPT-4o and CPA

- ... in collaboration with Elisabetta Ježek (University of Pavia)
- ... more at EURALEX





a unique point of access for anyone who wants to learn more about Dutch phraseology

Voorbeeldzinnen

Combinatiemogelijkheden

Patronen



bedanken

werkwoord

woordvormen

1. **Bedankt** voor uw medewerking.
2. Ik wil U hiervoor ten zeerste **bedanken**.
3. Namens het bestuur alvast **bedankt**.
4. Alvast **bedankt** voor de door u te nemen moeite.
5. Ook wilde ik je **bedanken** voor het gratis exemplaar dat je me stuurde.
6. Daarom wil ik hier deze mensen oprecht **bedanken**.
7. Bij voorbaat **bedankt** voor uw aandacht.

Thanks!

On behalf of the *Woordcombinaties* team:
Lut Colman, Jan Niestadt, Carole Tiberius