

# ***A Thesaurus of Old English*** as Linguistic Linked Data

---

bringing topical thesauri to the Semantic Web

# This paper...

**project:** Exploring Anglo-Saxon Eloquence  
(use cases for utilizing linguistic linked data)



Universiteit  
Leiden



/instituut voor  
de Nederlandse  
taal/



university of  
 groningen



University  
of Glasgow



Universiteit  
Leiden

# This paper...

**topic:**

Porting thesauri to linguistic linked data

**case study:**

*A Thesaurus of Old English*



# Outline

---

## Introduction

- A Thesaurus of Old English

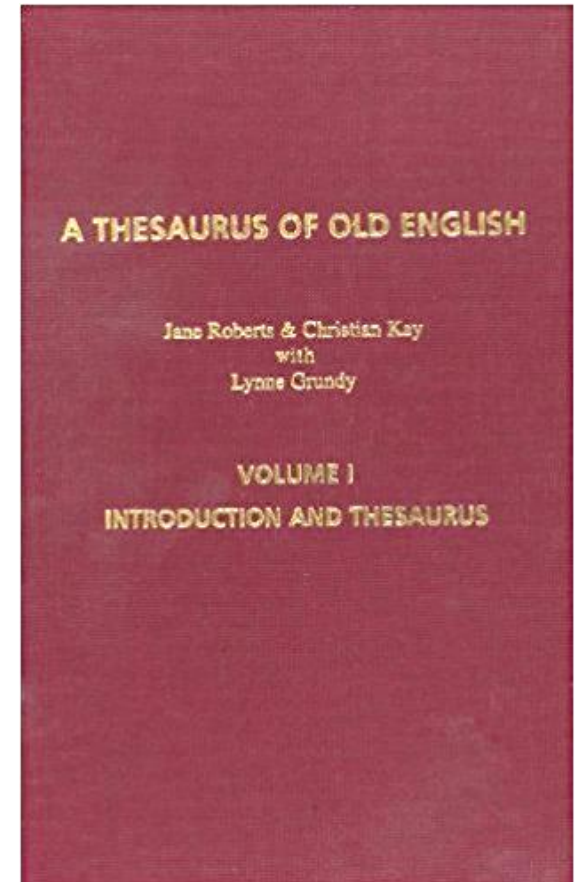
## From legacy to linked

- Legacy form
- Linguistic linked data form
- Conversion process

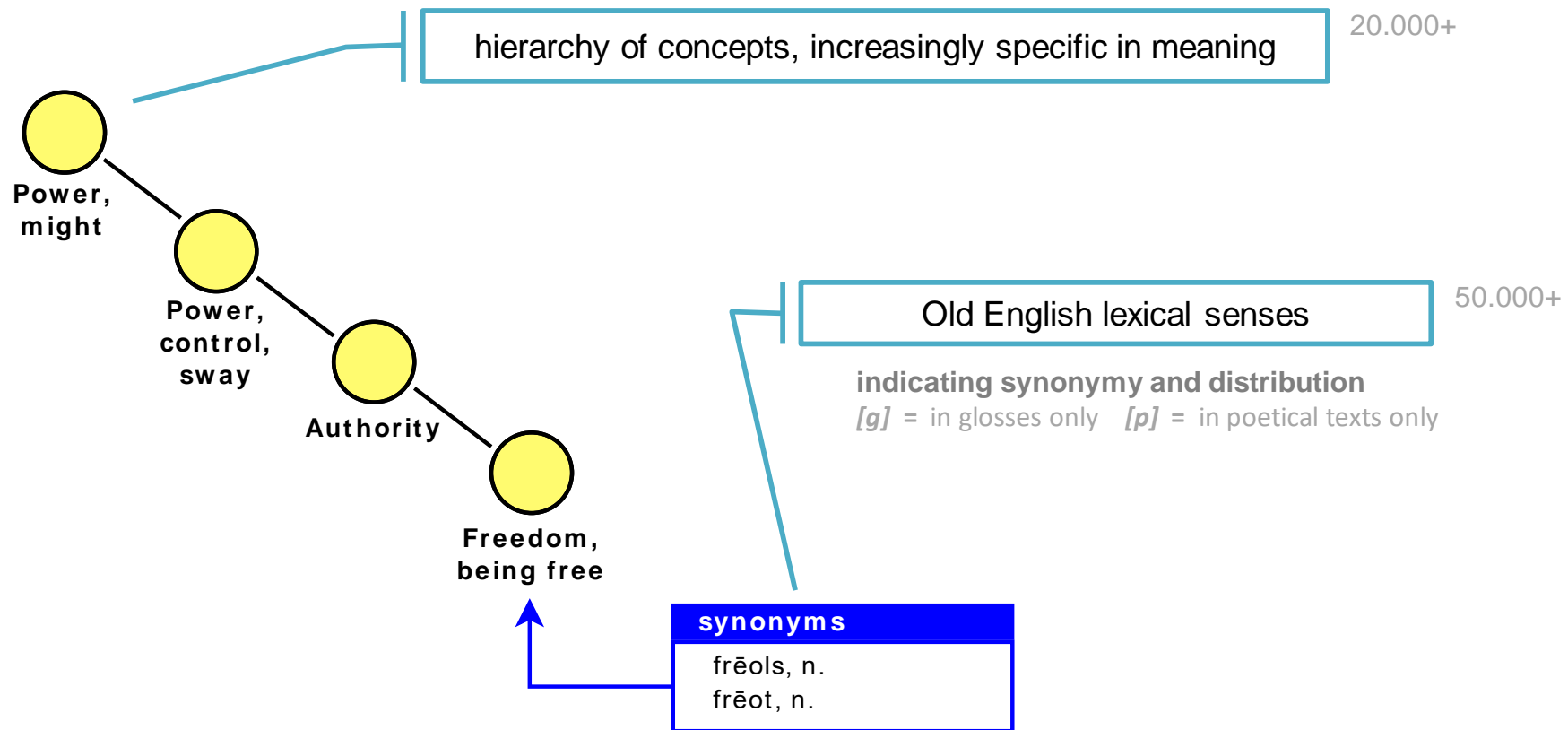
## Conclusion

# A Thesaurus of Old English

- **editors:**  
Jane Roberts, Christian Kay,  
Lynne Grundy
- **captures:**  
Old English lexis (~ 500-1100)
- **published:**  
1995, 2000 (book); 2005 (website)



# A Thesaurus of Old English



# From legacy to linked

---

legacy form

# Legacy form

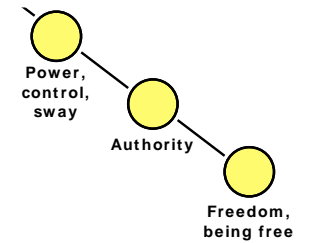
- **digital format:**  
Tabular / MySQL database
- **3 tables:**
  - ‘category’
  - ‘category-xref’
  - ‘lexeme’





# Legacy form

## table 1: 'category'



catid	t1	t2	t3	t4	t5	t6	t7	subcat	pos	heading	notes
1	1								N	Earth, world	
2	1							1	N	As God's creation	xr Religion
3	1							1.01	N	In the beginning	
...	...	...	...	...	...	...	...	...	...	...	...
17187	12	1	1	9				18	V	To accept as a slave	
17188	12	1	1	9				19	V	To bring into bondage	
17189	12	1	1	10					N	Freedom, being free	
17190	12	1	1	10				1	N	Citizenship	
17191	12	1	1	10				2	N	A free man	
17192	12	1	1	10				3	N	A free woman	
17193	12	1	1	10				4	N	Freeman of lowest class	

# Legacy form

table 2: 'category-xref'

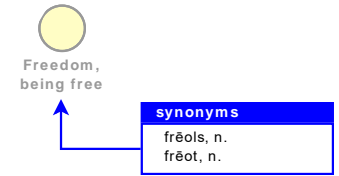


xid	catid	refid	tnum
1	18	588	01.03.01.05.01
2	18	9166	05.10.05.04.09
3	45	478	01.02.01.01.03
...	...	...	...
839	17189	16858	11.12.01
840	17189	18102	12.07.03

highlight: cross-references at category "Freedom, being free"

# Legacy form

table 3: 'lexeme'



lid	catid	prefix	word	catorder	et	notes	oflag	pflag	gflag	qflag
1	1		brytengrundas	1		ChristA 355	Y	Y	N	N
2	1		brytenwangas	2		ChristA 380	Y	Y	N	N
3	1		eormengrund	3		Beo 859	Y	Y	N	N
...	...	...	...	...	...	...	...	...	...	...
39486	17187		hēafod niman	1			N	N	N	N
39487	17188	=	(ge)hæftan	1			N	N	N	N
39488	17189		frēols	1			N	N	N	N
39489	17189		frēot	2			N	N	N	N
39490	17190		burhræden	1			Y	N	Y	N
39491	17190		burhscipe	2			N	N	N	N
39492	17191		bonda	1	bond		N	N	N	N
39493	17191		ceorl	2	churl		N	N	N	N

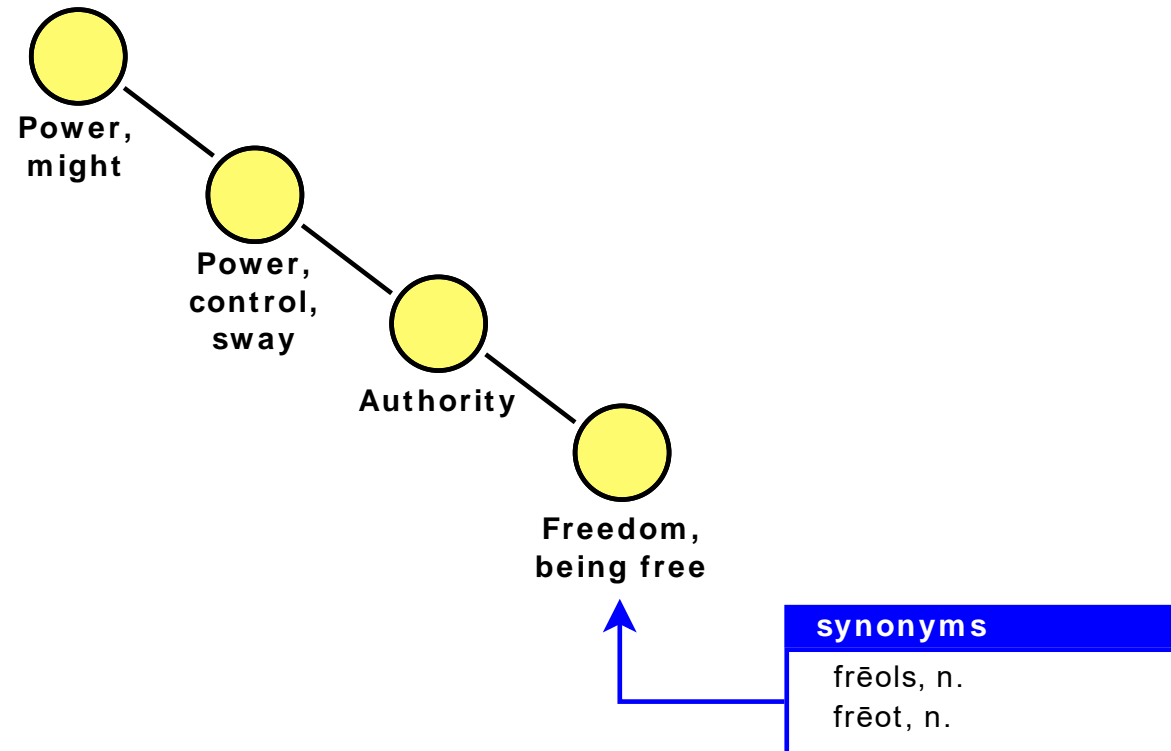
highlight: lexeme *frēols* found at category "Freedom, being free"

# From legacy to linked

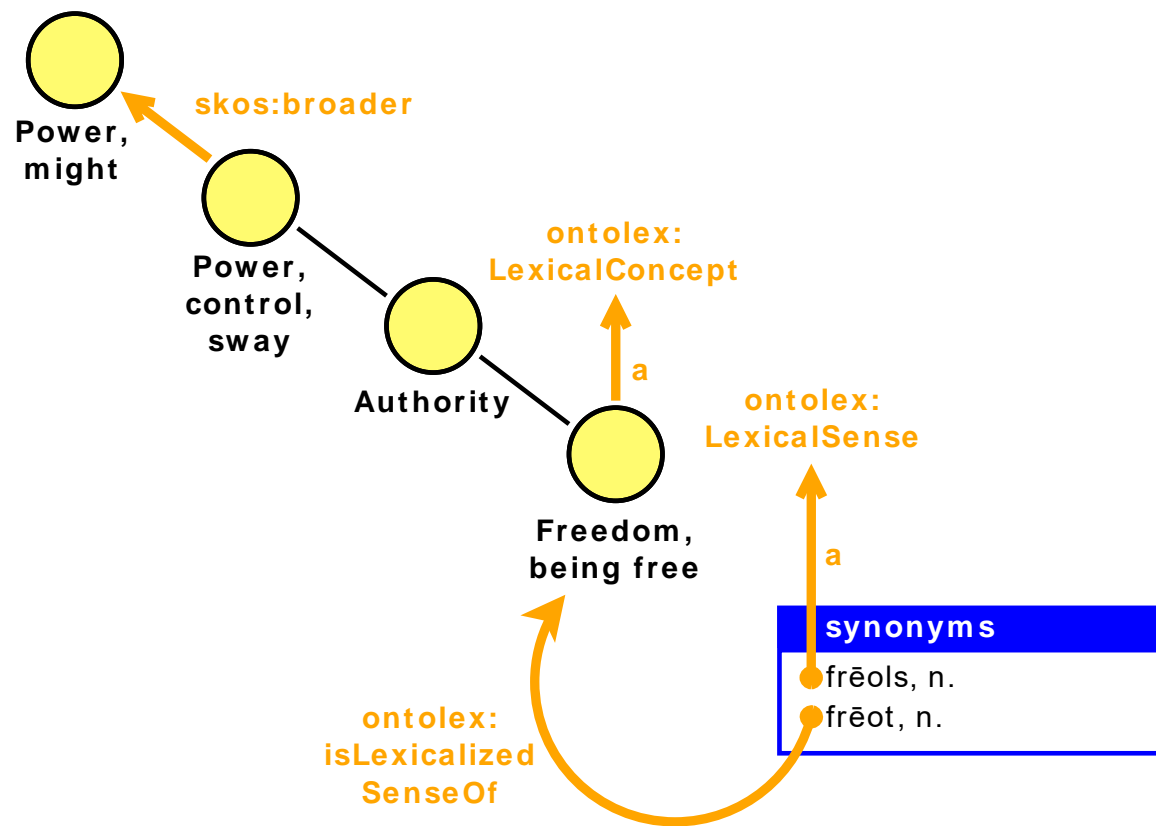
---

linguistic linked data form

# Linguistic linked data form



# Linguistic linked data form



## ontologies used:

- OntoLex
- SKOS
- Lemon-tree

# Linguistic linked data form

- mapping to linked data:  
straightforward... or, mostly so
- exceptions:
  - editorial or superseded

er	et	notes	oflag
1		ChristA 355	Y
2		ChristA 380	Y
3		Beo 859	Y
...	...	...	...
1			N
1			N
1			N
2			N
1			Y
2			N
1	bond		N
2	churl		N

*lexeme table*

# Linguistic linked data form

- mapping to linked data:  
straightforward... or, mostly so
- exceptions:
  - editorial or superseded
  - category identification

catid	t1	t2	t3	t4	t5	t6
1	1					
2	1					
3	1					
...	...	...	...	...	...	.
17187	12	1	1	9		
17188	12	1	1	9		
17189	12	1	1	10		
17190	12	1	1	10		
17191	12	1	1	10		
17192	12	1	1	10		
17193	12	1	1	10		

*category table*



# Linguistic linked data form

- mapping to linked data:  
straightforward... or, mostly so
- exceptions:
  - editorial or superseded
  - category identification
  - ‘lexeme’ ambiguity  
(sense vs entry)

catid	r	oflag	pflag	gflag	qflag
1		Y	Y	N	N
1		Y	Y	N	N
1		Y	Y	N	N
...		...	...	...	...
17187		N	N	N	N
17188		N	N	N	N
17189		N	N	N	N
17189		N	N	N	N
17190		Y	N	Y	N
17190		N	N	N	N
17191		N	N	N	N
17191		N	N	N	N

*lexeme table*

# From legacy to linked

---

conversion process

# Conversion process

step 1: express structure in (non-linguistic) linked data

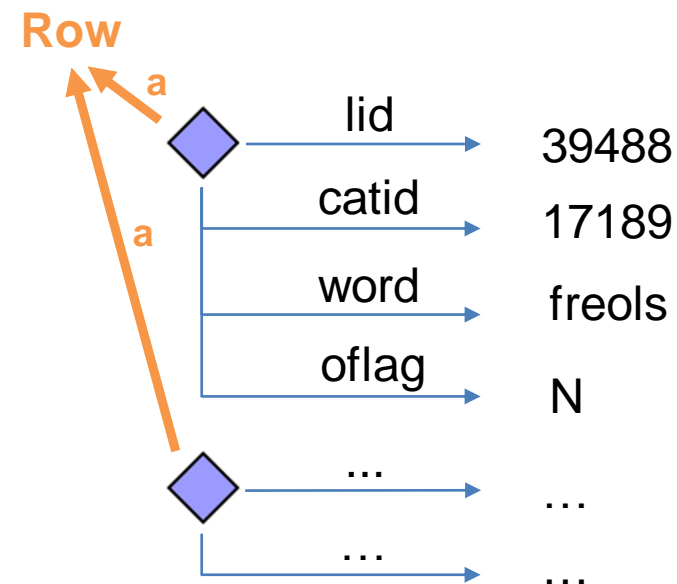
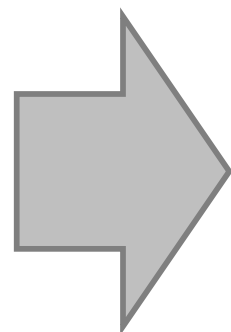
catid	t1	t2	t3
1	1		
2	1		
3	1		
...	...	...	...
17187	12	1	
17188	12	1	
171			
171			
171			
171			
171			

lid	catid	prefix	word	catorder	et	notes	oflag	pflag	gflag	qflag
1	1		brytengrundas	1		ChristA 355	Y	Y	N	N
2	1		brytenwagas	2		ChristA 380	Y	Y	N	N
3	1		eormengrund	3		Beo 859	Y	Y	N	N
...	...	...	...	...	...	...	...	...	...	...
39486	17187		hēafod niman	1			N	N	N	N
39487	17188	=	(ge)hæftan	1			N	N	N	N
39488	17189		frēols	1			N	N	N	N
39489	17189		fréot	2			N	N	N	N
39490	17190		burhræden	1			Y	N	Y	N
39491	17190		burhscipe	2			N	N	N	N
39492	17191		bonda	1	bond		N	N	N	N

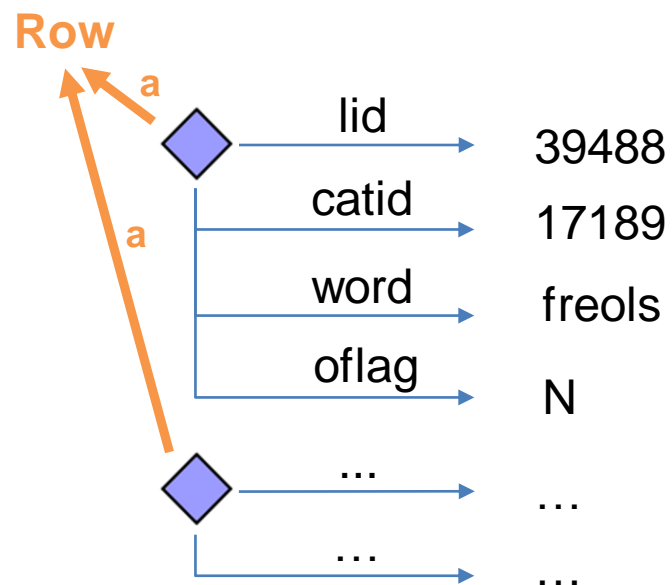
  

xid	catid	refid	tnum
...	...	...	...
839	17189	16858	11.12.01
840	17189	18102	12.07.03



# Conversion process

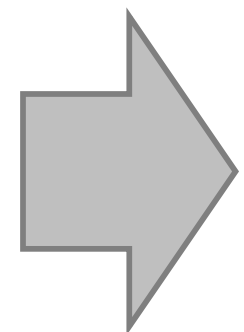
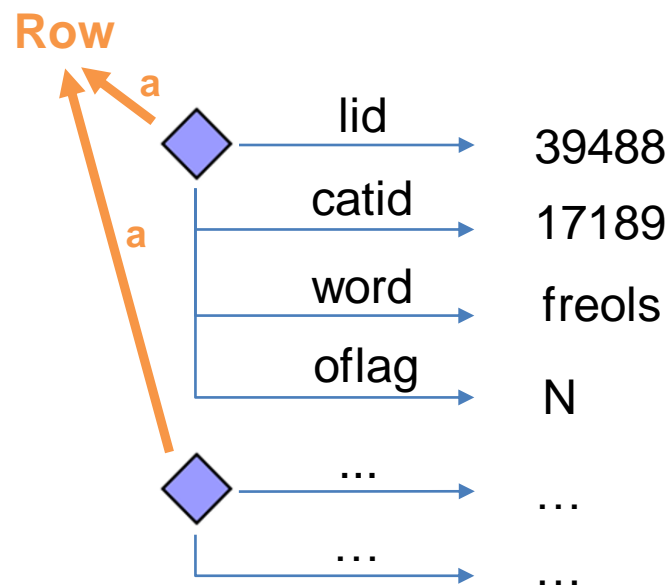
step 2: perform mapping using SPARQL



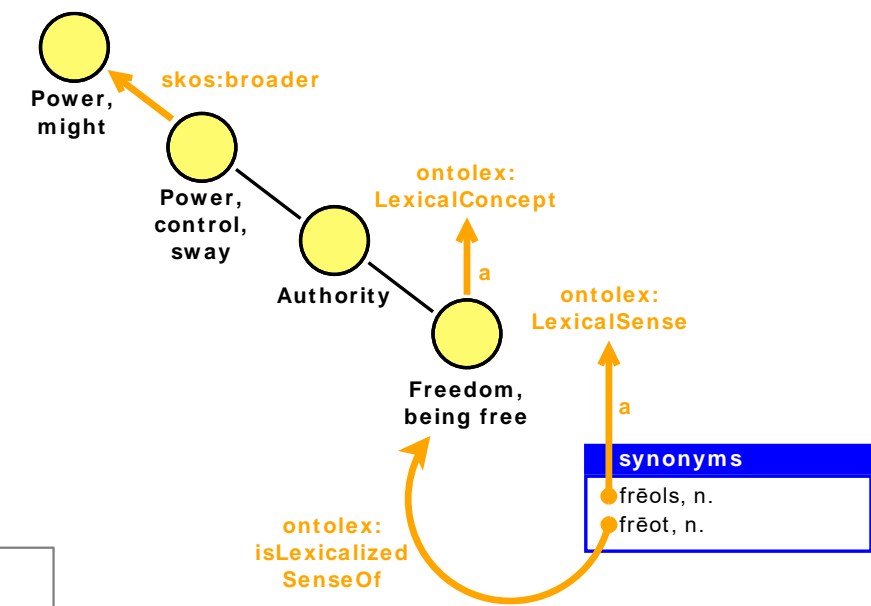
```
CONSTRUCT {  
  # construct LL(O)D pattern  
  ?tgtSense a ontolex:LexicalSense .  
  ...  
}  
WHERE {  
  # select pattern in structure  
  ?row a :Row .  
  ?row :lid ?lid .  
  ...  
}
```

# Conversion process

step 2: perform mapping using SPARQL



```
CONSTRUCT {  
  # construct LL(O)D pattern  
  ?tgtSense a ontolx:LexicalSense  
  .  
  ...  
}  
WHERE {  
  # select pattern in structure  
  ?row a :Row .  
  ?row :lid ?lid .  
  ...  
}
```



# Conclusion

---

# Conclusion

topical thesauri can be ported to the Semantic Web

*with expressivity and interoperability provided  
through OntoLex, SKOS and Lemon-tree*

conversion process can be relatively straightforward

*with the possibility to separate  
translation of format and meaning*



# Conclusion

## recommendations

- establish fit of source and target  
*is terminology in line with content?*
- establish exceptions and nuances  
*what is missing, what should be disregarded?*
- capture mapping in standardized language  
*reduces dependency on a single tool*





# Result + Future work

evoke Thesaurus of Old English

freedom

all > ● Power, might > ● Power, control, sway > ● Authority > ● Freedom, being free

Open

Freedom, being free ●  
*open*

- A free man
- A free woman
- Citizenship
- Free, not in bondage
- Freeing, manumission
- Freeman of lowest class
- Half-free
- Having full rights
- Man of class between slave and 'ceorl'
- One able to choose his lord

click on a concept to browse; press it longer to open it and view its contents

evoke

- **App for exploring**  
evoke.ullet.net
- **Research using result**  
12 papers @ ICEHL-21  
Leiden, 8 June 2020