

# Een combinatie van Distributionele Semantiek & Linked Open Data voor detectie van veranderende concepten

---

Laura Hollink, **Astrid van Aggelen**, Jacco van Ossenbruggen  
Information Access Group



VIII sec. - 1802 - 1820 - 1835 - 1848 - 1860 - 1868 - 1872 - 1876 - 1880 - 1885 - 1890 - 1894 - 1898 - 1903 - 1906 - 1909 - 1912 - 1914 - 1916 - 1919 - 1923 - 1927 - 1930

# Veranderende concepten

---

Hoeveel is het concept

Kei  
Milieu  
Holocaust  
Propaganda  
Europese Unie  
Democratie  
...

veranderd

in de afgelopen

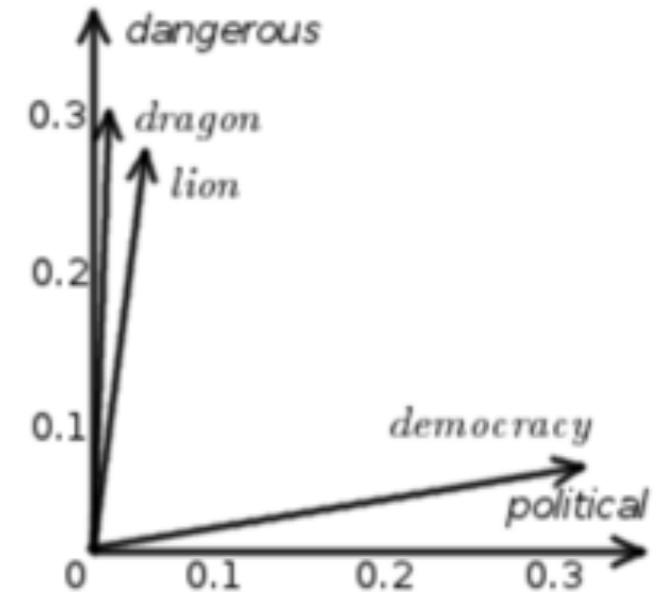
jaren  
decennia  
eeuwen  
...

?

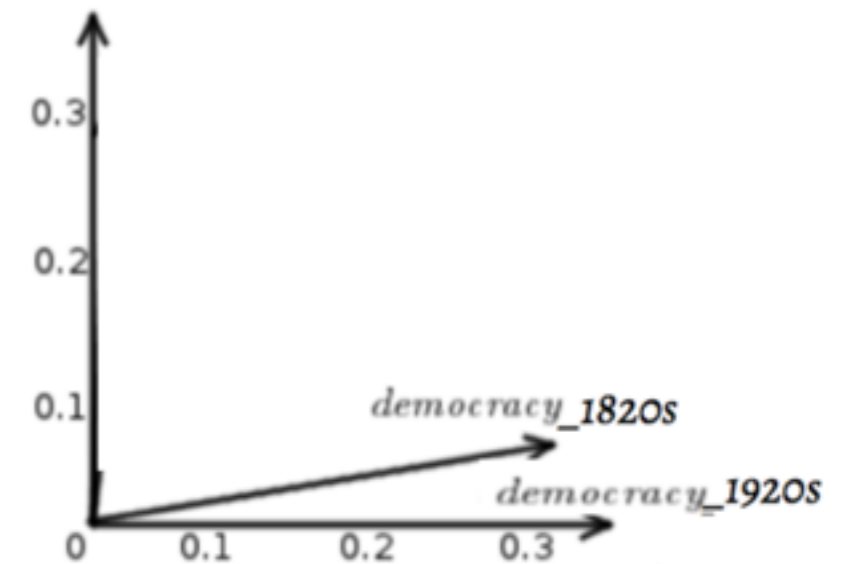
# Distributionele semantiek werkt goed



	<i>I</i>	<i>like</i>	<i>enjoy</i>	<i>deep</i>	<i>learning</i>	<i>NLP</i>	<i>flying</i>	<i>.</i>
<i>I</i>	0	2	1	0	0	0	0	0
<i>like</i>	2	0	0	1	0	1	0	0
<i>enjoy</i>	1	0	0	0	0	0	1	0
<i>deep</i>	0	1	0	0	1	0	0	0
<i>learning</i>	0	0	0	1	0	0	0	1
<i>NLP</i>	0	1	0	0	0	0	0	1
<i>flying</i>	0	0	1	0	0	0	0	1
<i>.</i>	0	0	0	0	1	1	1	0



	<i>I</i>	<i>like</i>	<i>enjoy</i>	<i>deep</i>
<i>I</i>	0	2	1	0
<i>like</i>	2	0	0	1
<i>enjoy</i>	1	0	0	0
<i>deep</i>	0	1	0	0
<i>learning</i>	0	0	0	1
<i>NLP</i>	0	1	0	0
<i>flying</i>	0	0	1	0
<i>.</i>	0	0	0	0



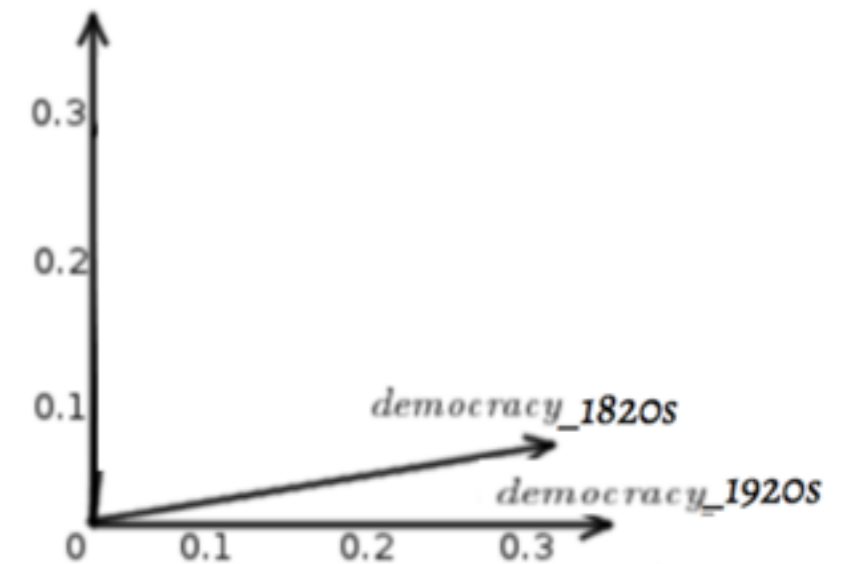
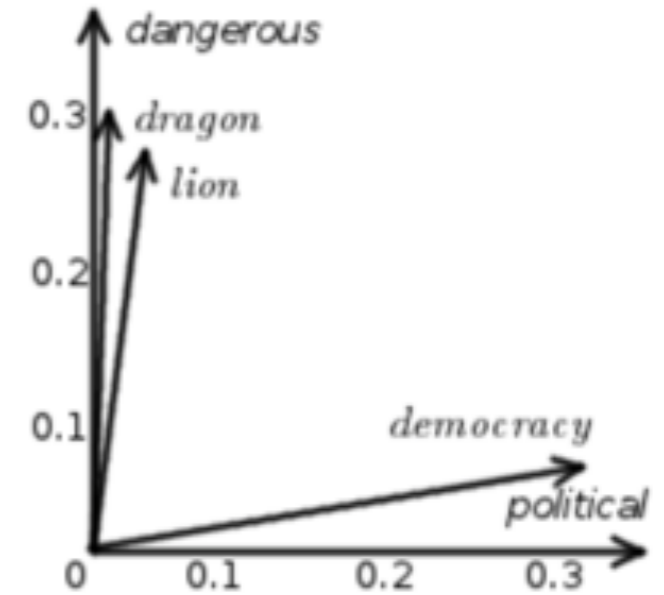
# Distributionele semantiek werkt goed



	<i>I</i>	<i>like</i>	<i>enjoy</i>	<i>deep</i>	<i>learning</i>	<i>NLP</i>	<i>flying</i>	.
<i>I</i>	0	2	1	0	0	0	0	0
<i>like</i>	2	0	0	1	0	1	0	0
<i>enjoy</i>	1	0	0	0	0	0	1	0
<i>deep</i>	0	1	0	0	1	0	0	0
<i>learning</i>	0	0	0	1	0	0	0	1
<i>NLP</i>	0	1	0	0	0	0	0	1
<i>flying</i>	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

	<i>I</i>	<i>like</i>	<i>enjoy</i>	<i>deep</i>
<i>I</i>	0	2	1	0
<i>like</i>	2	0	0	1
<i>enjoy</i>	1	0	0	0
<i>deep</i>	0	1	0	0
<i>learning</i>	0	0	0	1
<i>NLP</i>	0	1	0	0
<i>flying</i>	0	0	1	0
.	0	0	0	0



Evaluated e.g. in Frermann & Lapata. *A Bayesian Model of Diachronic Meaning Change*.

examples by Aurelie Herbelot,  
<http://aurelieherbelot.net/research/distributional-semantics-intro/>  
 matrices from [https://cs224d.stanford.edu/lecture\\_notes/notes1.pdf](https://cs224d.stanford.edu/lecture_notes/notes1.pdf)

Maar...

---

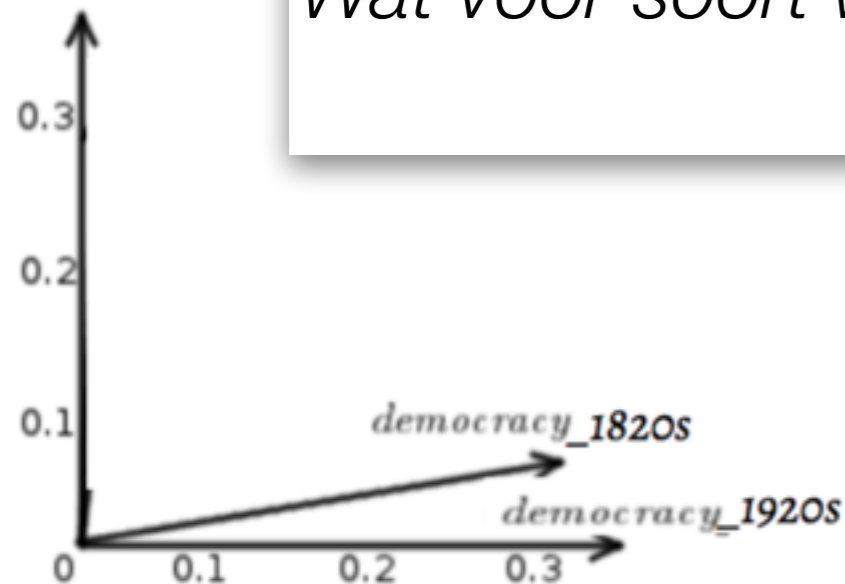
... deze techniek geeft alleen informatie over individuele woorden.

*Open vragen:*

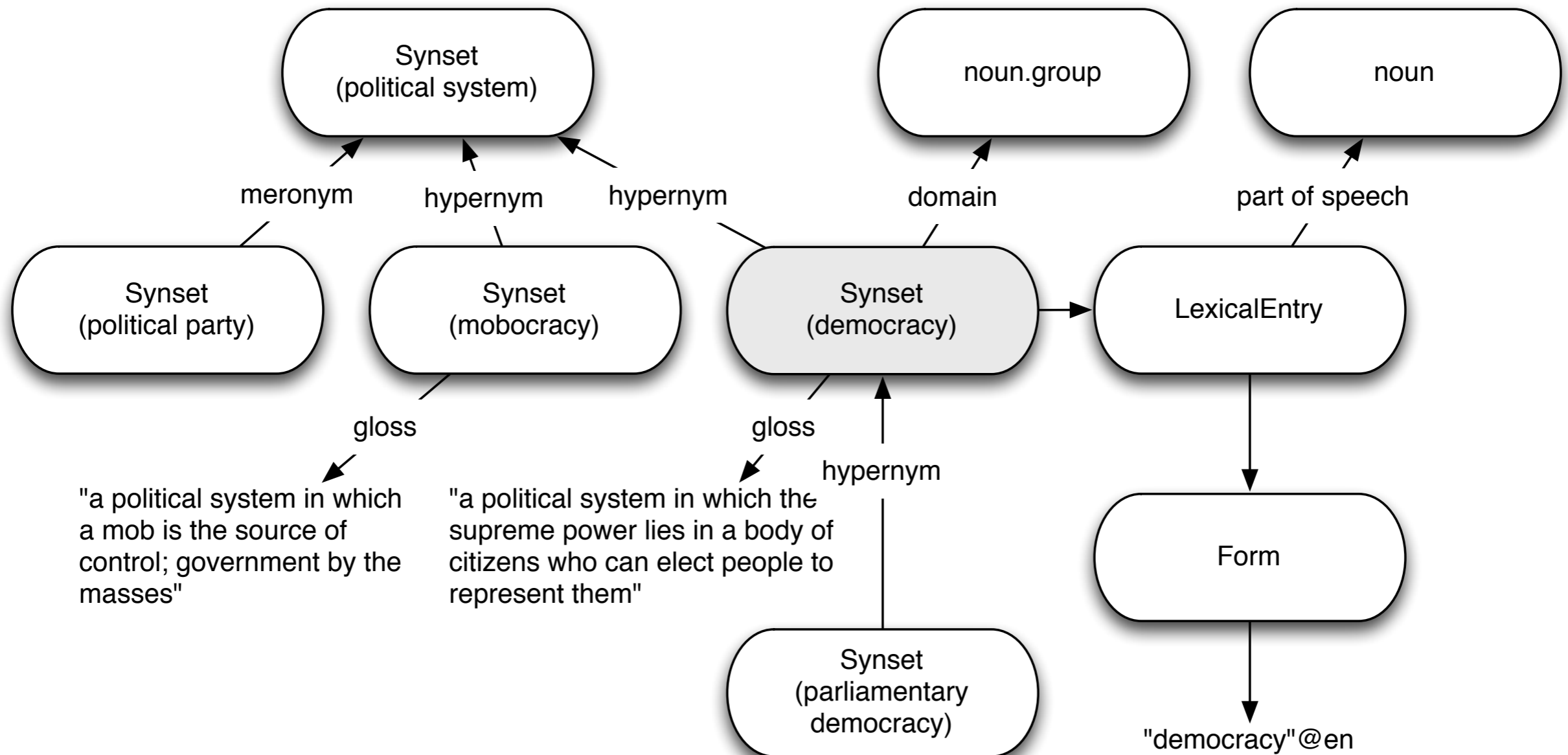
*Zijn synoniemen ook meeveranderd? En hyponiemen?*

*Zijn alle woorden voor politieke systemen veranderd?*

*Wat voor soort woorden/concepten zijn het meest veranderd?*



# We vinden deze relaties in Linked Open Data



example of data from WordNet RDF

Maar...

---

... Linked Open Data is vaak statisch en bevat geen info over verandering.

Maar...

---

... Linked Open Data is vaak statisch en bevat geen info over verandering.

*(Hoewel, er is LOD sensor-data, veranderende rollen van mensen, nieuwe lidstaten van de Europese Unie, enz.)*



Maar...

---

... Linked Open Data is vaak statisch en bevat geen info over verandering.

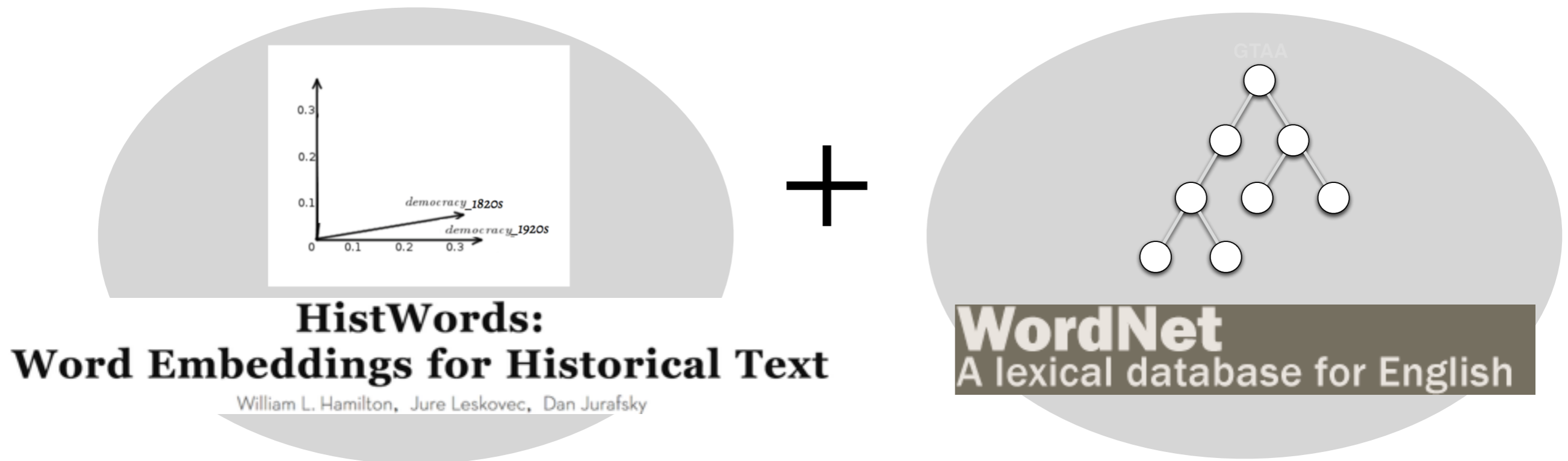
*(Hoewel, er is LOD sensor-data, veranderende rollen van mensen, nieuwe lidstaten van de Europese Unie, enz.)*

Subtiele veranderingen in context en frequentie waarin concepten voorkomen zijn nauwelijks gerepresenteerd in Linked Open Data.

# Onze bijdrage

---

- \* Een methode om twee datasets te koppelen



- \* Een data model om de combinatie te representeren
- \* Een RDF dataset waarop SPARQL queries gedaan kunnen worden



Disclaimer: tot nu toe 10k woorden gekoppeld

# Voorbeeldquery

---

In WordNet zijn alle woorden ingedeeld in 46 ‘domeinen’.

*Welk domein is het meest veranderd in de afgelopen twee eeuwen?*

<b>Mean change score</b>	<b>Domain</b>
0.909	noun.process
0.872	noun.phenomenon
0.869	noun.event
0.867	noun.act
0.86	noun.possession
⋮	
0.814	verb.body
0.791	noun.animal
0.784	noun.food
0.778	noun.feeling
0.737	verb.weather

# Voorbeeldquery

---

In WordNet zijn alle woorden ingedeeld in 46 ‘domeinen’.

*Welk domein is het meest veranderd in de afgelopen twee eeuwen?*

<b>Mean change score</b>	<b>Domain</b>
0.909	noun.process
0.872	noun.phenomenon
0.869	noun.event
0.867	noun.act
0.86	noun.possession

Vervolgvraag: welk subconcept van ‘psychologische toestand’ is het meest veranderd?

⋮

0.814	verb.body
0.791	noun.animal
0.784	noun.food
0.778	noun.feeling
0.737	verb.weather

# Voorbeeldquery

---

In WordNet zijn alle woorden ingedeeld in 46 ‘domeinen’.

*Welk domein is het meest veranderd in de afgelopen twee eeuwen?*

Mean change score	Domain
0.909	noun.process
0.872	noun.phenomenon
0.869	noun.event
0.867	noun.act
0.86	noun.possession

Vervolgvraag: welk subconcept van ‘psychologische toestand’ is het meest veranderd?

➔ Stress

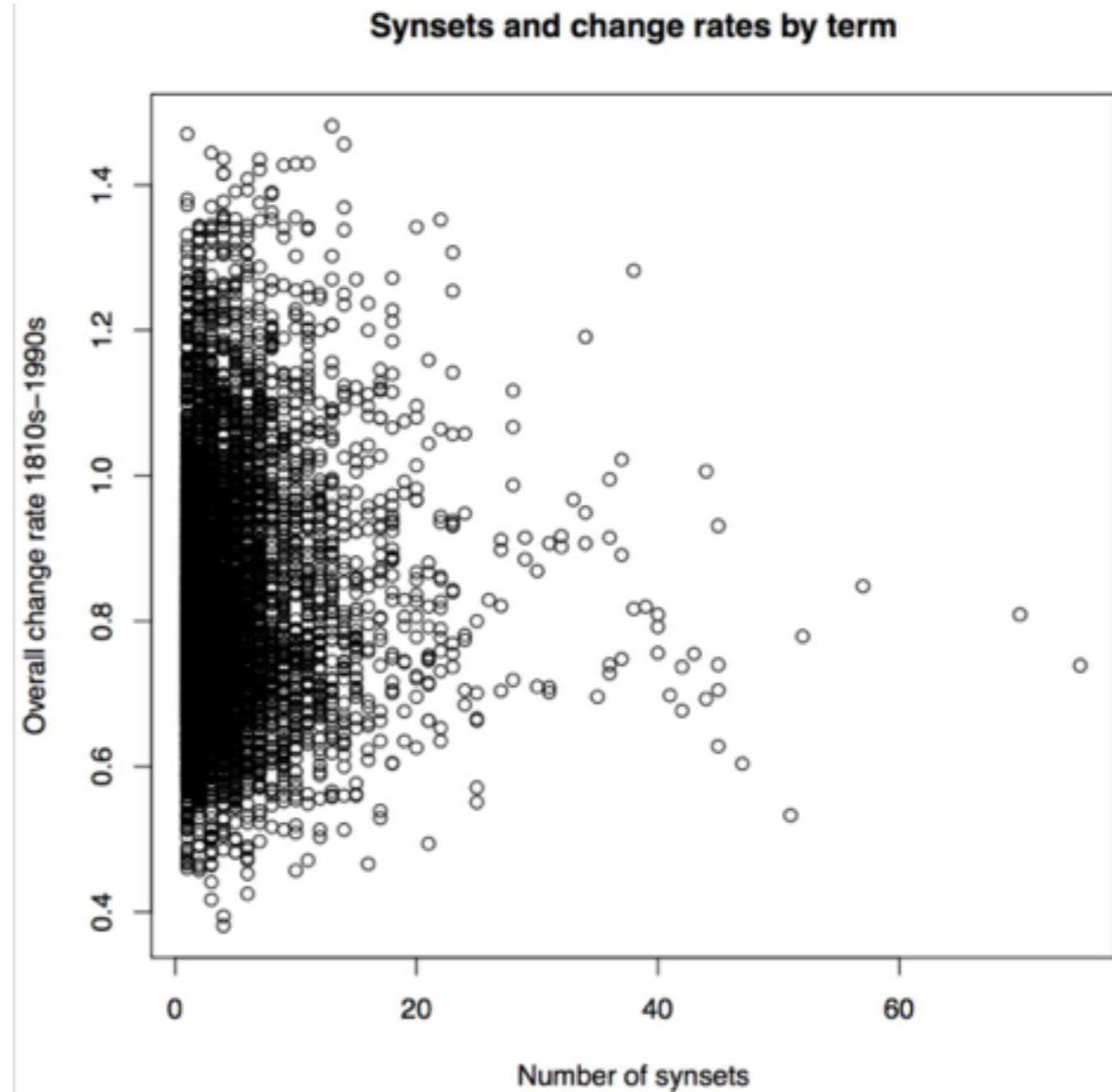
⋮

0.814	verb.body
0.791	noun.animal
0.784	noun.food
0.778	noun.feeling
0.737	verb.weather

# Voorbeeldquery

---

*Is er een relatie tussen polysemie en de hoeveelheid verandering van een woord?*



# Conclusie

---

Een eerste stap op weg naar een verrijking van de gestructureerde data in de Linked Open Data cloud met informatie uit de ongestructureerde, rijke, interpretabele bronnen van de geesteswetenschappen en sociale wetenschappen.

