



L

# The Prolex Project

A multilingual platform  
to Proper Name Processing



# Motivations

L



# Proper Names

- The common denominator of many definitions of proper names proposed by linguists may be

Proper name refers

as opposed to

Common noun is defined.



# Proper Names



- The prototypes of proper names are **personal names** and **toponyms**...

Lech Kaczynski, Jacques Chirac,  
Poland, France  
Warsaw, Paris



# Proper Names



- But we want also accept as proper names:  
names of organizations, events, etc. as in  
Named Entity task...

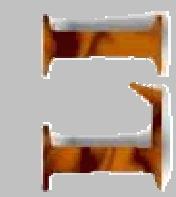
the independent trade union Solidarity  
the Warsaw Pact  
the Eiffel Tower

L

# Unknown words

- Somebody considers capitalized unknown words to proper names...
- It is not exact... above all multiword units!
- See four examples extracted to newspaper  
*Le Monde*





# Unknown words

Proper names	Unknown words
<i>Banque Centrale Européenne</i>	0
<i>Banco Real</i>	1
<i>Banca commerciale italiana</i>	2
<i>Banca di Roma</i>	3



# Relations

- Texts often play with anaphoric relation between Proper names.

John Paul II  
Karol Jozef Wojtyla  
the Polish pope



# Translation



- Somebody thinks that Proper Names are not translated but it is not true...

Warszawa

Warsaw

Varsovie

Чехов

Chekhov

Чехов, Тчекхов

Karol Józef Wojtyła  
Karol Jozef Wojtyla

Чехов (nominative)

Чехова (accusative)

Чеховowi (dative)



# Some realized works





# Inhabitant Names in French

- The derivation from City Name to Inhabitant Name in French is not obvious (because, we frequently use an auxiliary base)

Inhabitant of Metz

Messin

Metzois

Messois



# Ambiguity of dot in French

- In French, dot is used at the end of sentences, but also as abbreviation...

M. (*Mr*) L. Kaczynski a rencontré (*has met*)  
M. J. Chirac à l'O.N.U. (*in UNO*)  
ce matin (*this morning*).  
↓  
Abbreviation

First name      Acronym  
End of sentence      12

# French Proper Noun recognition



```
<Person>
  <Status> M. </Status>
  <First> L. </First>
  <Name> Kaczynski </Name>
</Person>
```



# French Proper noun recognition



Newspaper	Le Monde
Recall	93,2
Precision	94,4



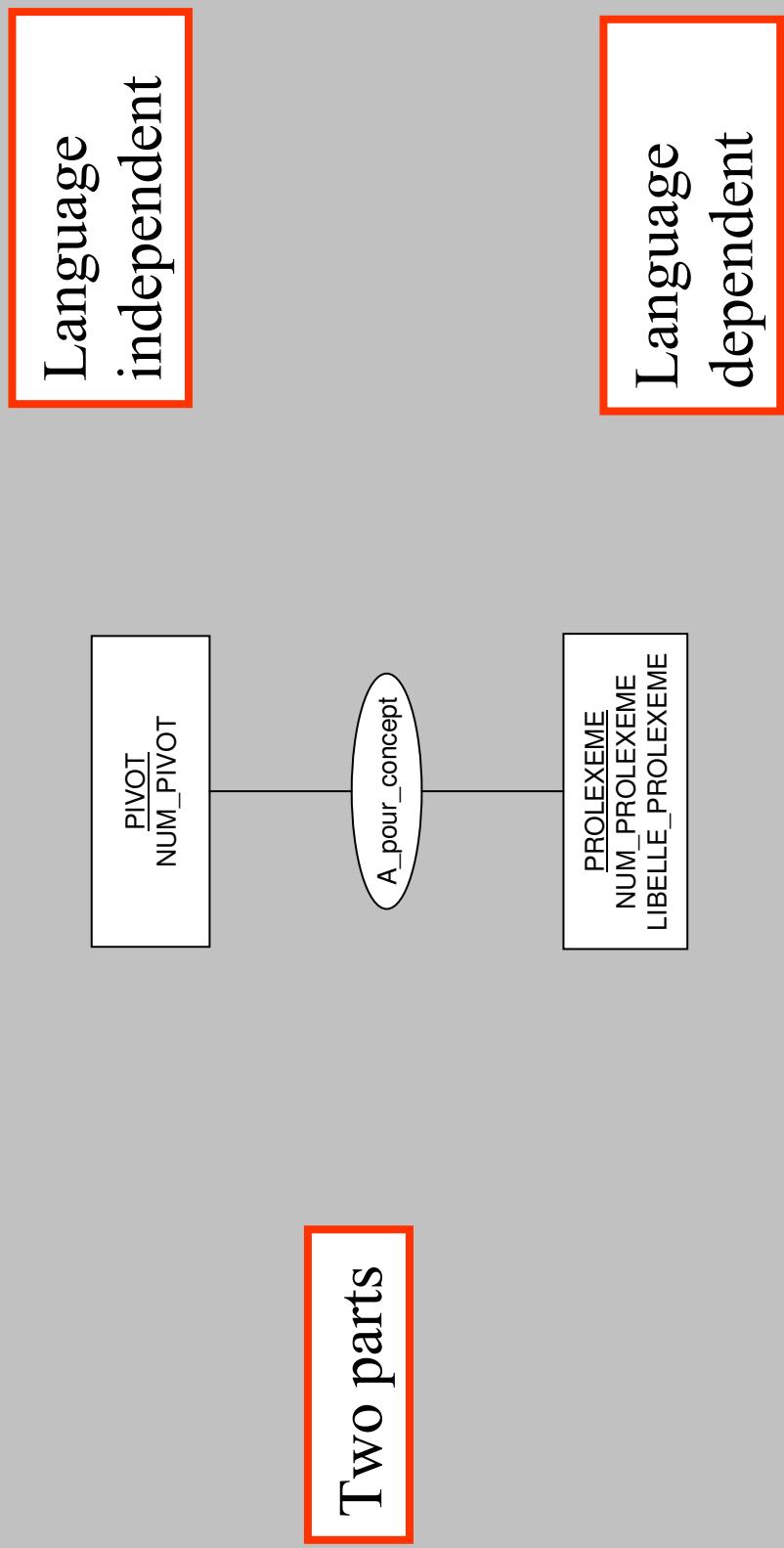
# Prolexbase

A multilingual relational database  
of Proper Name





# Two main concepts





# Two main concepts

Conceptual

Level

PIVOT  
NUM\_PIVOT

A\_pour\_concept

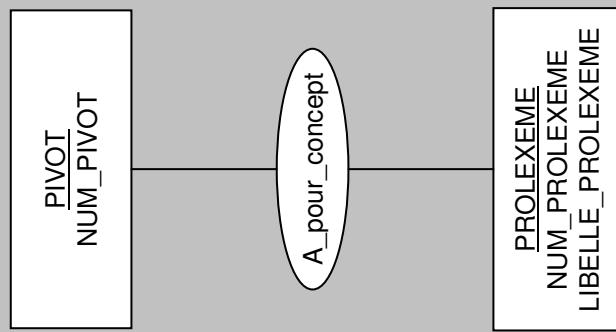
PROLEXEME  
NUM\_PROLEXEME  
LIBELLE\_PROLEXEME

Two parts



# Two main concepts

The conceptual  
Proper Name

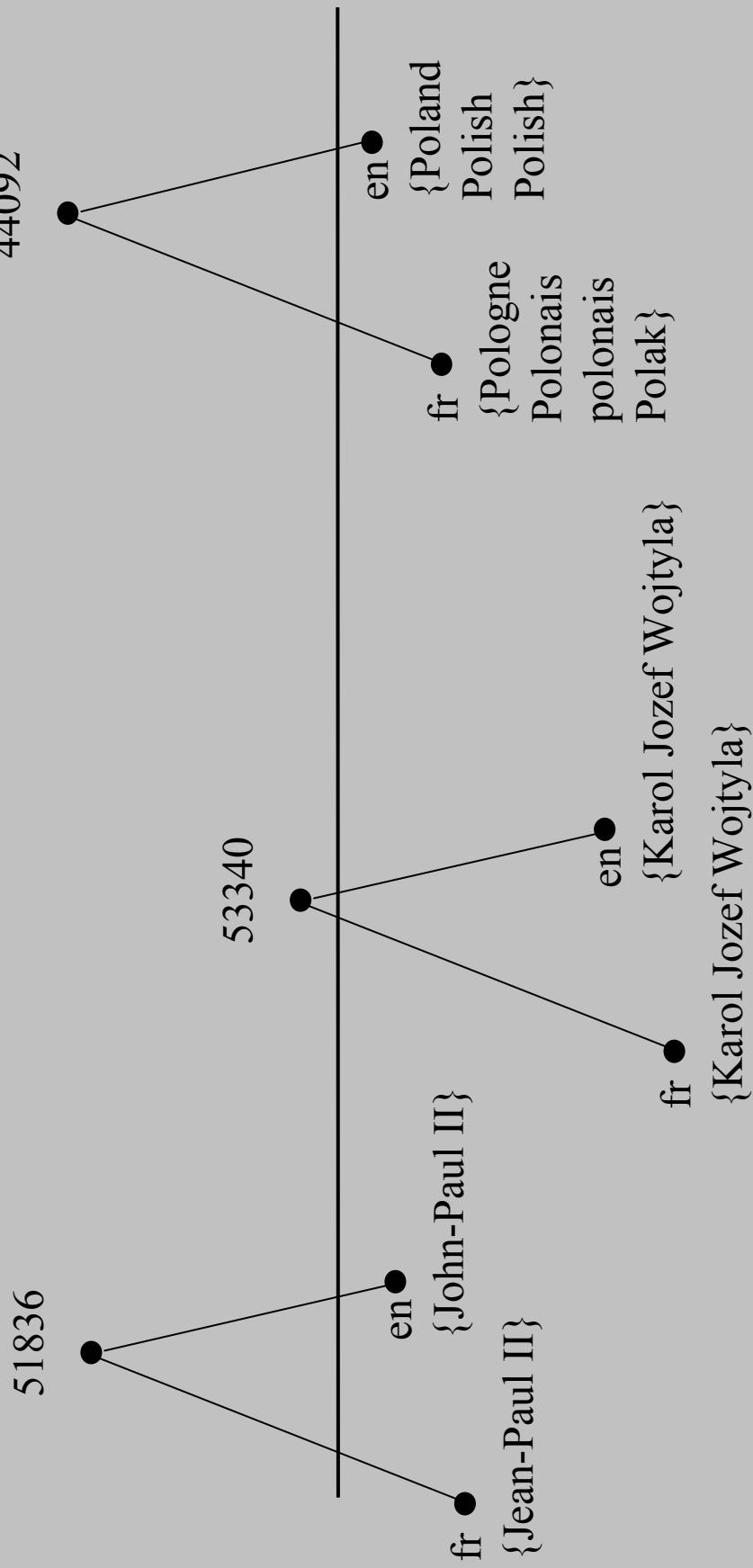


Two parts

A set of word  
semantically linked  
to the Proper Name



# Two main concepts





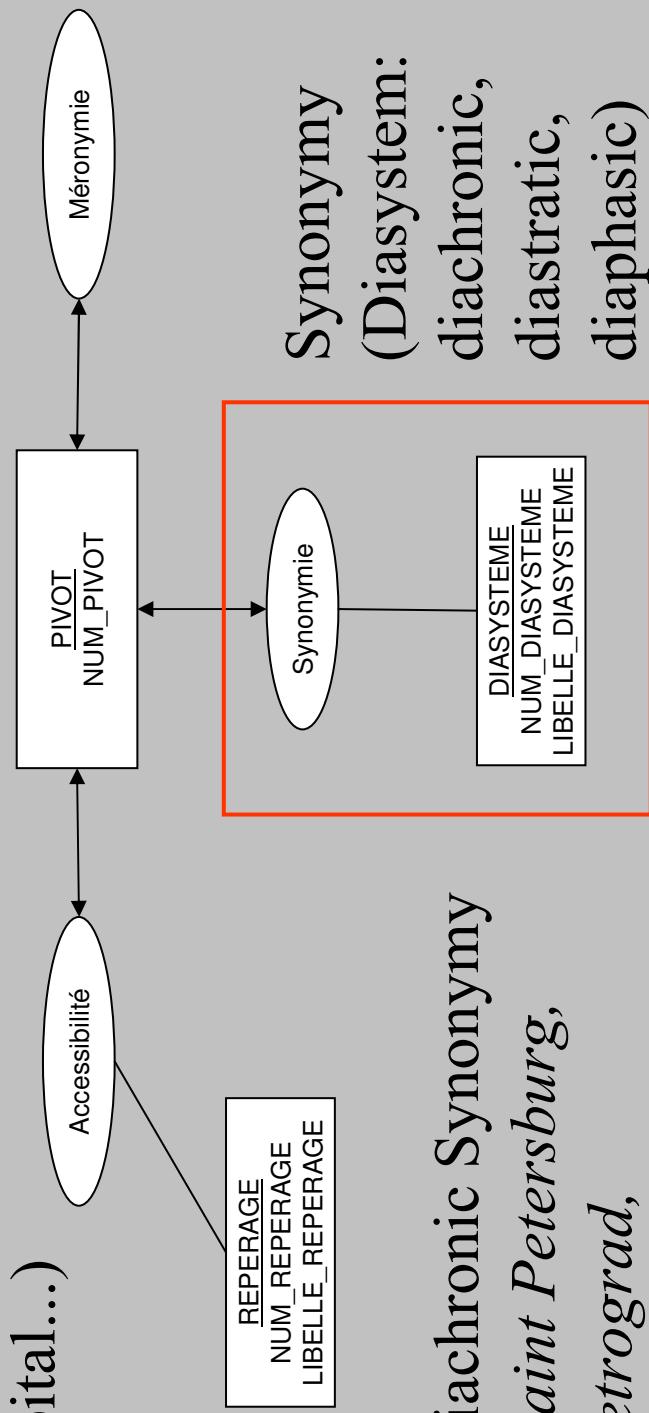
# Conceptual Level

Accessibility

(Location:  
family,  
authors,  
capital...)

*Metaconceptual Level  
(hyperonymies)*

Meronymy



*Diachronic Synonymy  
(Saint Petersburg,  
Petrograd,  
Leningrad)*

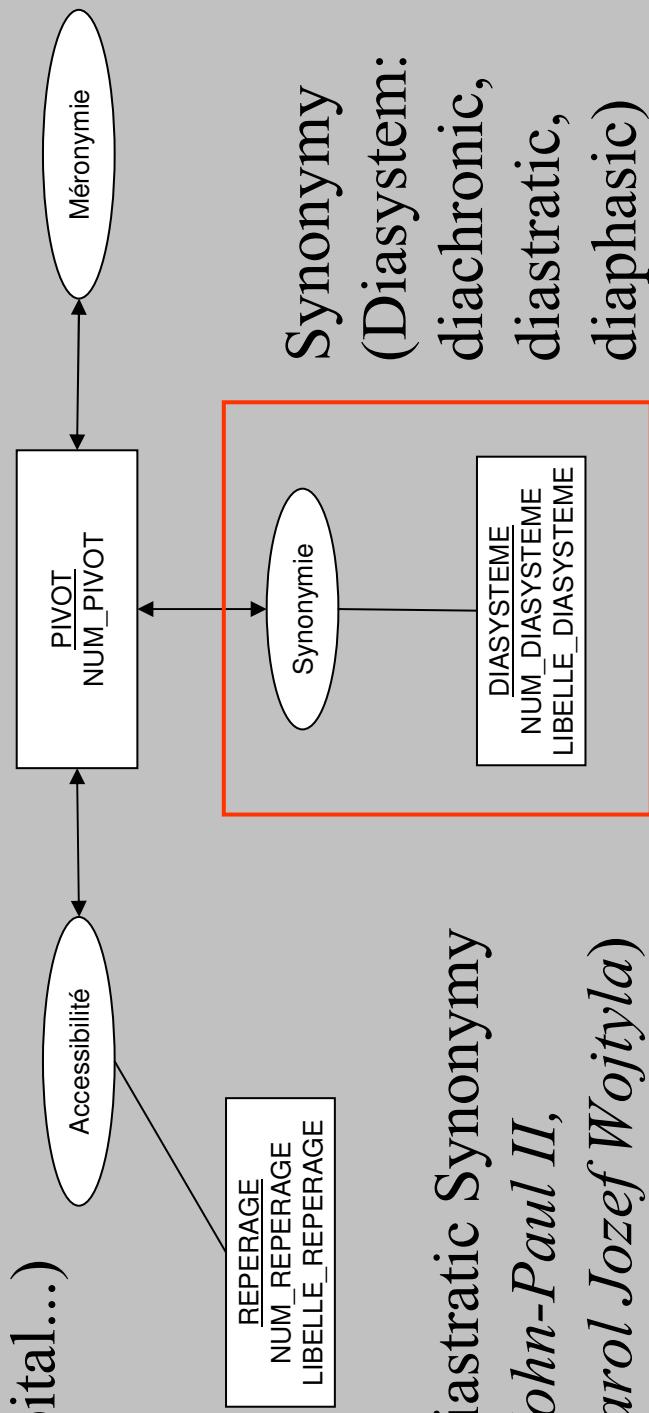


# Conceptual Level

Accessibility  
(Location:  
family,  
authors,  
capital...)

*Metaconceptual Level  
(hyperonymies)*

Meronymy



Diastatic Synonymy  
(*John-Paul II,*  
*Karol Jozef Wojtyla*)

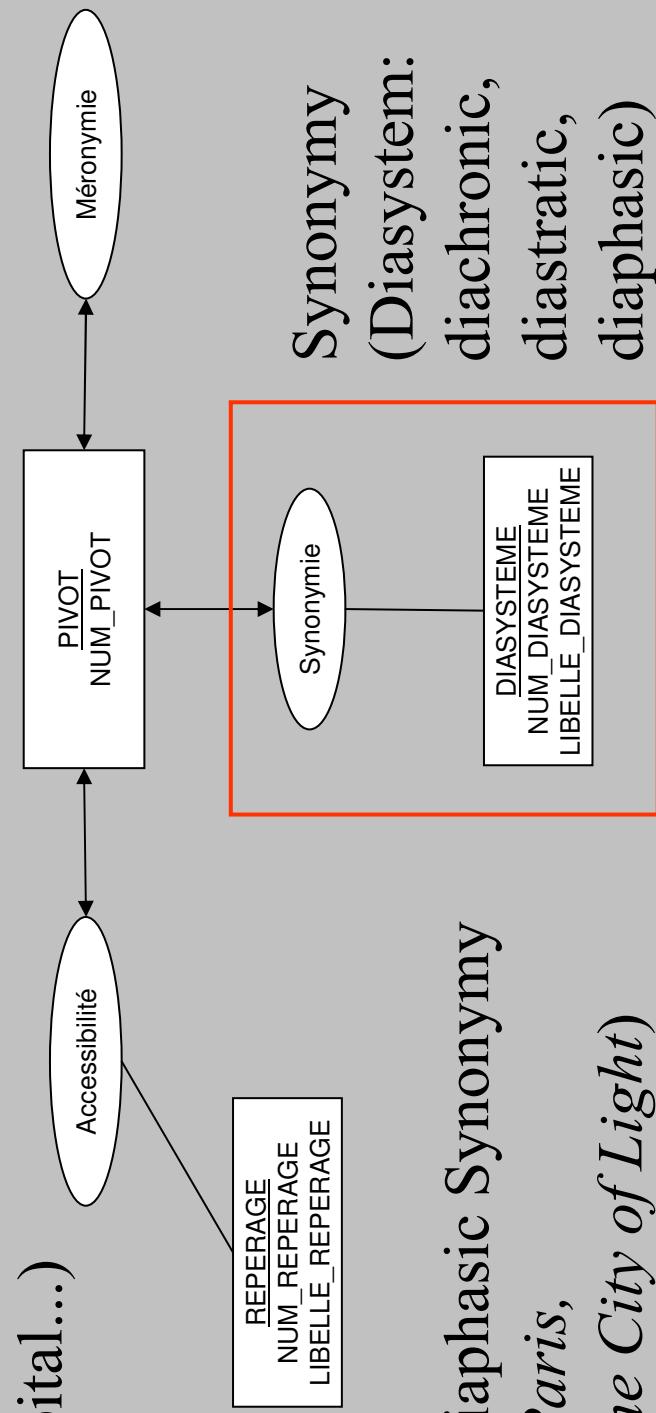


# Conceptual Level

Accessibility  
(Location:  
family,  
authors,  
capital...)

Metaconceptual Level  
(hyperonymies)

Meronymy



Diaphasic Synonymy  
(*Paris*,  
*The City of Light*)



# Conceptual Level

Accessibility

(Location:  
family,  
authors,  
capital...)

*Metaconceptual Level  
(hyperonymies)*

Meronymy

Accessibilité

PIVOT  
NUM\_PIVOT

REPERAGE  
NUM\_REPERAGE  
LIBELLE\_REPERAGE

Synonymie

Meronymy  
(*Paris and France*  
*John-Paul II and*  
*Poland*)

DIASYSTEME  
NUM\_DIASYSTEME  
LIBELLE\_DIASYSTEME

Synonymy  
(Diasytem:  
diachronic,  
diastatic,  
diaphasic)



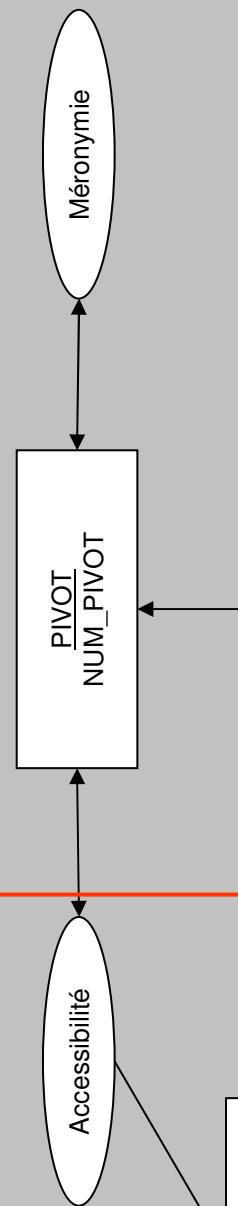
# Conceptual Level

Accessibility

(Location:  
family,  
authors,  
capital...)

*Metaconceptual Level  
(hyperonymies)*

Meronymy



Meronymy  
Synonymy

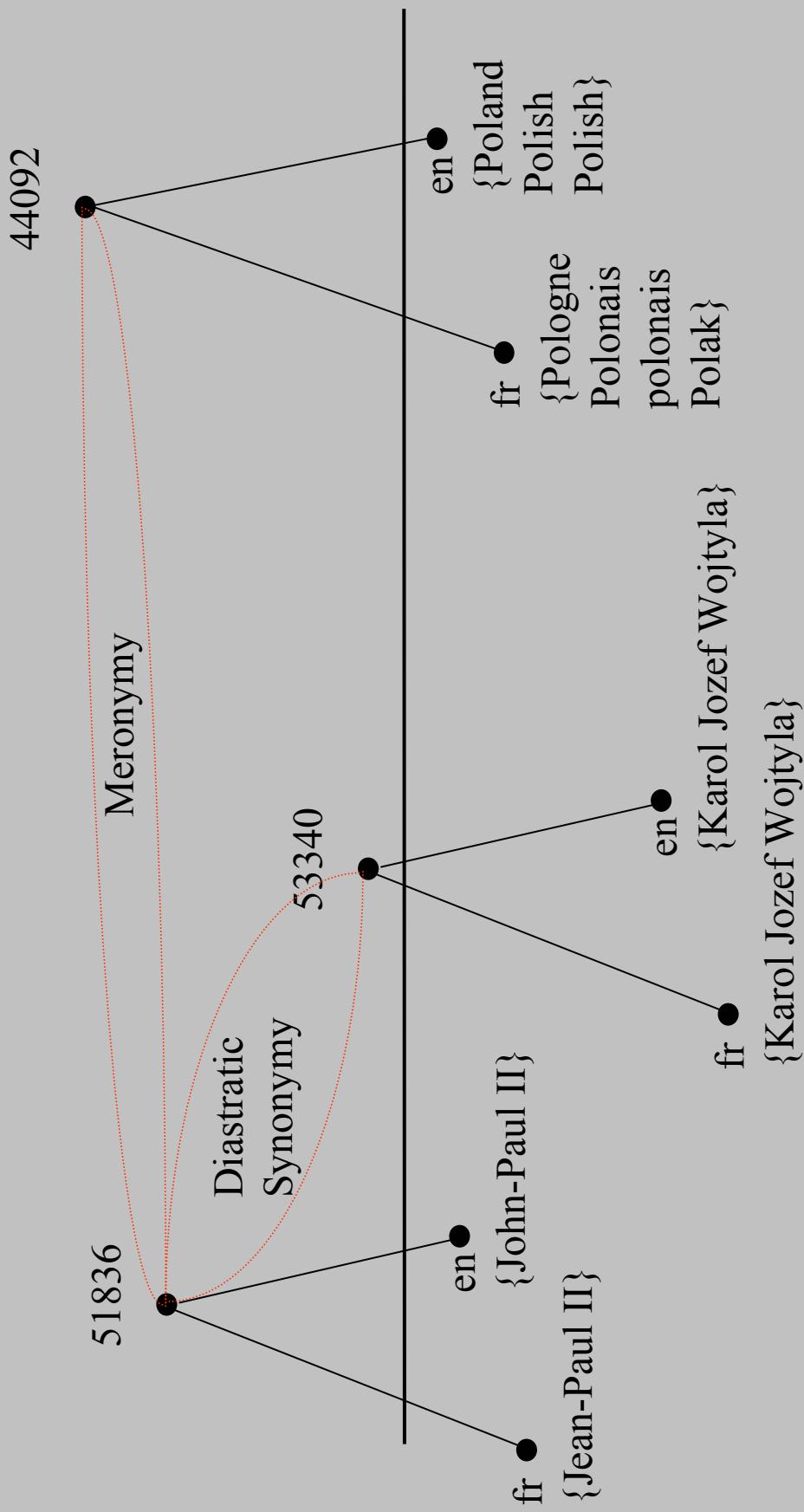
Synonymy  
(Diasytem:  
diachronic,  
diastratic,  
diaphasic)

DIASYSTEME  
NUM\_DIASYSTEME  
LIBELLE\_DIASYSTEME

Accessibility  
(*Paris, capital of France*  
*John-Paul II, pope of*  
*Catholic Church*)



# Conceptual Level





# Metaconceptual Level

Typology  
(Celebrity,  
Country,  
Hydronym...)

Existence  
(historical,  
fictive,  
religious)

Hyperonymie1

TYPE  
NUM\_TYPE  
LIBELLE\_TYPE

EXISTENCE  
NUM\_EXISTENCE  
LIBELLE\_EXISTENCE

Hyperonymie2

Hyperonymie3

PIVOT  
NUM\_PIVOT



# Typology

Proper Name		Toponym		Ergonym		Pragmonym
Anthroponym		Territory		Object		Disaster
Individual	Collective	Group	Astronym	Country	Product	Event
Celebrity	Dynastic	Association	Building	Region	Thought	Feast
First Name	Ethnonyme	Ensemble	City	Supra-national	Vessel	Hystory
Patronym	Pseudo-anthroponym	Firm	Geonym		Work	Meteorology
		Institution	Hydronym			
		Organization	Way			



# Linguistic Level

Derivate  
(Relative noun,  
Relative adjective,  
Prefix...)

A\_pour\_dérivé3

DERIVE  
NUM\_DERIVE  
LIBELLE\_DERIVE

Alias  
(variation,  
abbreviative,  
acronym...)

A\_pour\_dérivé2

ALIAS  
NUM\_ALIAS  
LIBELLE\_ALIAS

A\_pour\_alias

PROLEXEME  
NUM\_PROLEXEME  
LIBELLE\_PROLEXEME

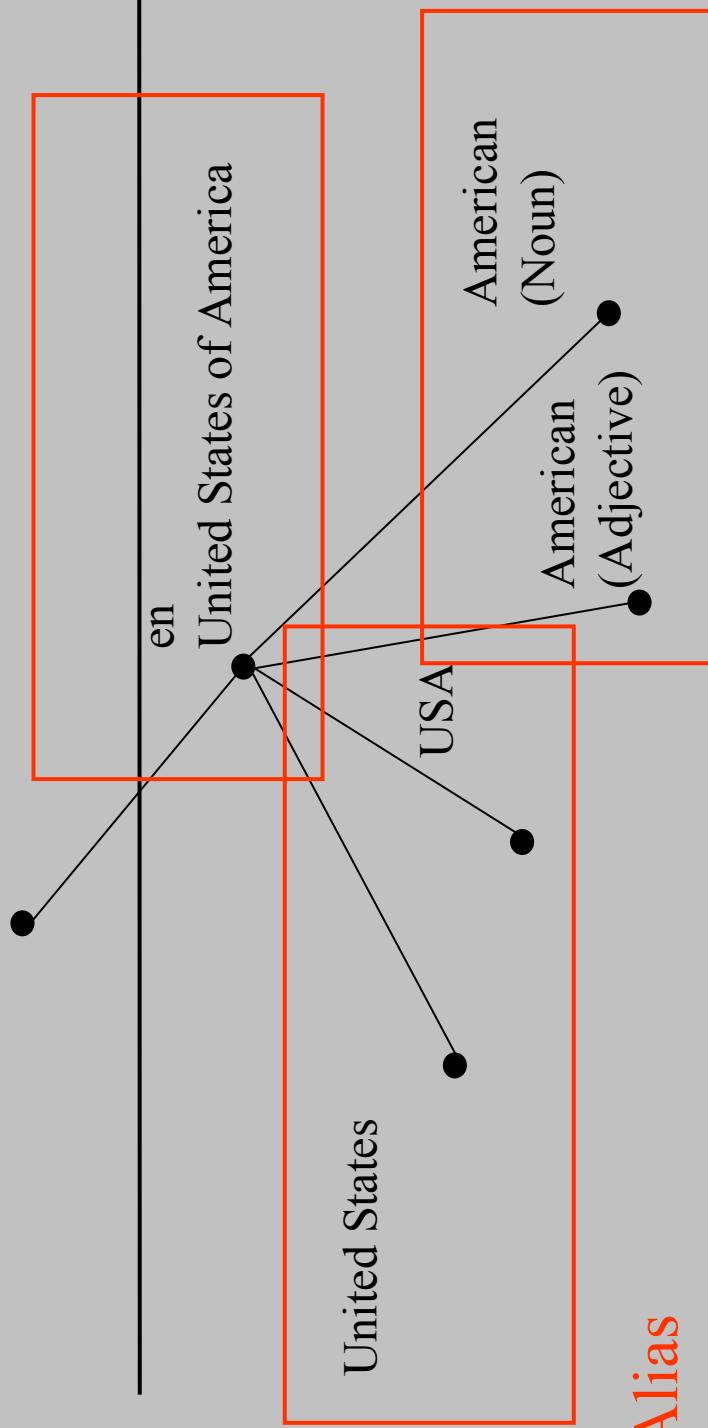
A representative  
Proper Noun



# Linguistic Level

46929

## Prolexeme



## Alias

## Derivatives



# Linguistic Level

Rules to  
generate  
derivatives

DERIVATION  
NUM\_DERIVATION  
LIBELLE\_DERIVATION

A\_pour\_dérivation3

A\_pour\_dérivation2

ALIASISATION  
NUM\_ALIASISATION  
LIBELLE\_ALIASISATION

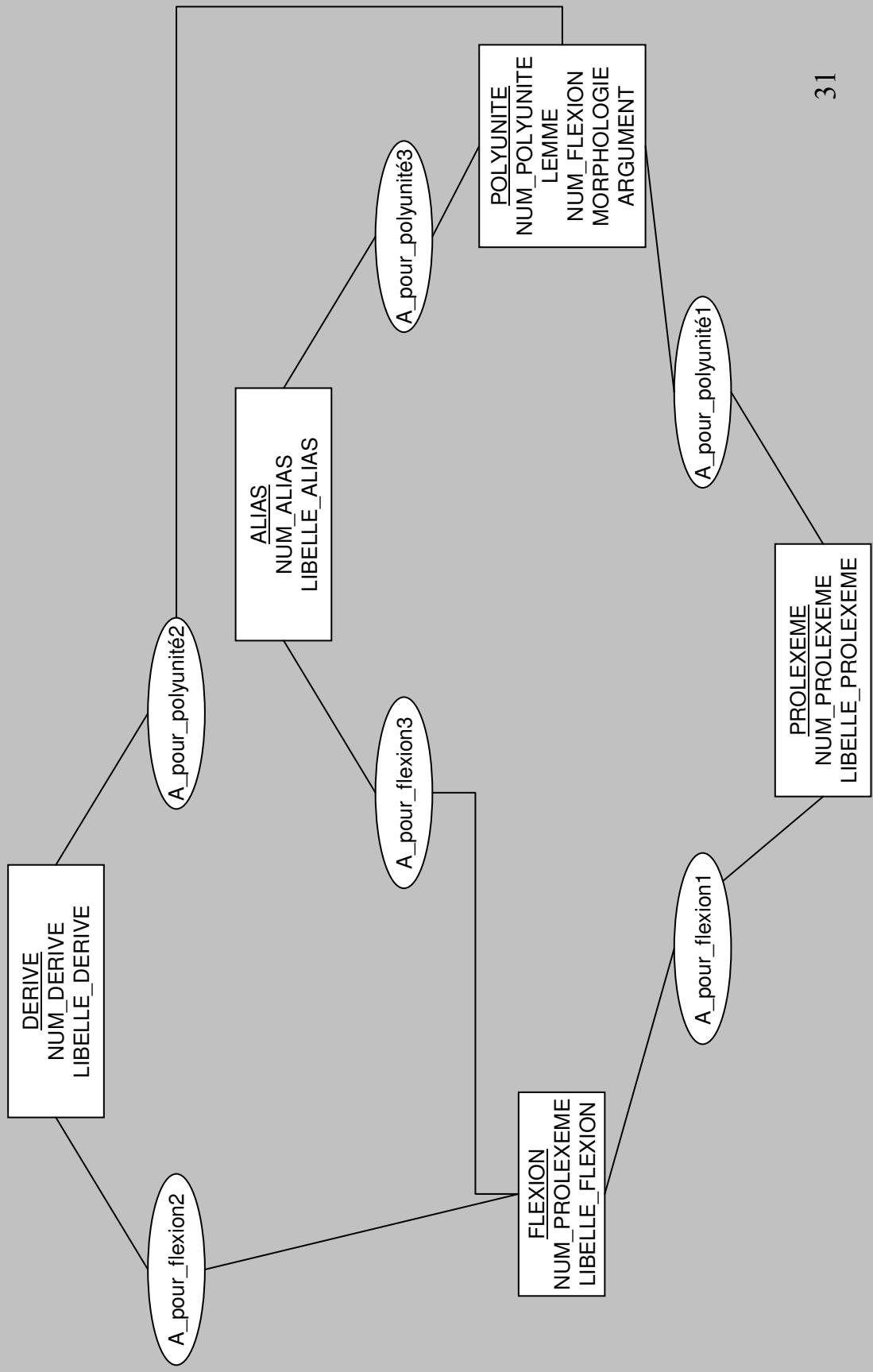
A\_pour\_dérivation1

PROLEXEME  
NUM\_PROLEXEME  
LIBELLE\_PROLEXEME

A\_pour\_aliasisation



# Inflexion





# Instance Level

46929

en  
United States of America

United States

USA

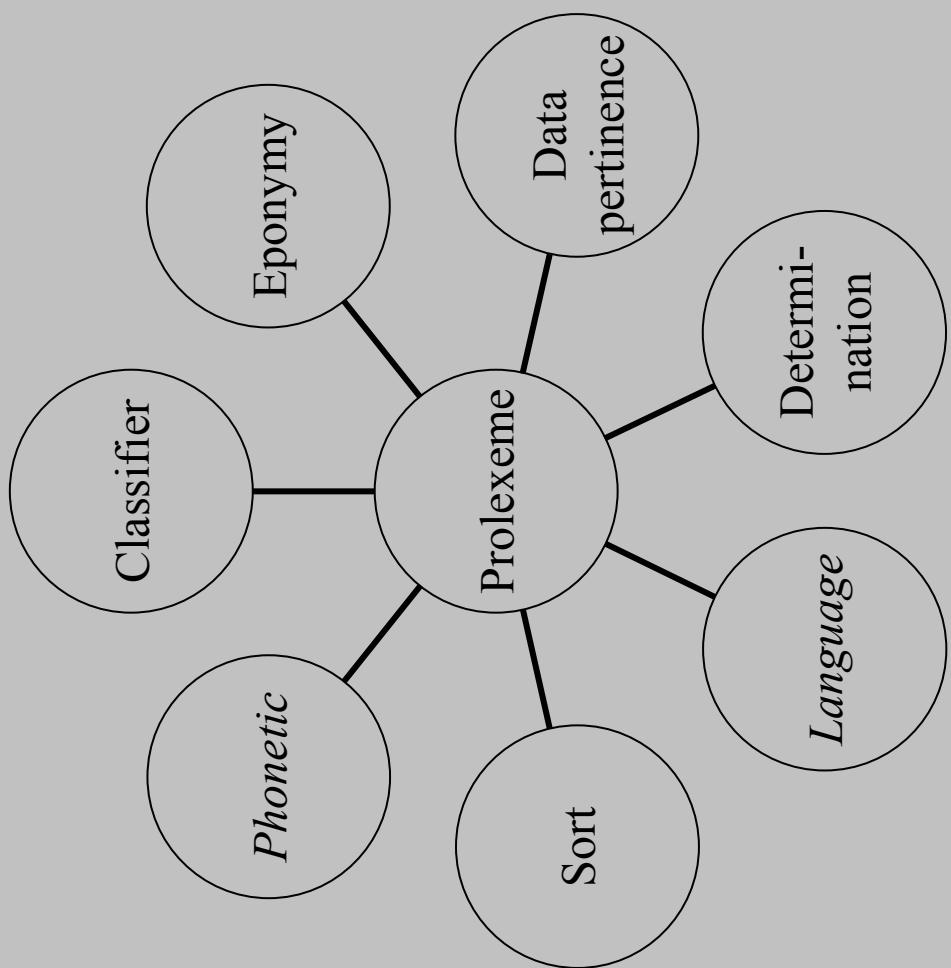
American  
(Noun)

American  
(Adjective)

Instances

American American  
Americans  
United States USA  
United States of America

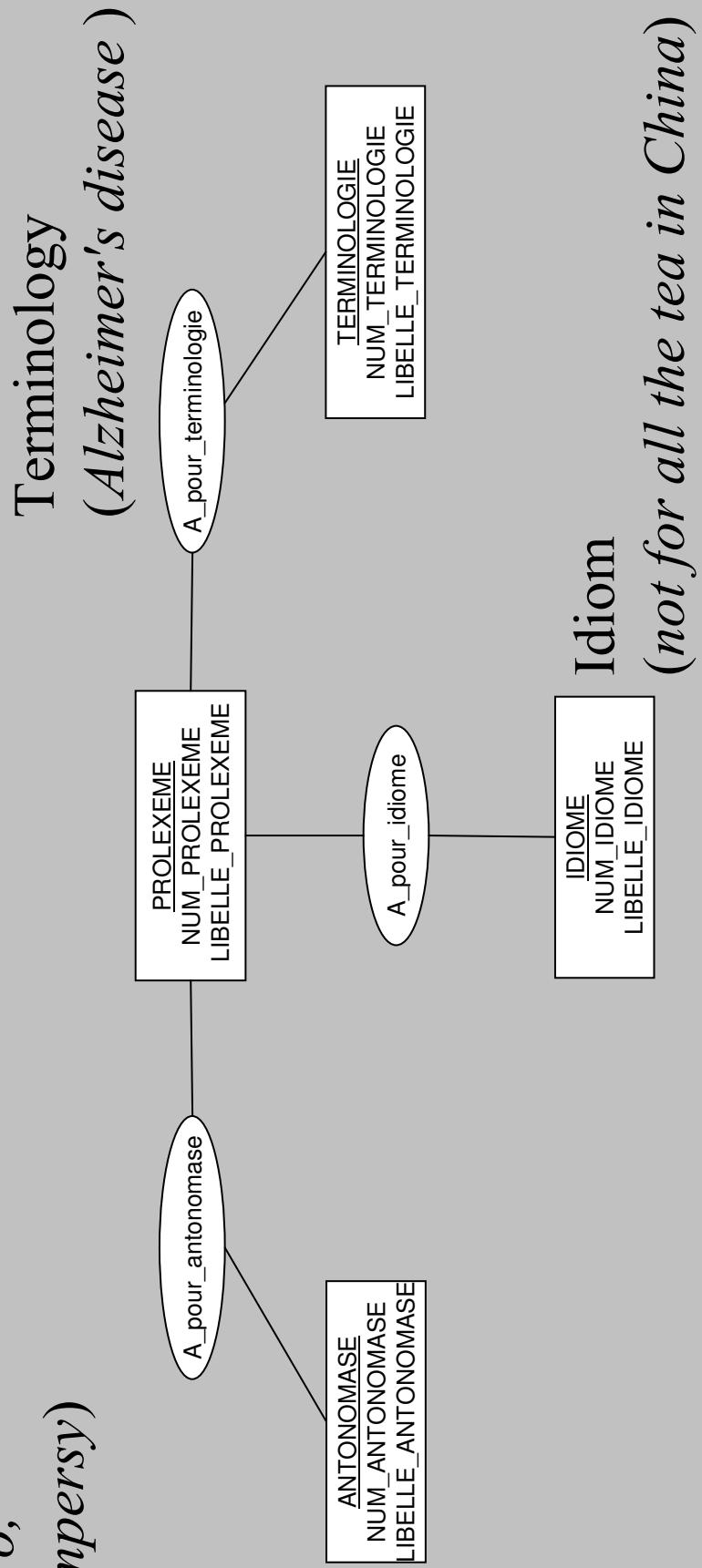
# Further relations





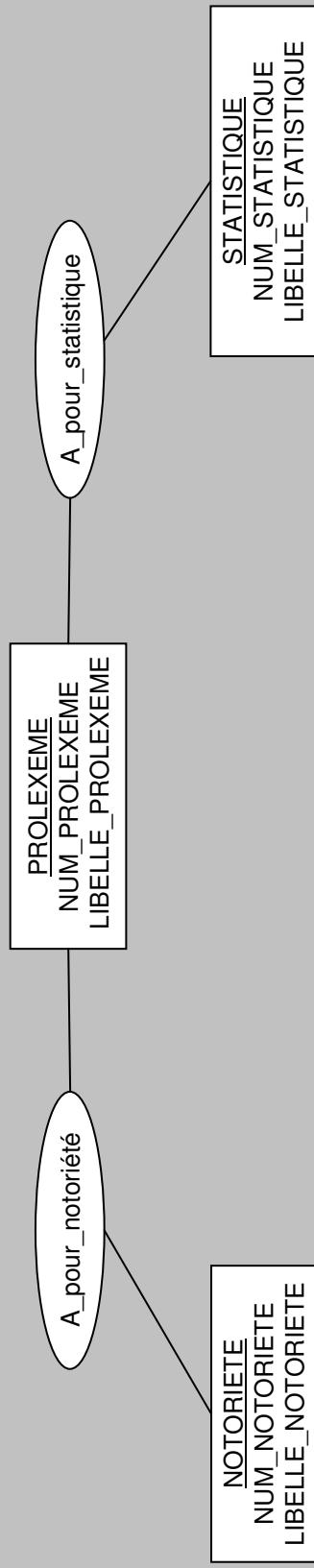
# Eponymy

Antonomasia  
(*biro,*  
*pampersy*)





# Data pertinence



Details,  
National,  
European,  
International

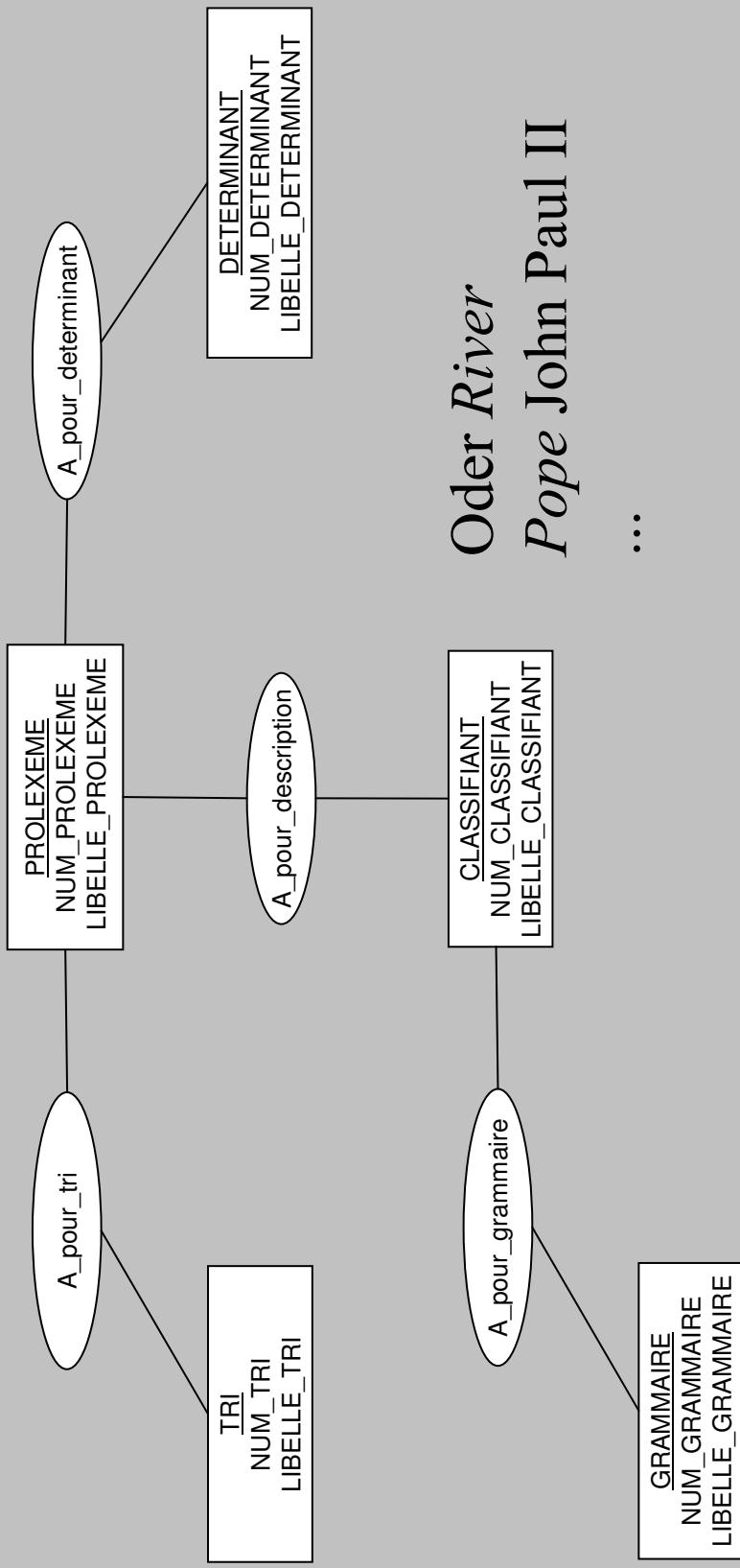
From a corpus study?



# Information

Karol Jozef Woytyla

La Pologne  
Jean-Paul II





## Some numbers...

Lemmas	4 092 anthroponyms	49 751 toponyms	174 ergonyms	222 pragmonyms
Relations	2 249 accesibilities	47 670 meronymies	648 synonymies	
Instances			123 859 instances	

Merci !

