

Verkenning van semasiologische variatie in verkeerstermen in het Nederlands

Jocelyne Daems

KU Leuven

OG Quantitative Lexicology and Variational Linguistics



KU LEUVEN

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Achtergrond

**Verkenning van semasiologische variatie in verkeestermen
in het Nederlands**

Achtergrond

Verkenning van semasiologische variatie in verkeestermen
in het Nederlands

Achtergrond

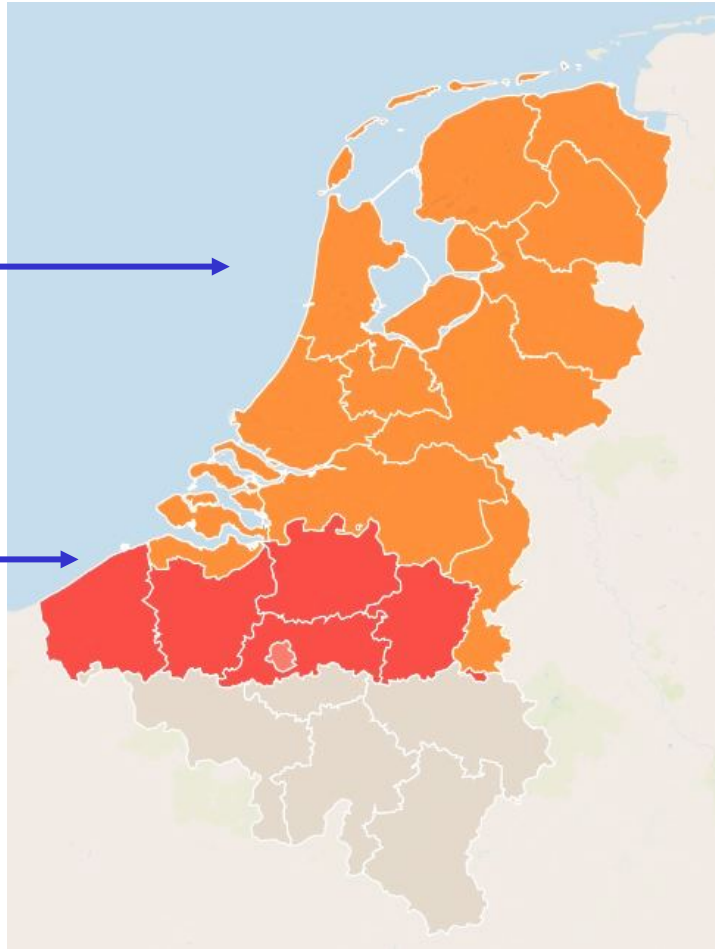
Verkenning van semasiologische variatie in verkeestermen
in het **Nederlands**

Nederlands

Nederlands Nederlands



Belgisch Nederlands



Nederlands

Standaardisatie

- Nederland: Renaissance (cf. Statenbijbel)
- België: vertraging (cf. Franse invloed), inhaalbeweging '60

Stratificatie

- Nederland: vage grens tussen standaardtaal en omgangstaal
- België: supra-regionale tussentaal duidelijk onderscheidbaar van het standaard Belgisch Nederlands

Achtergrond

Verkenning van **semasiologische** variatie in verkeerstermen
in het Nederlands

Achtergrond

Dit is een ...



*vrachtwagen – vrachtauto
camion - truck*

een voertuig

gevaar op de baan





Achtergrond



Wat is...

ring



QML

Achtergrond

2 fenomenen



truck vrachtauto camion vrachtwagen

ring



Achtergrond

2 fenomenen

CONCEPT

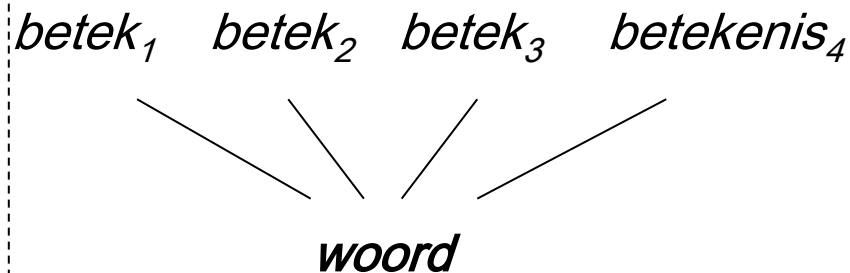
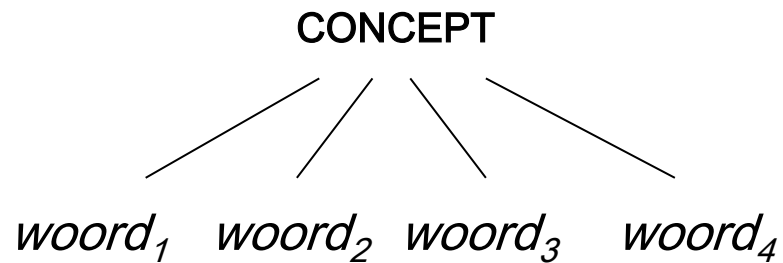
*woord*₁ *woord*₂ *woord*₃ *woord*₄

woord

*betek*₁ *betek*₂ *betek*₃ *betekenis*₄

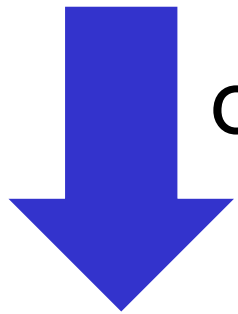
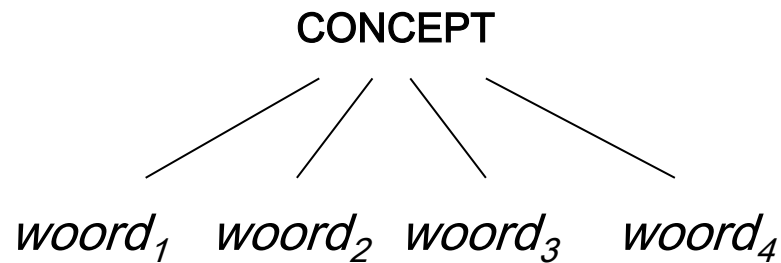
Achtergrond

2 fenomenen



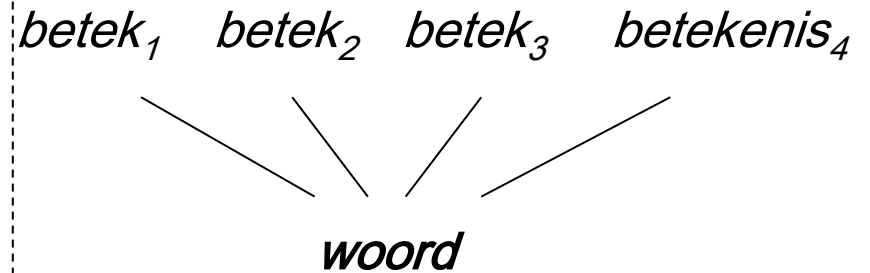
Achtergrond

2 fenomenen



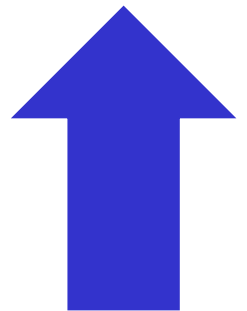
ONOMASIOLOGIE

“benoemen”



SEMASIOLOGIE

“betekenen”



Achtergrond

Lectale variatie

- Tijd: 1950 – 2014 / Oudnederlands – Middelnederlands
- Regio: België – Nederland / Oost-Vlaanderen – Limburg
- Register: spontane conversatie– geredigeerde speech / populaire krant – kwaliteitskrant
- ...

Achtergrond

Lectale variatie

- Tijd: 1950 – 2014 / Oudnederlands – Middelnederlands
- Regio: België – Nederland / Oost-Vlaanderen – Limburg
- Register: spontaan gesproken taal – geredigeerde speech / populaire krant – kwaliteitskrant – internetfora
- ...

Waarom semasiologische variatie?

Theoretisch: compleetheid (onomasiologie + semasiologie)

Methodologisch: aanpassing profielgebaseerde uniformiteitsmaat

Waarom semasiologische variatie?

Theoretisch: compleetheid (onomasiologie + semasiologie)

Methodologisch: aanpassing profielgebaseerde uniformiteitsmaat

Praktisch: schat aan informatie

Onderzoeksvragen en hypothesen

- Ligt de uniformiteit tussen Belgisch Nederlands en Nederlands Nederlands wat betreft woord*betekenis* hoger of lager dan voor woord*keuze*?
- Aan welke factoren is uniformiteit onderhevig? Emotionaliteit, semantisch veld, abstract vs. concreet, conceptfrequentie, ...
- Beïnvloedt de polysemie van de woorden de uniformiteit?
- Wat zijn de voornaamste andere betekenissen, treffen we dezelfde betekenisvernauwingen aan in de verschillende variëteiten en registers, welke concepten of velden laten eerder beeldspraak toe, ... ?

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Data

Vier stappen

- Selectie concepten
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie

1 - Selectie concepten



rondpunt
rotonde



pechstrook
vluchstrook



voetgangersoversteekplaats
zebra zebra-pad

1 - Selectie concepten



convertible
cabrio
cabriolet



fiets
rijwiel
velo
tweewieler



rondpunt
rotonde



pechstrook
vluchstrook



ondergrondse
metro
metrotrein



voetgangersoversteekplaats
zebra
zebrapad

2 - Gelemmatiseerde corpora

['Na', 'prep', 'na', 3849, 0], ['een', 'det', 'een', 3849, 1], ['jaar', 'noun', 'jaar', 3849, 2], ['-', 'punct', '-', 3849, 3], ['ik', 'pron', 'ik', 3849, 4], ['deed', 'verb', 'doe', 3849, 5], ['alles', 'noun', 'alles', 3849, 6], [',', 'punct', ',', 3849, 7], ['van', 'prep', 'van', 3849, 8], ['was', 'noun', 'was', 3849, 9], ['ophangen', 'verb', 'hang op', 3849, 10], ['tot', 'prep', 'tot', 3849, 11], ['de', 'det', 'de', 3849, 12], ['beenhouwerij', 'noun', 'beenhouwerij', 3849, 13], ['kuisen', 'verb', 'kuis', 3849, 14], ['-', 'punct', '-', 3849, 15], ['vroeg', 'verb', 'vraag', 3849, 16], ['ik', 'pron', 'ik', 3849, 17], ['of', 'comp', 'of', 3849, 18], ['ik', 'pron', 'ik', 3849, 19], ['meer', 'det', 'meer', 3849, 20], ['kon', 'verb', 'kan', 3849, 21], ['verdienen', 'verb', 'verdien', 3849, 22], ['.', 'punct', '.', 3849, 23], ['De', 'det', 'de', 3850, 0], ['slager', 'noun', 'slager', 3850, 1], ['wees', 'verb', 'wijs af', 3850, 2], ['dat', 'det', 'dat', 3850, 3], ['verzoek', 'noun', 'verzoek', 3850, 4], ['af', 'part', 'af', 3850, 5], ['.', 'punct', '.', 3850, 6], ['Ik', 'pron', 'ik', 3851, 1], ['pakte', 'verb', 'pak', 3851, 2], ['m'n', 'det', 'mijn', 3851, 3], ['velo', 'noun', 'velo', 3851, 4], ['en', 'vg', 'en', 3851, 5], ['ik', 'pron', 'ik', 3851, 6], ['was', 'verb', 'ben', 3851, 7], ['weg', 'adv', 'weg', 3851, 8], ['.', 'punct', '.', 3851, 9]



2 - Gelemmatiseerde corpora

['Na', 'prep', 'na', 3849, 0], ['een', 'det', 'een', 3849, 1], ['jaar', 'noun', 'jaar', 3849, 2], ['-', 'punct', '-', 3849, 3], ['ik', 'pron', 'ik', 3849, 4], ['deed', 'verb', 'doe', 3849, 5], ['alles', 'noun', 'alles', 3849, 6], [',', 'punct', ',', 3849, 7], ['van', 'prep', 'van', 3849, 8], ['was', 'noun', 'was', 3849, 9], ['ophangen', 'verb', 'hang op', 3849, 10], ['tot', 'prep', 'tot', 3849, 11], ['de', 'det', 'de', 3849, 12], ['beenhouwerij', 'noun', 'beenhouwerij', 3849, 13], ['kuisen', 'verb', 'kuis', 3849, 14], ['-', 'punct', '-', 3849, 15], ['vroeg', 'verb', 'vraag', 3849, 16], ['ik', 'pron', 'ik', 3849, 17], ['of', 'comp', 'of', 3849, 18], ['ik', 'pron', 'ik', 3849, 19], ['meer', 'det', 'meer', 3849, 20], ['kon', 'verb', 'kan', 3849, 21], ['verdienen', 'verb', 'verdien', 3849, 22], ['.', 'punct', '.', 3849, 23], ['De', 'det', 'de', 3850, 0], ['slager', 'noun', 'slager', 3850, 1], ['wees', 'verb', 'wijs af', 3850, 2], ['dat', 'det', 'dat', 3850, 3], ['verzoek', 'noun', 'verzoek', 3850, 4], ['af', 'part', 'af', 3850, 5], ['.', 'punct', '.', 3850, 6], ['Ik', 'pron', 'ik', 3851, 1], ['pakte', 'verb', 'pak', 3851, 2], ['m'n', 'det', 'mijn', 3851, 3], ['**velo**', 'noun', 'velo', 3851, 4], ['en', 'vg', 'en', 3851, 5], ['ik', 'pron', 'ik', 3851, 6], ['was', 'verb', 'ben', 3851, 7], ['weg', 'adv', 'weg', 3851, 8], ['.', 'punct', '.', 3851, 9]



2 - Gelemmatiseerde corpora

Na een jaar – ik deed alles, van was ophangen tot de beenhouwerij kuisen – vroeg ik of ik meer kon verdienen. De slager wees dat verzoek af. Ik pakte m'n **velo** en ik was weg.

3 - Script

- Frequenties
- Concordanties
(welk corpus, welke tekst, (welke positie))

4 – Manuele desambiguatie

Stevaert vindt de sluiting van de kleine **ring** rondom Antwerpen een topprioriteit.

Al draagt een aap een gouden **ring**, het is en blijft een lelijk ding.

Wat is de kick ? Als ik in de **ring** oog in oog sta met mijn tegenstander.

4 – Manuele desambiguatie

Stevaert vindt de sluiting van de kleine **ring** rondom Antwerpen een topprioriteit.

=> ringbaan, ringweg rond een stad

Al draagt een aap een gouden **ring**, het is en blijft een lelijk ding.

=> sieraad + vaste uitdrukking

Wat is de kick ? Als ik in de **ring** oog in oog sta met mijn tegenstander.

=> boksring

Dataset

- Semantisch veld: **verkeer**
- Concepten: **20** (AFRIT, AMBULANCE, AUTO, AUTOBUS, AUTOSNELWEG, BESTELWAGEN, CABRIOLET, FIETS, KRUISPUNT, METRO, MOTOR, RINGWEG, ROTONDE, TREIN, TROTTOIR, VERKEERSKNOOPPUNT, VLUCHTHEUVEL, VLUCHTSTROOK, VRACHTWAGEN, ZEBRAPAD)
- 6 subcorpora:
 - **België – Nederland**
 - **kwaliteitskrant – populaire krant – internetfora (Usenet)**
 - **(1,9 miljard woorden)**

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Onomasiologisch profiel

Concreet voorbeeld

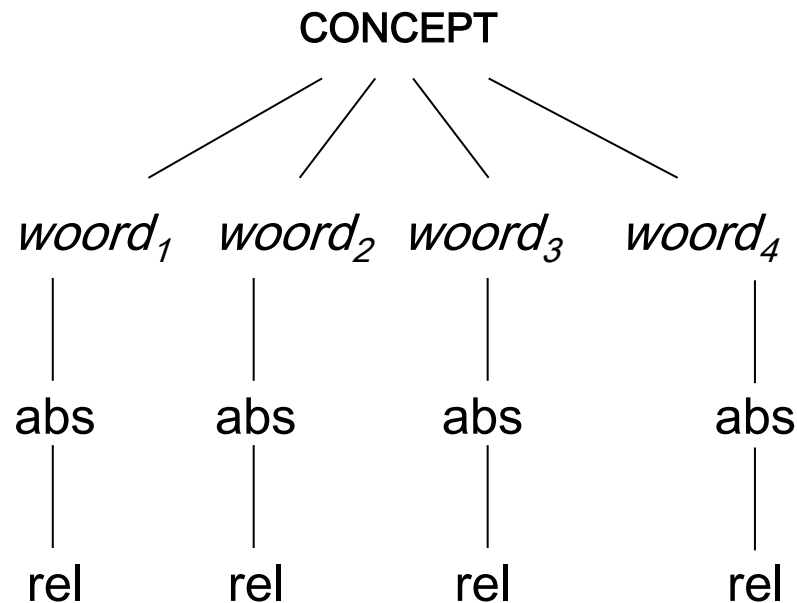


<i>truck</i>	<i>vrachtauto</i>	<i>camion</i>	<i>vrachtwagen</i>
140	630	350	2380
4%	18%	10%	68%

Onomasiologisch profiel

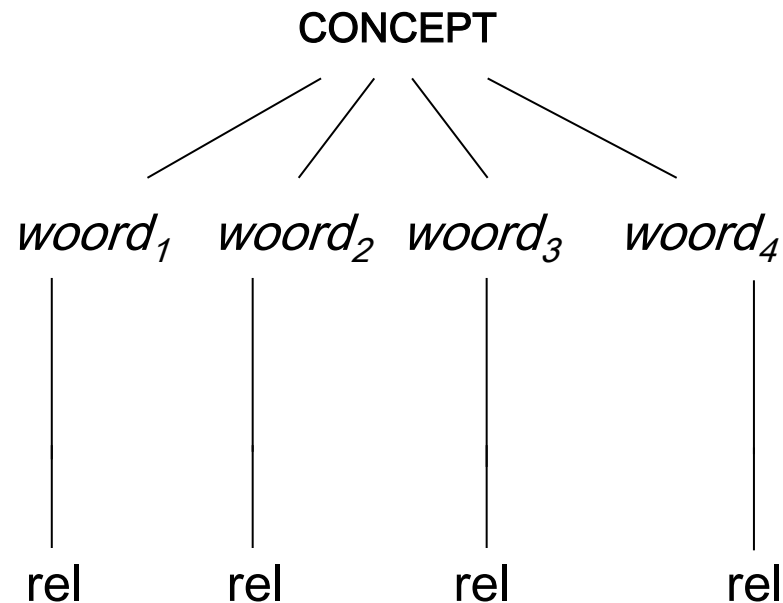
profiel =

geheel van alternatieve lexicalisering met frequentiedistributie



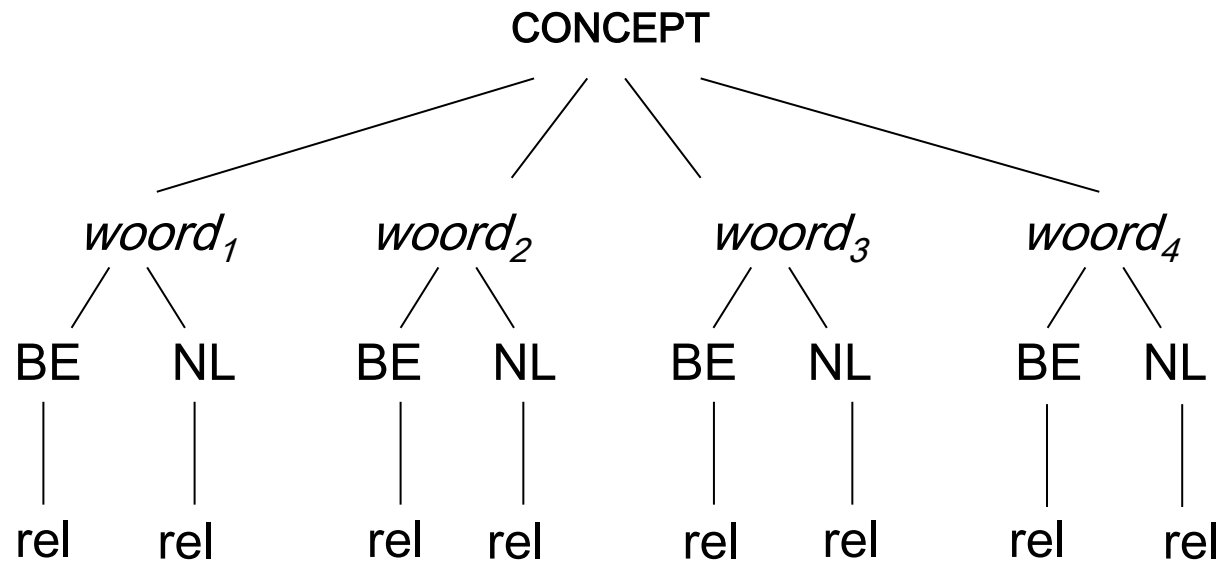
Onomasiologisch profiel

profiel =
geheel van alternatieve lexicalisering met frequentiedistributie



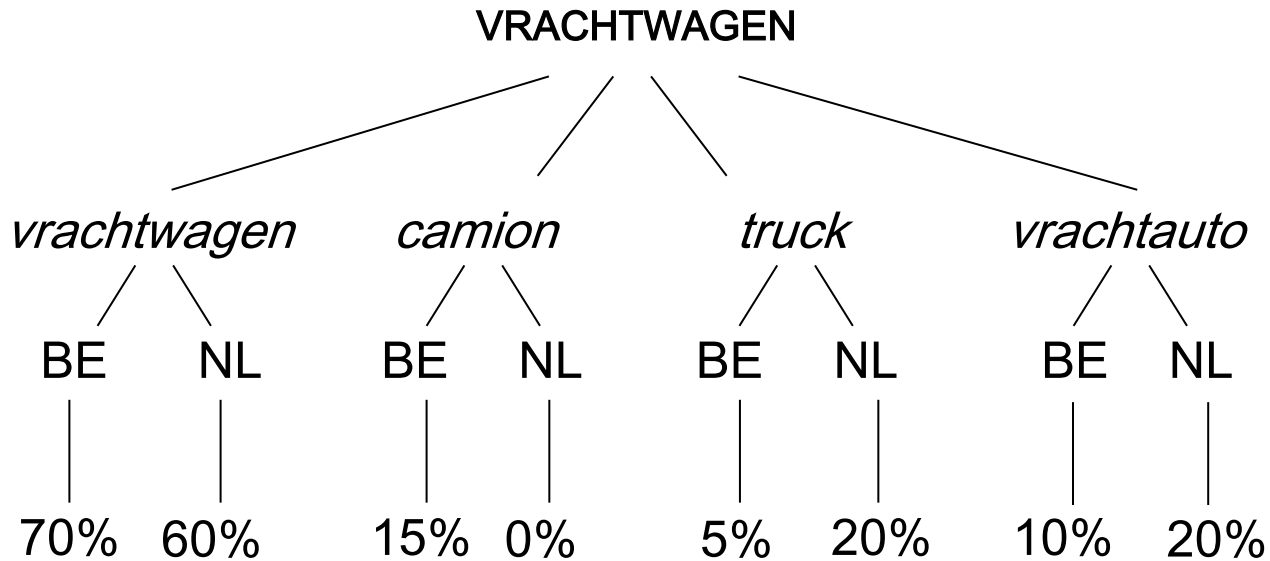
Onomasiologisch profiel

profiel voor beide regio's



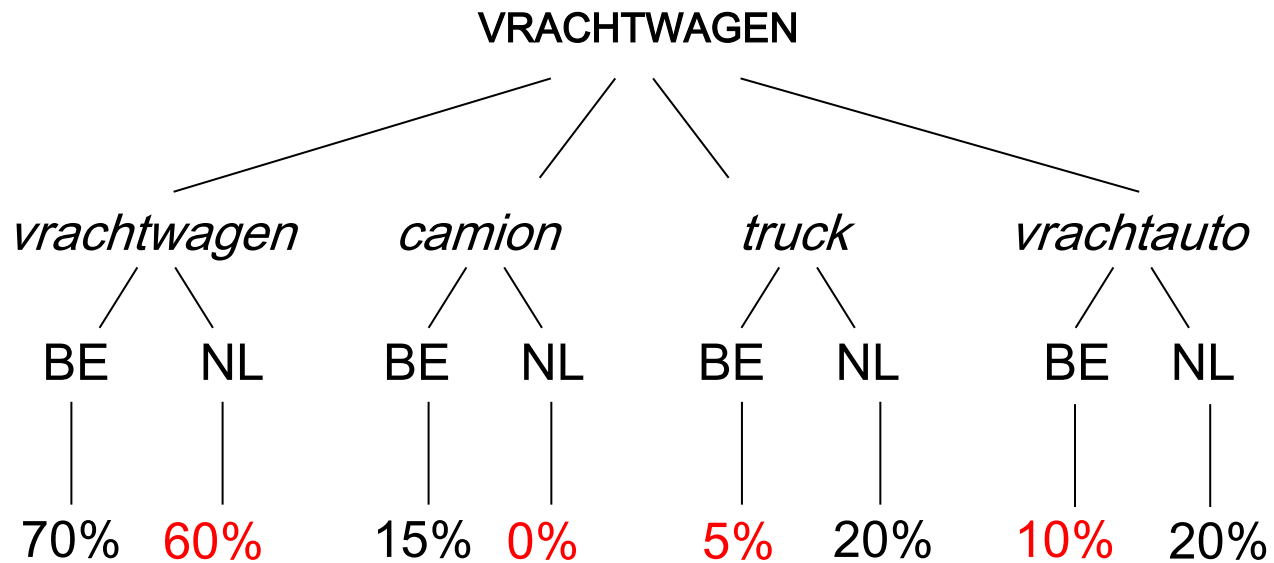
Onomasiologisch profiel

concreet profiel voor BE & NL voor VRACHTWAGEN



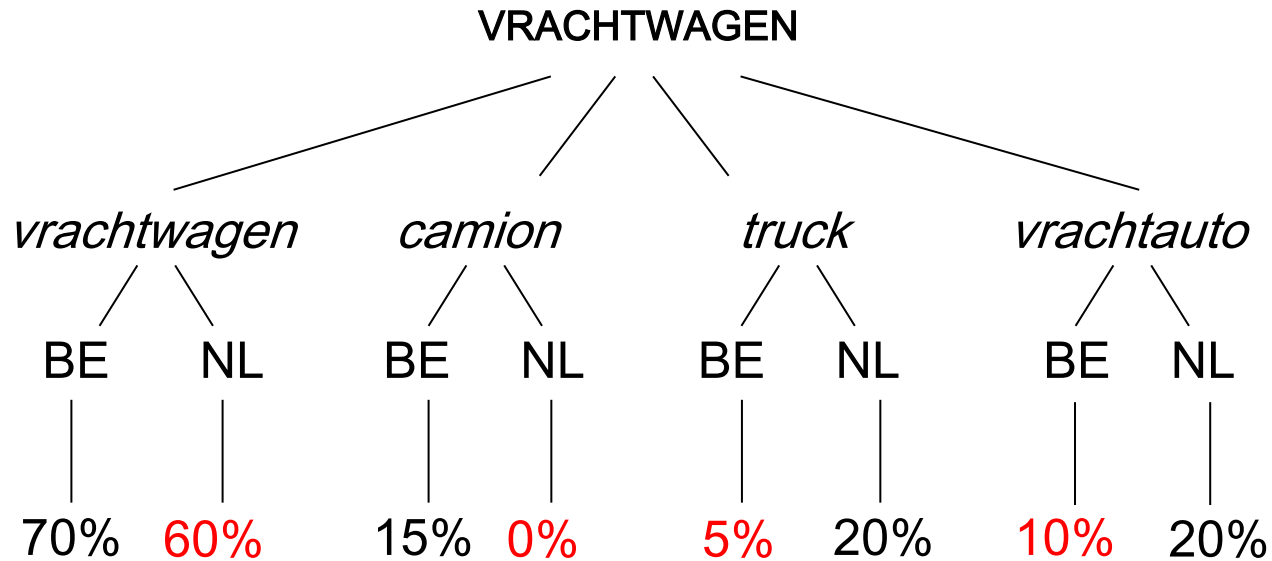
Onomasiologisch profiel

overlapping = de minima



Onomasiologisch profiel

uniformiteit = gehele overlap = som van de minima



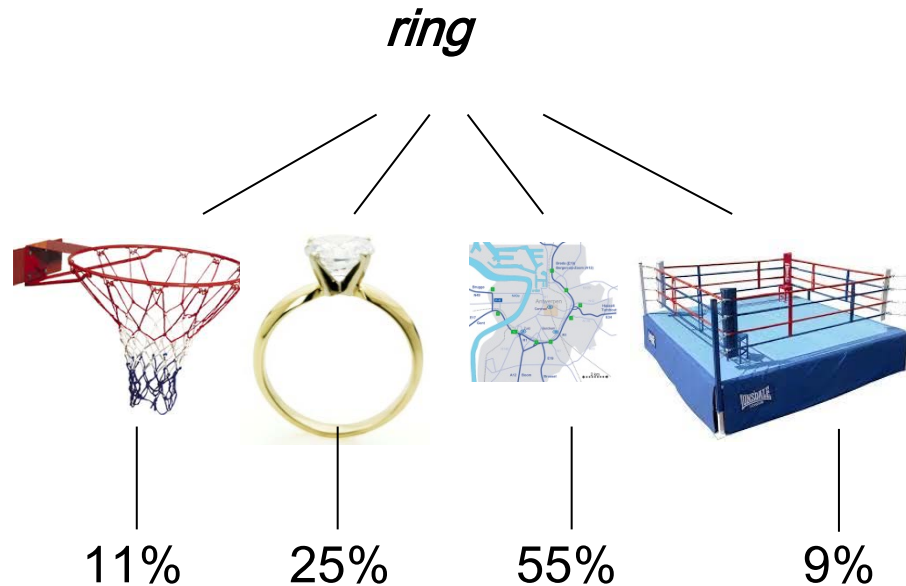
$$\rightarrow 60\% + 0\% + 5\% + 10\% = 75\%$$

Semasiologische variatie

Semasiologische variatie: 1 woord → verschillende betekenissen

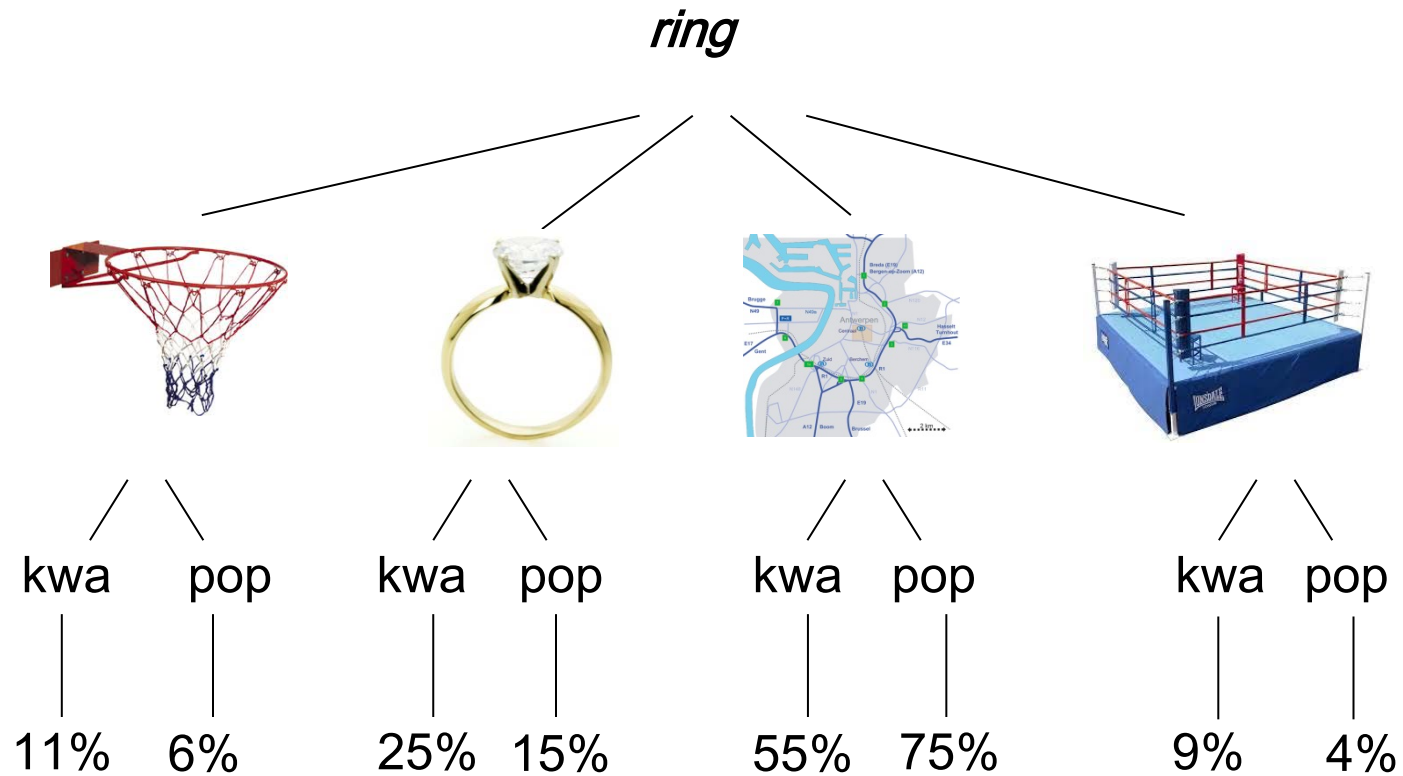
Wat gaan we meten?

- Binnen de context van het semantische **veld verkeer**, wanneer heeft een **woord** een **verkeersbetekenis**?



Semasiologisch profiel

concreet profiel voor *ring* voor kwaliteits- en populaire kranten



Semasiologisch profiel

enkel verkeersbetekenis

ring



kwa 11%
pop 6%



kwa 25%
pop 15%



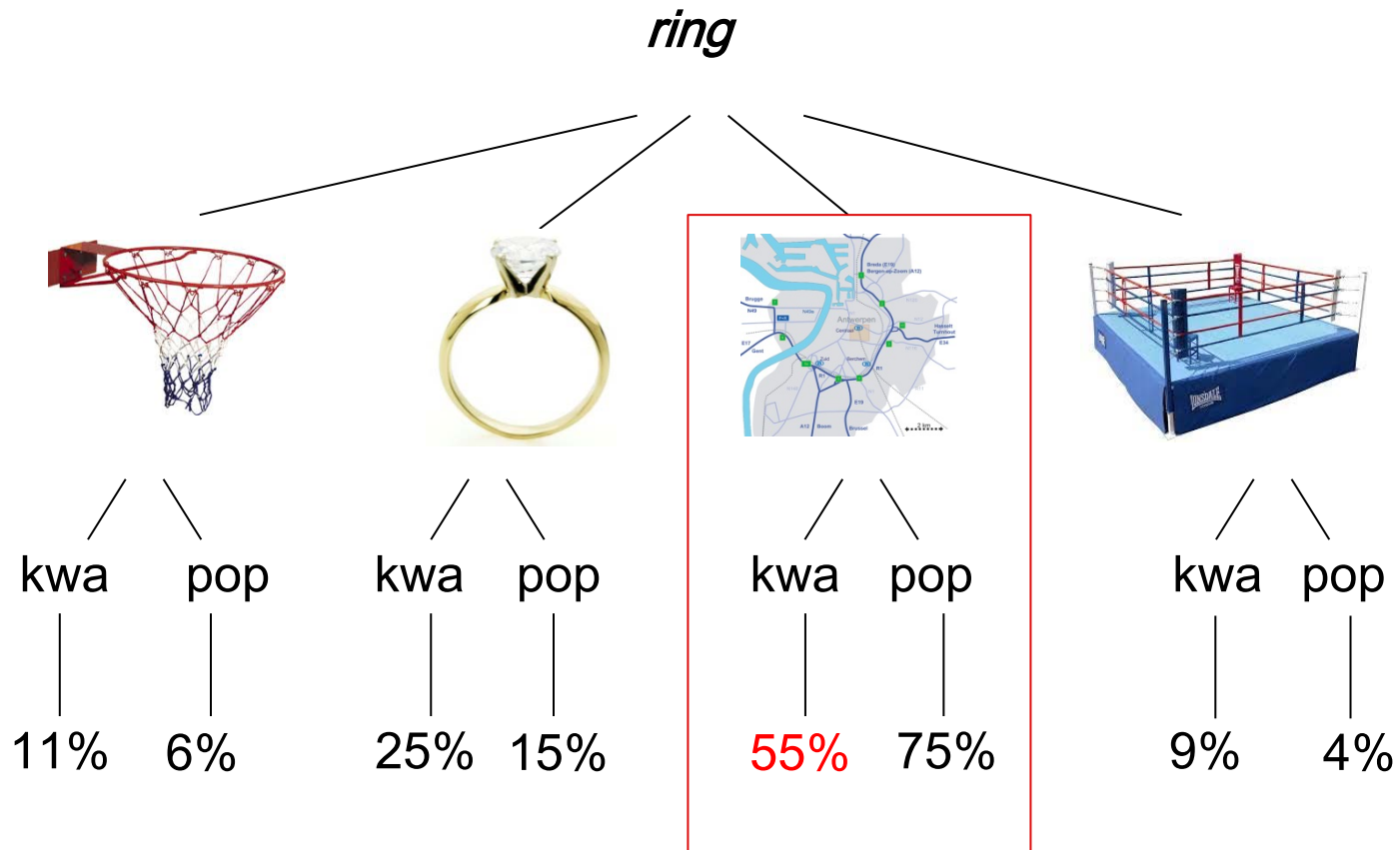
kwa 55%
pop 75%



kwa 9%
pop 4%

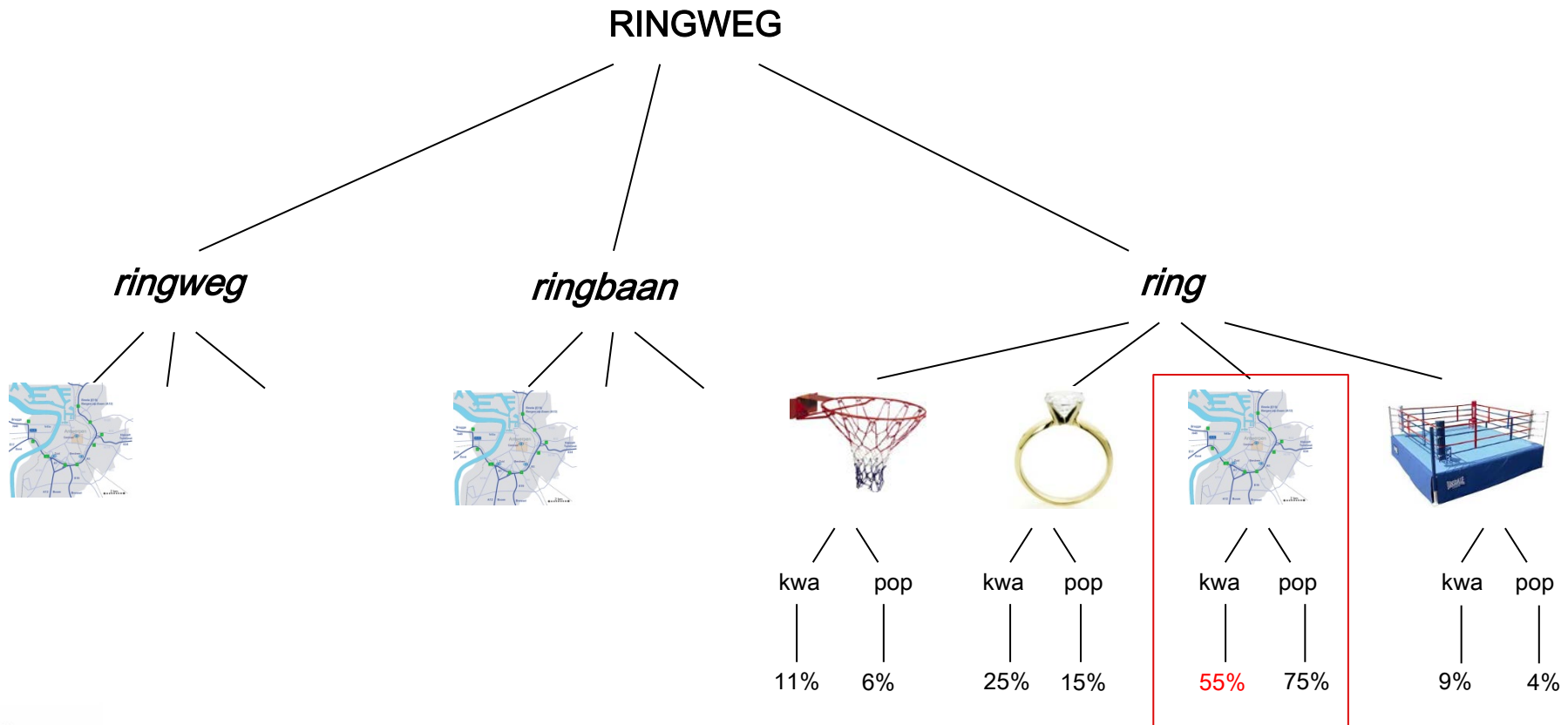
Semasiologisch profiel

overlapping = het minimum



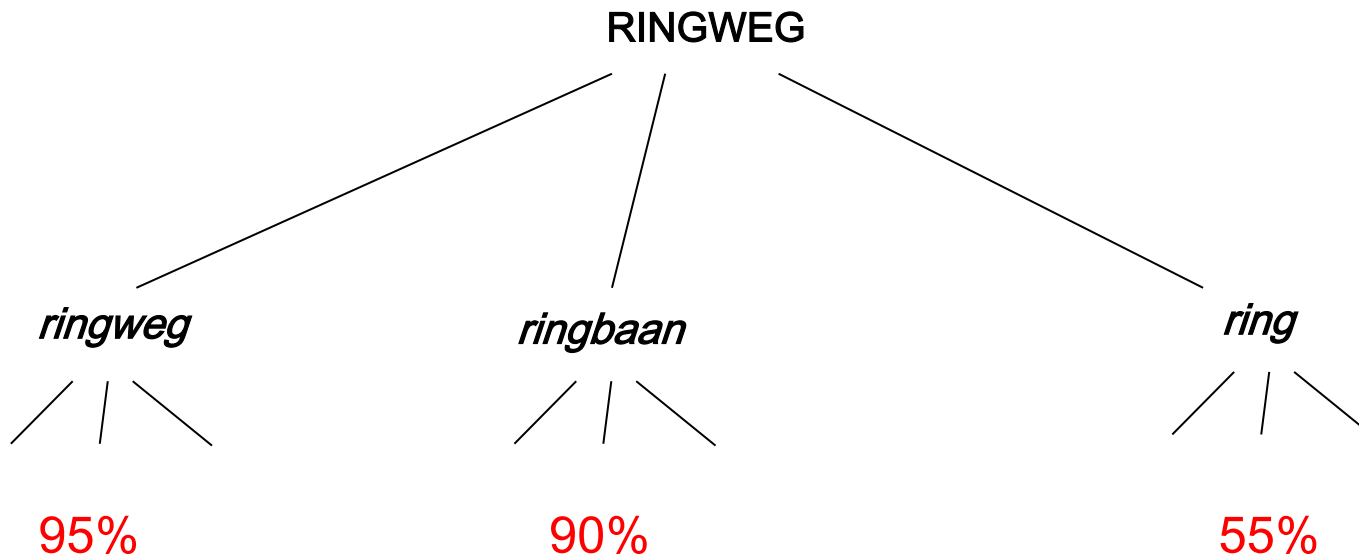
Semasiologisch profiel

dubbele link met onomasiologische profiel



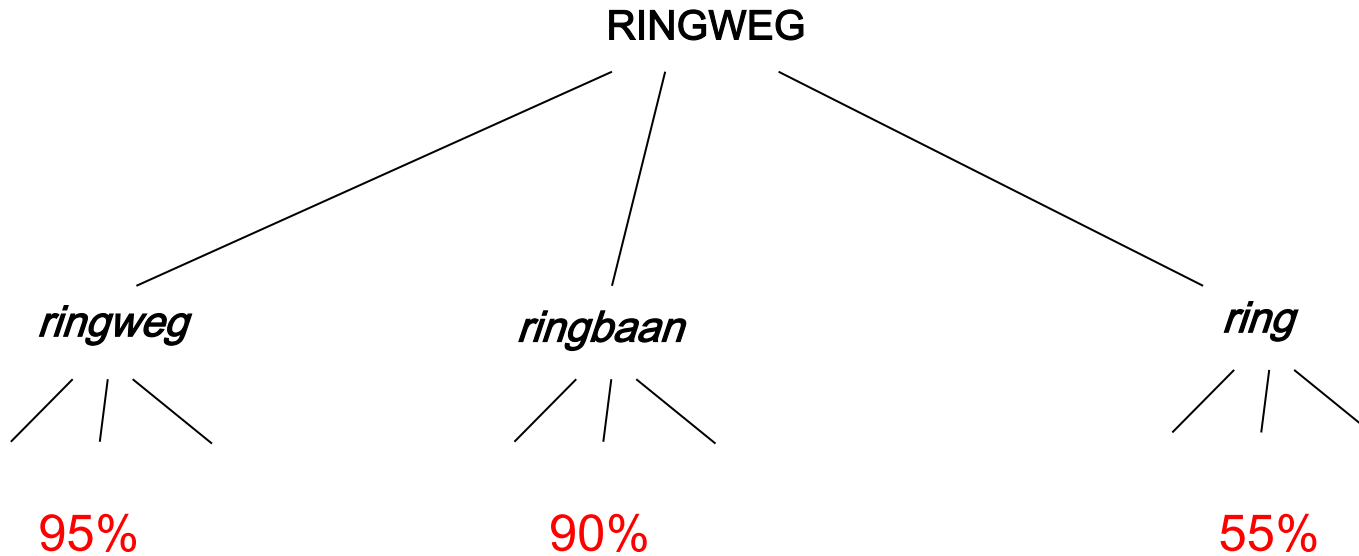
Semasiologisch profiel

uniformiteit per woord



Semasiologisch profiel

uniformiteit per profiel



$$\rightarrow \frac{95\% + 90\% + 55\%}{3} = 80\%$$

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Resultaten

Hoe interpreteren wat we vinden?

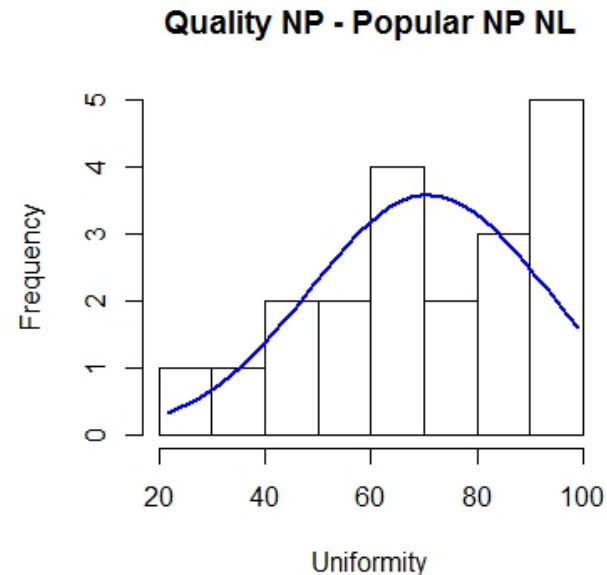
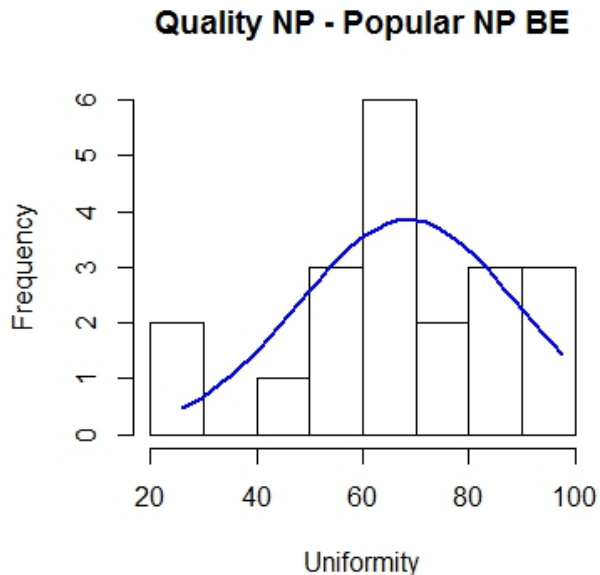
- In welke mate betekenen woorden [(in een onomasiologisch profiel) binnen het verkeersveld] hetzelfde in twee corpora?
 - → onderling begrip
- **!! caveat**
 - **Enkel voor de verkeersbetekenis**

Resultaten

Belgisch Nederlands vs. Nederlands Nederlands

$U(\text{quanp-be}, \text{popnp-be}) = 71\%$

$U(\text{quanp-nl}, \text{popnp-nl}) = 71\%$



Resultaten

Belgisch Nederlands vs. Nederlands Nederlands

$U(\text{quanp-be}, \text{popnp-be}) = 71\%$

$U(\text{quanp-nl}, \text{popnp-nl}) = 71\%$

$U(\text{quanp-be}, \text{usenet-be}) = 66\%$

$U(\text{quanp-nl}, \text{usenet-nl}) = 66\%$

→ Woorden hebben in dezelfde mate dezelfde betekenis in Belgisch Nederlands als in Nederlands Nederlands

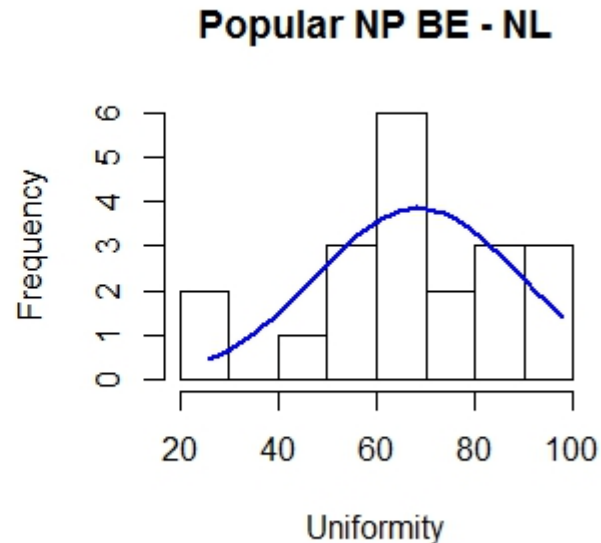
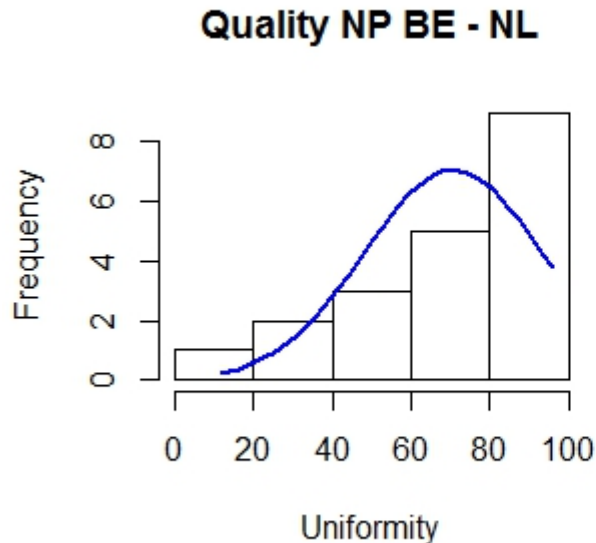
→ De afstand tussen kwaliteitskranten en populaire kranten is kleiner dan tussen kwaliteitskranten en internetfora

Resultaten

Register: kwaliteitskranten – populaire kranten – internetfora

$U(\text{quanp-be}, \text{quanp-nl}) = 70\%$

$U(\text{popnp-be}, \text{popnp-nl}) = 68\%$



Resultaten

Register: kwaliteitskranten – populaire kranten – internetfora

$U(\text{quanp-be}, \text{quanp-nl}) = 70\%$

$U(\text{popnp-be}, \text{popnp-nl}) = 68\%$

$U(\text{usenet-be}, \text{usenet-nl}) = 64\%$

→ De afstand tussen kwaliteitskranten is kleiner dan (tussen populaire kranten is kleiner dan) tussen internetfora

Resultaten

Invloed semantisch veld: verkeer – emoties – ICT

Verkeer $U(\text{quanp-be}, \text{quanp-nl}) = 70\%$

ICT $U(\text{quanp-be}, \text{quanp-nl}) = 67\%$

EmotiesZN $U(\text{quanp-be}, \text{quanp-nl}) = 86\%$

EmotiesBN $U(\text{quanp-be}, \text{quanp-nl}) = 73\%$

→ Semantisch veld speelt een belangrijke rol

?concrete vs. abstracte concepten, emotionaliteit, ...

Overzicht

Achtergrond

Data

Methodologie

Resultaten

Conclusies

Conclusies

Geen verschil Belgisch Nederlands en Nederlands Nederlands

Wel verschil register: kranten vs. internet

Invloed semantisch veld

? Methodologie

? Invloed type concept (concreet/abstract, expressiviteit, frequent, ...)

? Invloed polysemie van het woord

Bedankt! Vragen? Suggesties? Inzichten?



Meer informatie:

<http://www.ling.arts.kuleuven.be/qlv>

Jocelyne.Daems@arts.kuleuven.be



Referenties

Geeraerts, Dirk, Stefan Grondelaers & Dirk Spielman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. P.J. Meertens-Instituut.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter.