

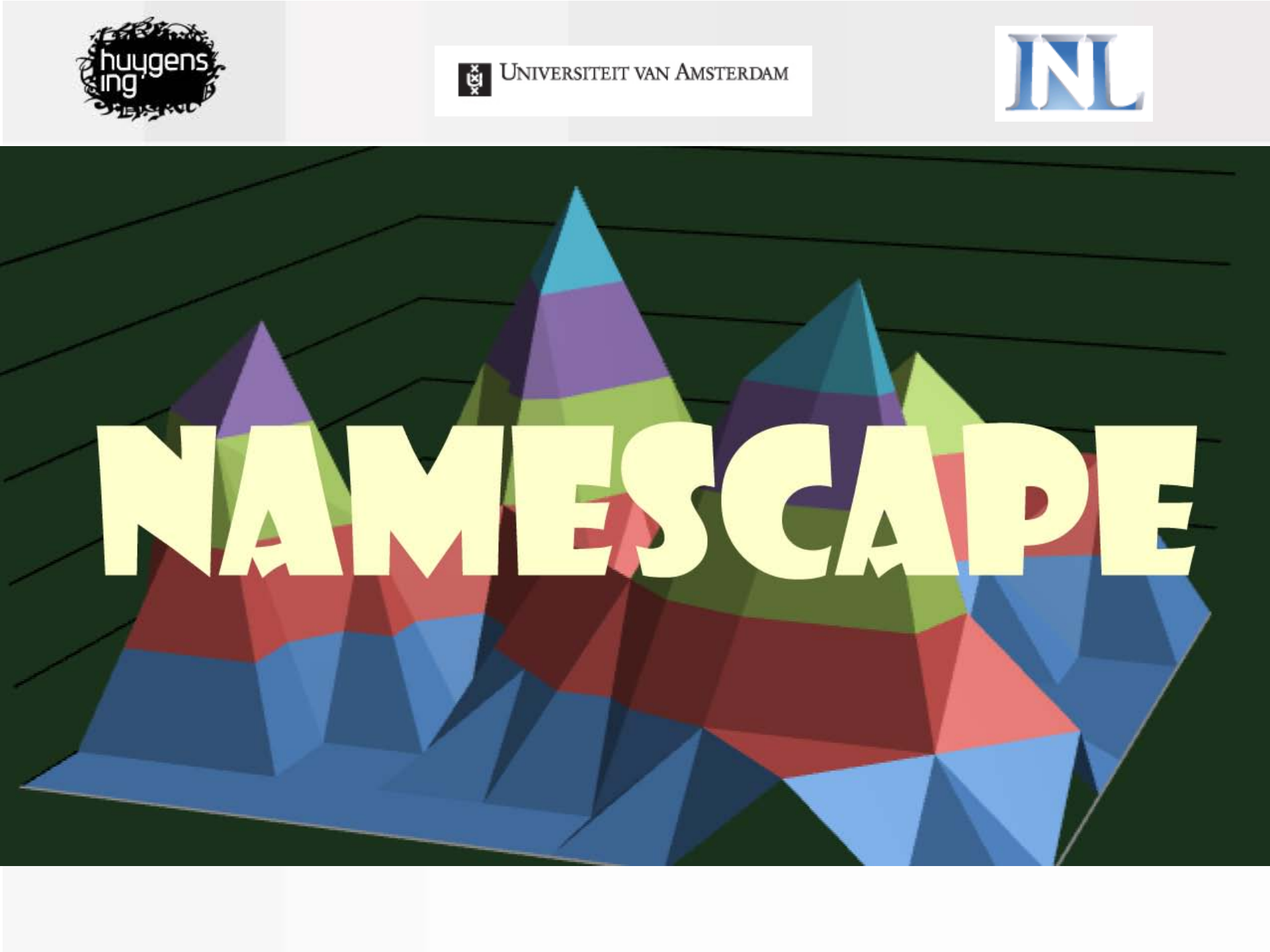


Namescape

Karina van Dalen-Oskam

6 maart 2012

NAMESCAPE

The background of the slide is a dark green, textured surface with a series of horizontal lines. In the foreground, there is a complex, abstract geometric landscape composed of many colorful, faceted shapes, including triangles and polygons in shades of blue, red, purple, green, and yellow. The word "NAMESCAPE" is written in large, bold, yellow capital letters across the center of the image.

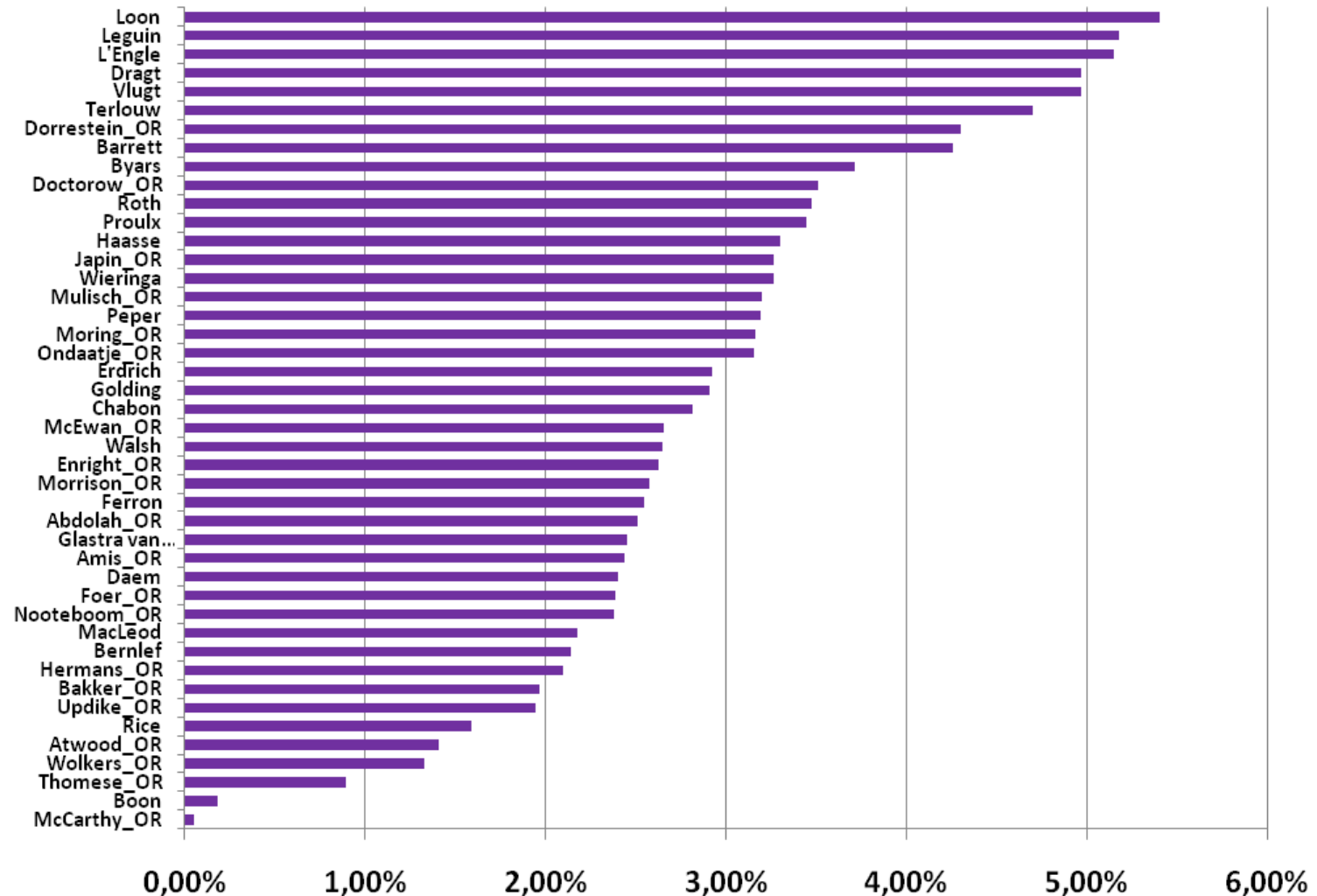
Letterkundig stilistisch onderzoek

Wat is de rol van eigennamen in romans?

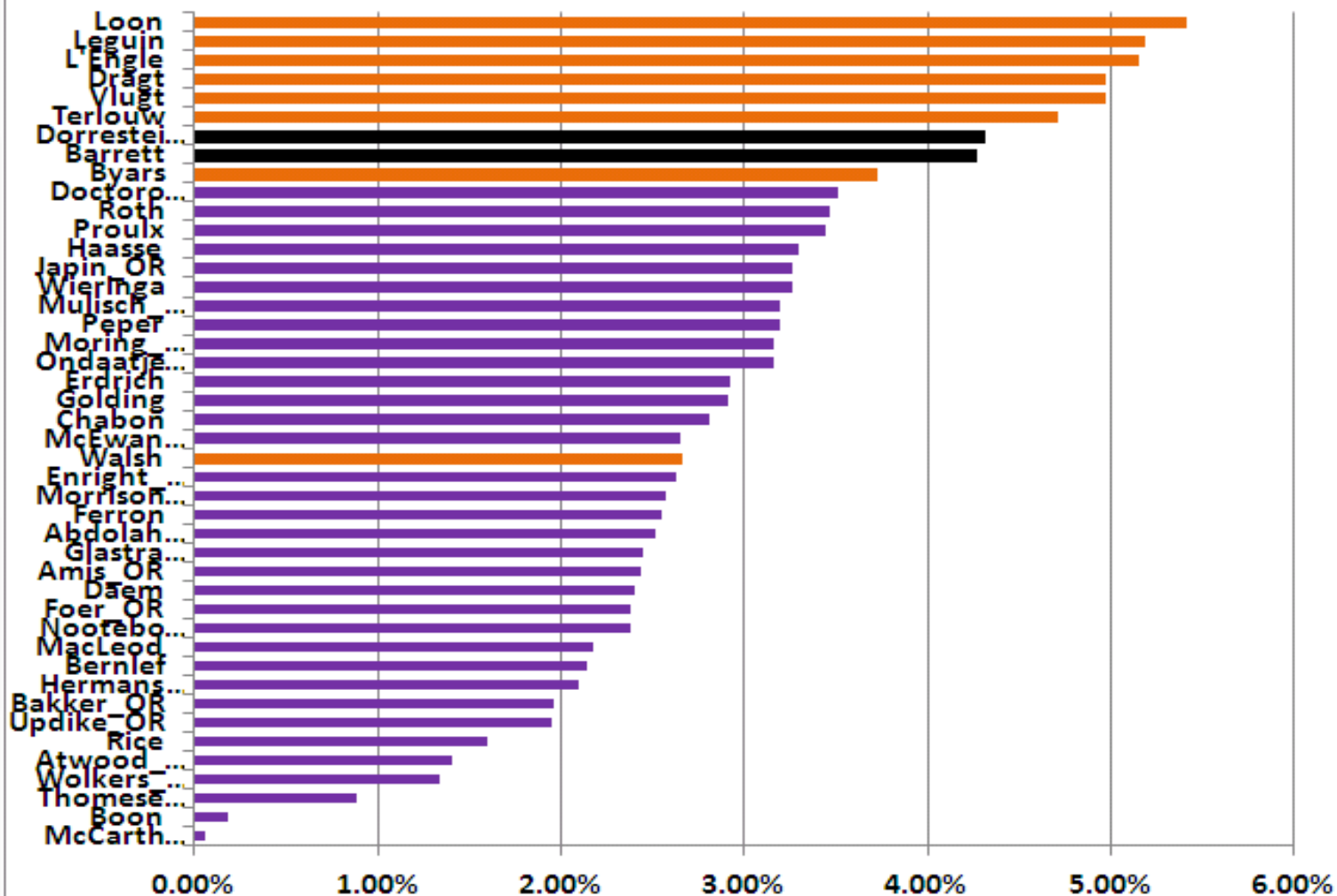
Hoe vaak komen ze voor en wat zijn hun functies?



name tokens % of total tokens



name tokens % of total tokens



Naamtypen

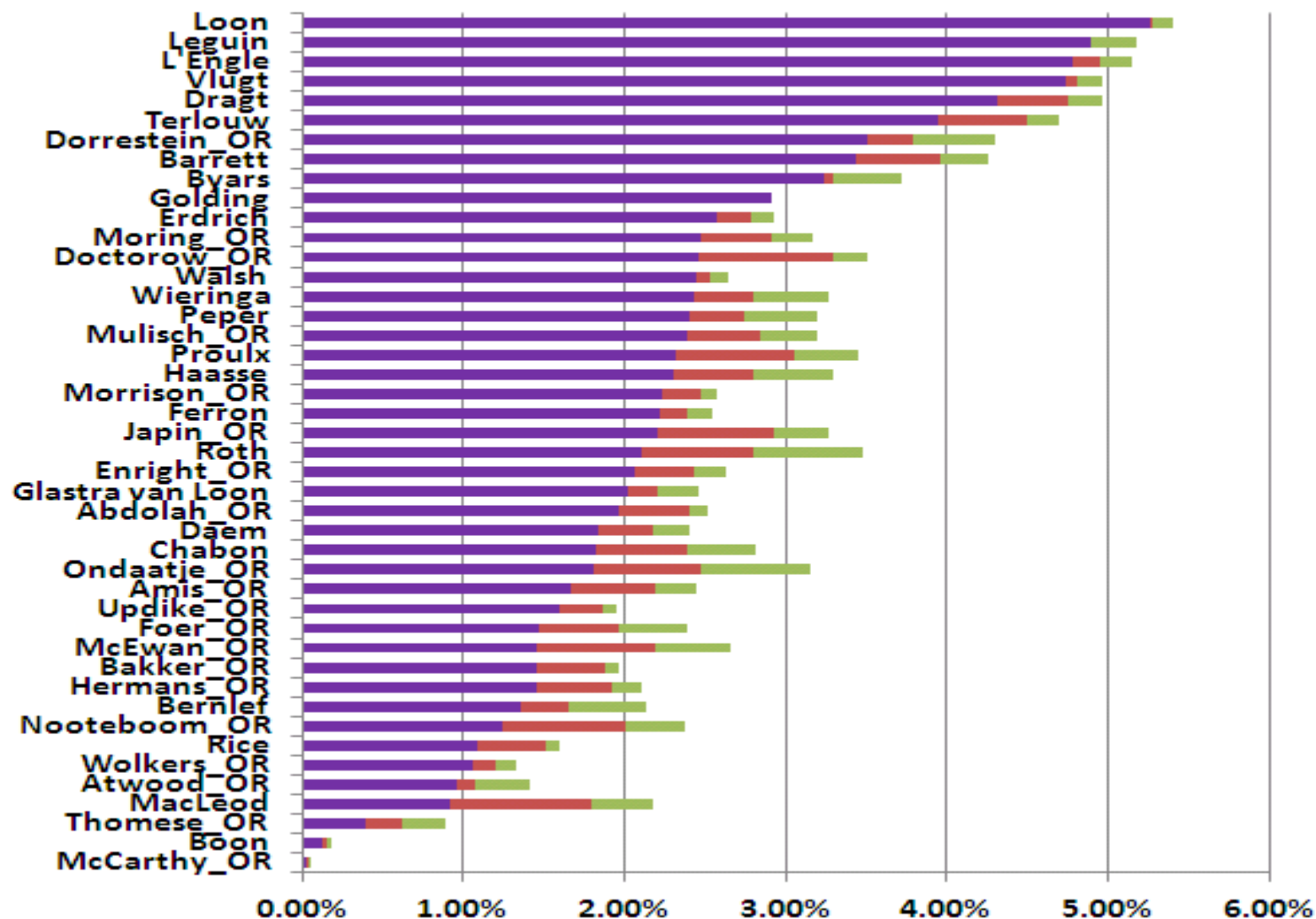
Persoonsnamen

Geografische namen

Overige namen (gebouwen, gebeurtenissen,
instituten, boektitels, etc.)



% PERS, % GEO, % OTH



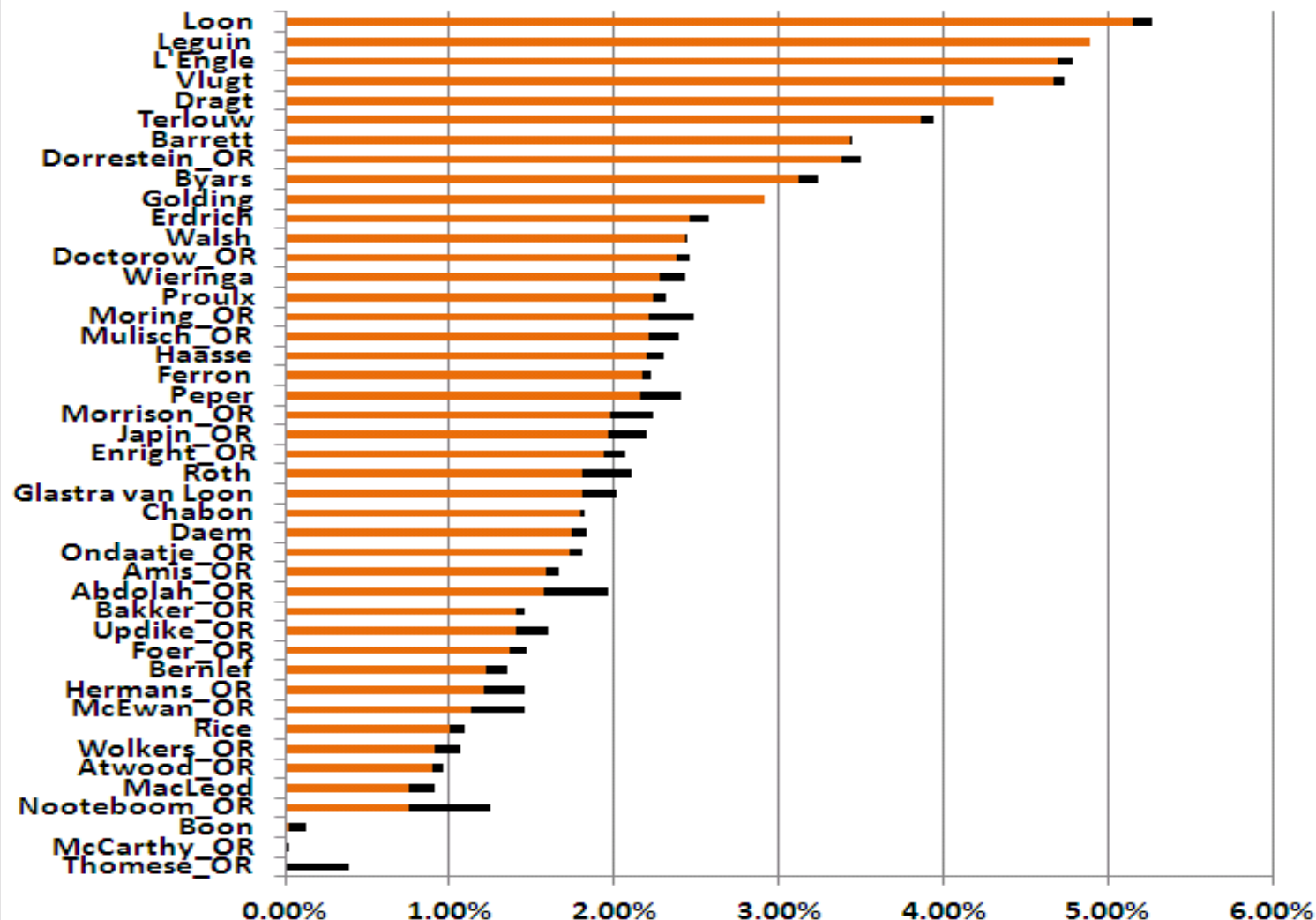
Plot intern versus plot extern

Plot intern: namen die verwijzen naar personages, plaatsen of zaken die uitsluitend 'bestaan' binnen de fictie van de roman.

Plot externe namen verwijzen naar personen, plaatsen of zaken die in de werkelijke wereld bestaan.



% INT PERS, % EXT PERS



Subtypen van persoonsnamen

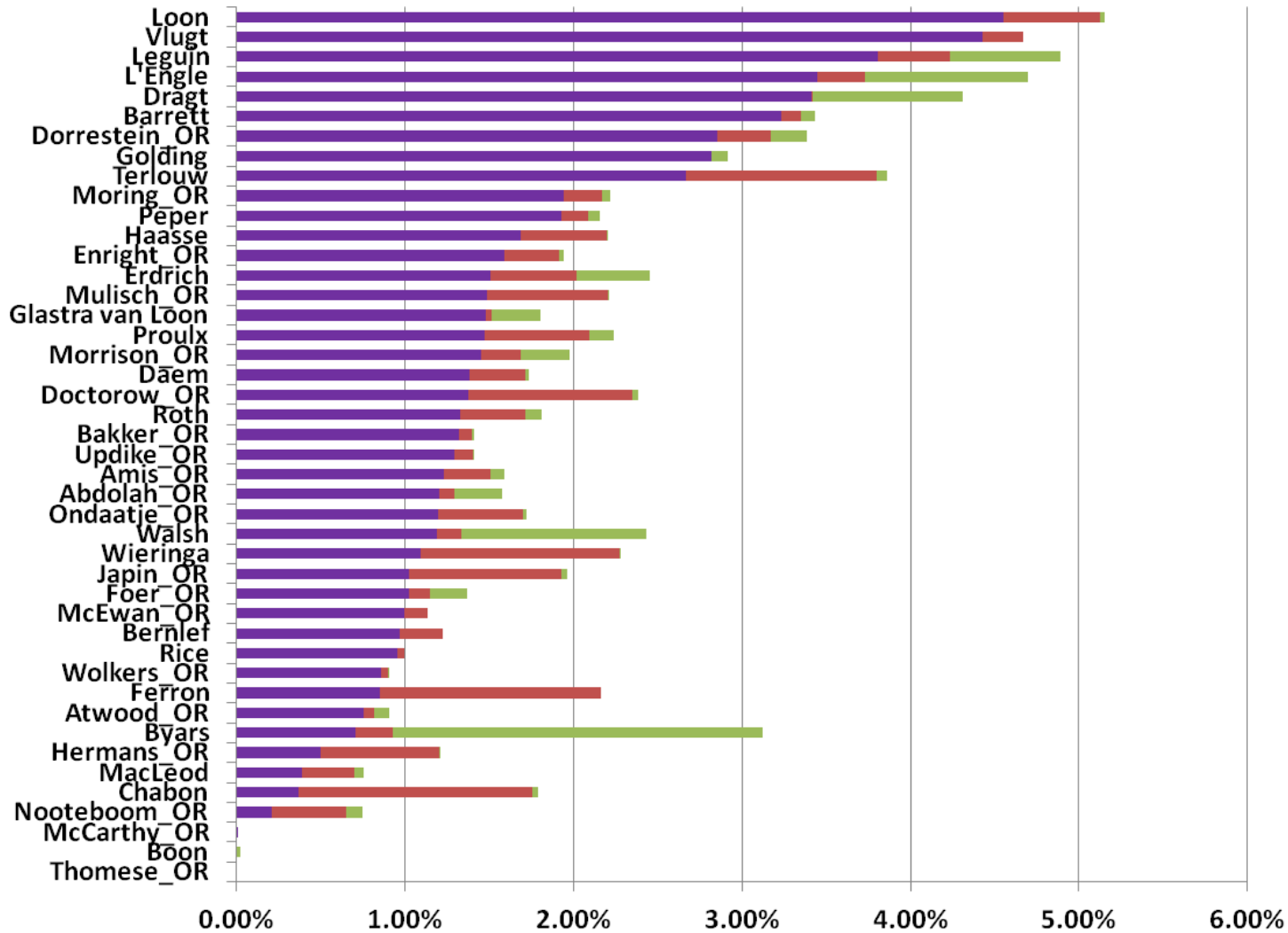
Voornamen

Familienamen

Bijnamen



% FIRST, % FAM, % BY

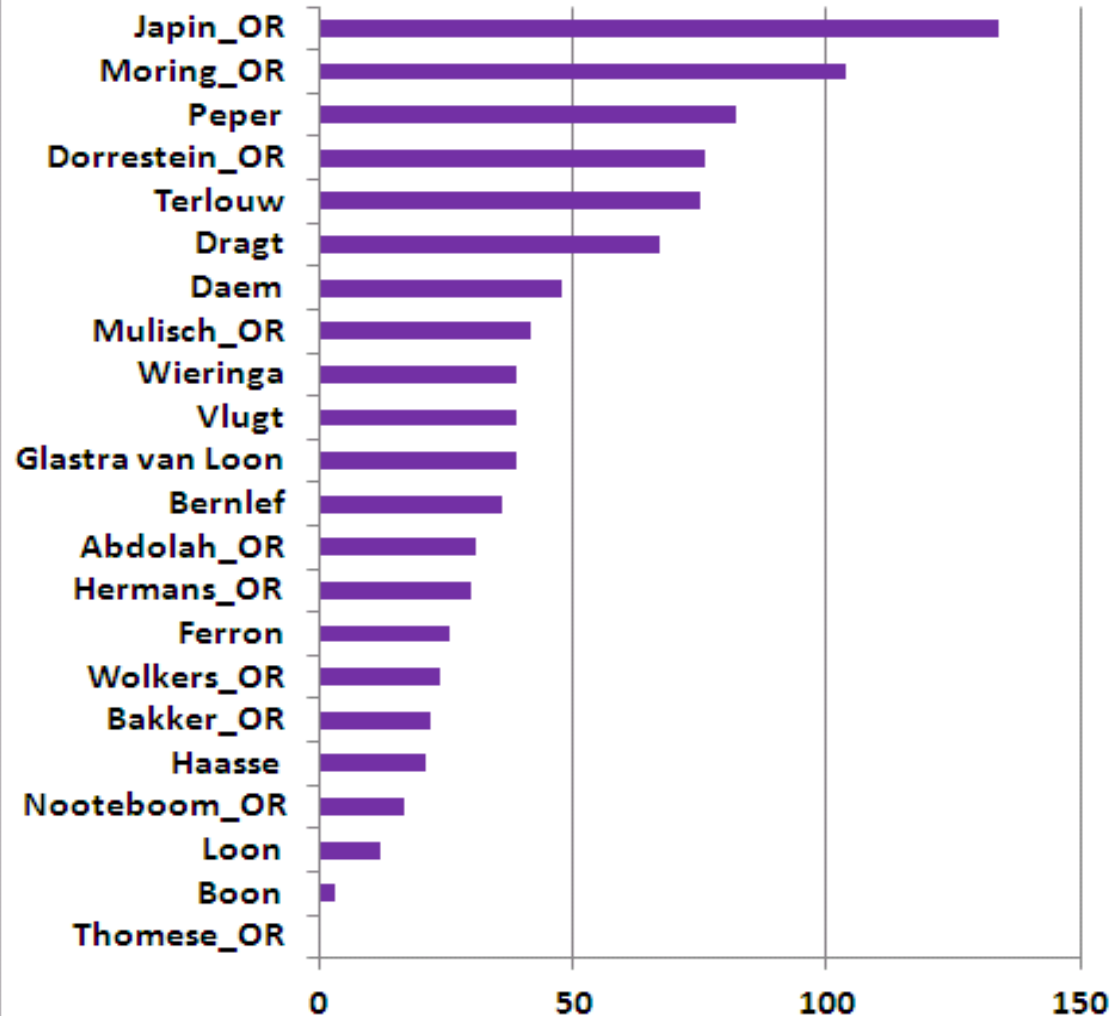


Aantal personen met een naam

Onafhankelijk van naamvorm of naamtype
(voornaam, familienaam en bijnaam zijn
verschillende naamtypen)



Named entities (Dutch corpus)



Namescape werkzaamheden

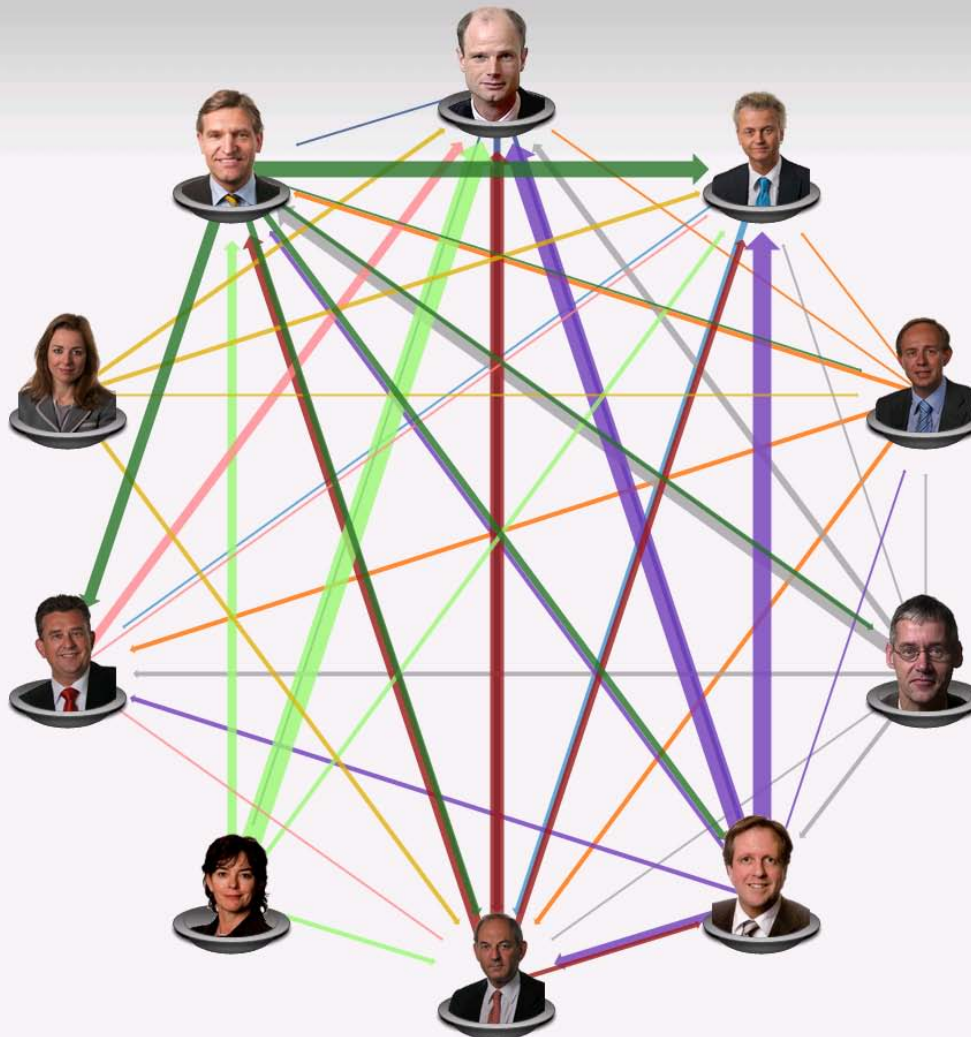
- aanpassing Stanford NER INL om de eerder genoemde tagset te kunnen gebruiken
- met gebruiksvriendelijk online annotatie- en correctie-mogelijkheid
- aanpassing techniek ILLC voor Named Entity Resolution (mapping naar 'real entities') gebruikmakend van Wikipedia
- aanbrengen metadata romans (auteur, titel, uitgever, jaar, aantal pagina's, ISBN) CLARIN compliant (CMDI)

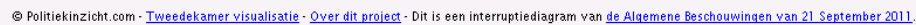
Namescape resultaten (1)

- lexicon van namen in context uit 582 Nederlandstalige romans
- online demonstrator met complexe zoekmogelijkheden
- barcode browsing
- weergave van de entities in een roman in een netwerkvisualisatie en met clusteranalyse

Zoek naam in /bb-open/EwoudSanders  zoek

1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND	1991 SAND
3 hits	6 hits	2 hits	0 hits	7 hits	14 hits	2 hits	13 hits	18 hits	19 hits	0 hits	2 hits	39 hits	26 hits



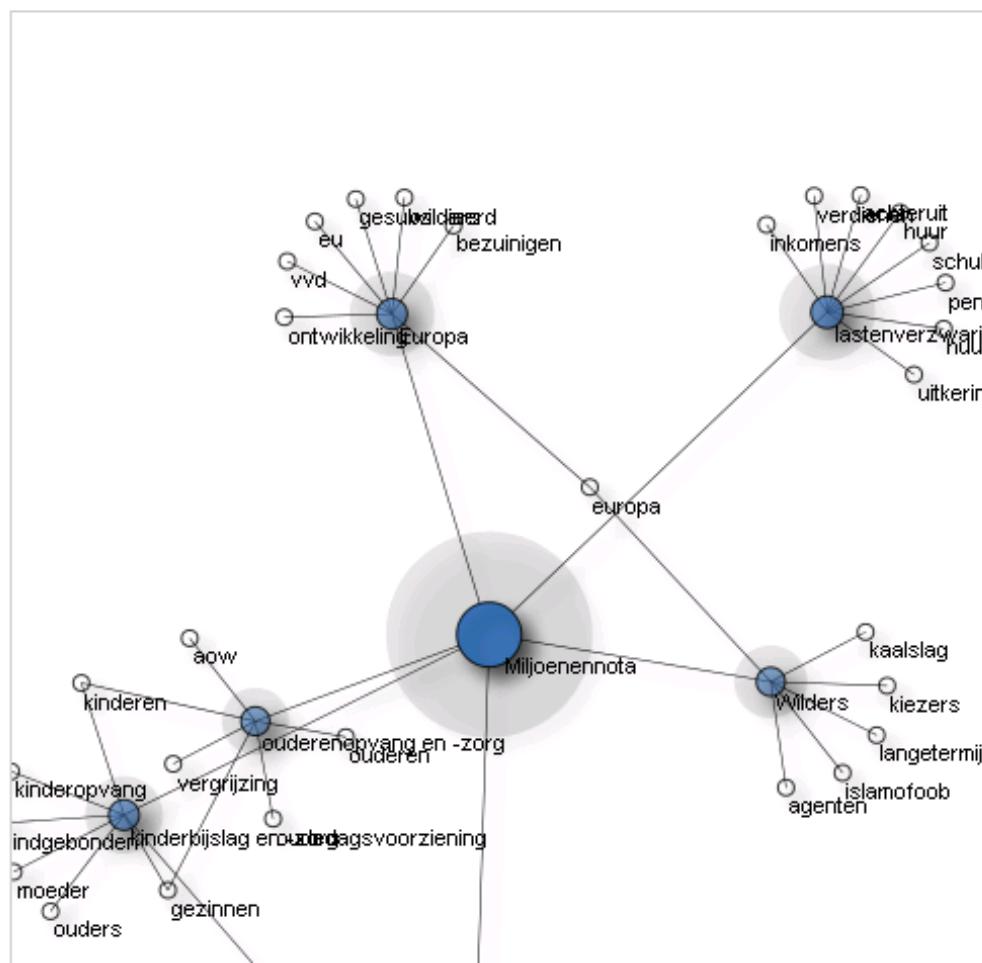


Prinsjesdagwoordenboekje

Donderdag voor Prinsjesdag lekte de miljoenennota uit. Iedereen kon de tekst al lezen ver voordat er in de Tweede Kamer over gesproken werd. Waar gingen de reacties over? Welke onderwerpen in de miljoenennota hielden de mensen bezig?

Het INL verzamelde berichten van nieuwssites en de reacties daarop. Met behulp van moderne statistische methoden werden de reacties gegroepeerd naar onderwerp.

Hieronder zie je een visualisatie van de verschillende onderwerpen en hun onderlinge relaties.



Namescape resultaten (2)

- lexicon en tools beschikbaar in CLARIN compliant vorm
- enhanced publication waarin getoond wordt wat er met de data en tools kan worden gedaan
- nut voor letterkundige onderzoeker: grotere corpora sneller te taggen, en door de tools is het corpus te exploreren op allerlei manieren, wat tot vele nieuwe inzichten kan leiden over gebruik en functie van eigennamen in romans
 - en wellicht ook in andere tekstgenres



karina.van.dalen@huygens.knaw.nl