

# On the way to a Relation Registry for ISOcat data categories

Irina Nevskaya – Frankfurt University  
Menzo Windhouwer – MPI for Psycholinguistics

# Outline

- Introduction to
  - ISOcat data categories
  - Relations in ISOcat
- RELISH and ISOcat
- Introduction to
  - Relation Registry
  - RELcat

# A glimpse of ISOcat



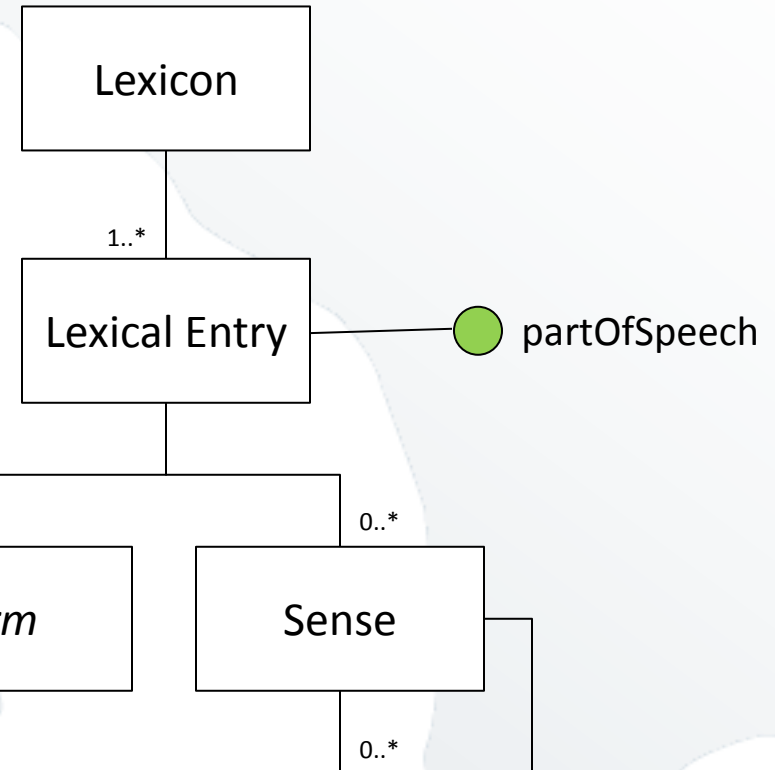
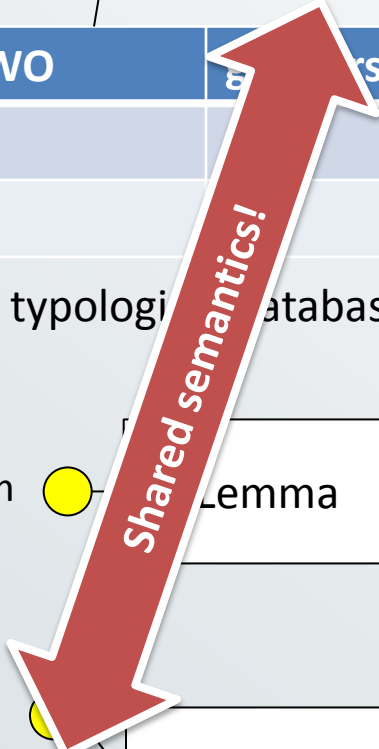
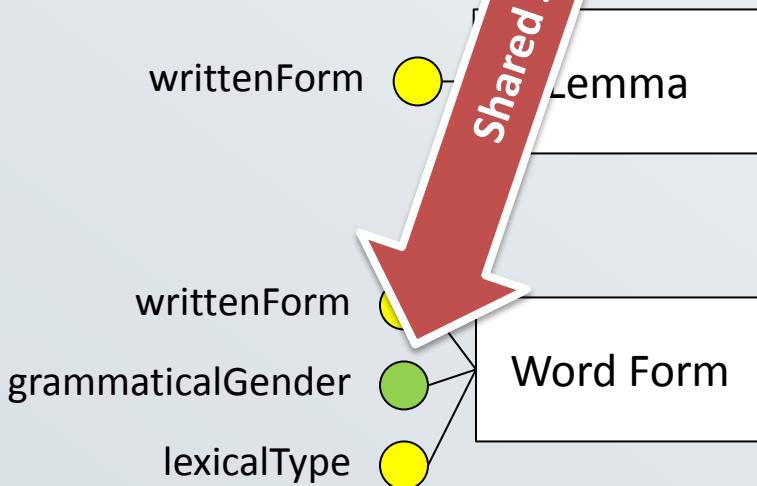
# ISOcat data categories

- ISOcat is a Data Category Registry (DCR)
  - an implementation of ISO 12620:2009
- A data category is
  - the result of the specification of a given data field
  - an elementary descriptor in a linguistic structure or an annotation scheme

# Goal: semantic interoperability

Language	BWO	gram	rs

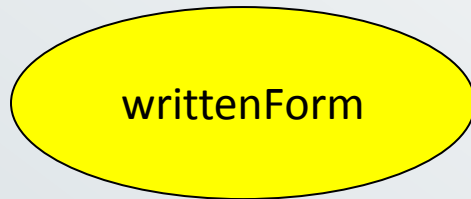
A (schema for a) typological database



A (schema for a) LMF lexicon

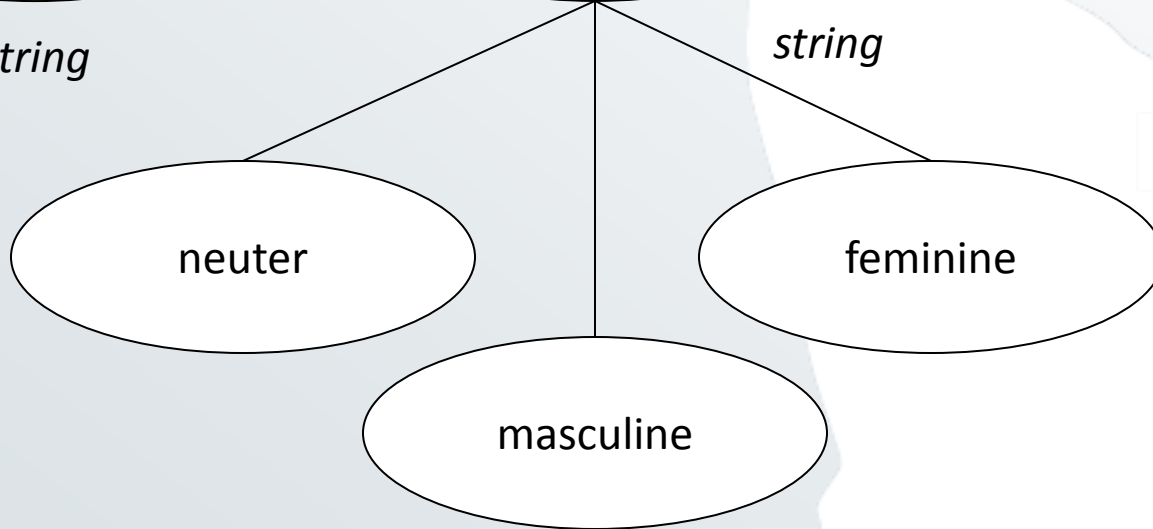
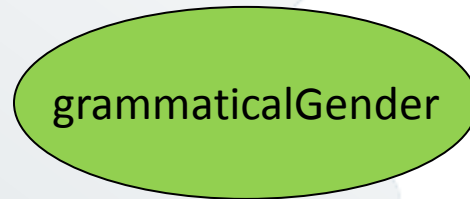
# Data category types

complex: open



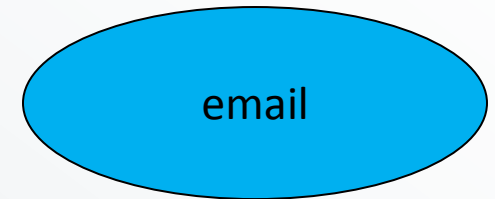
*string*

closed



*string*

constrained



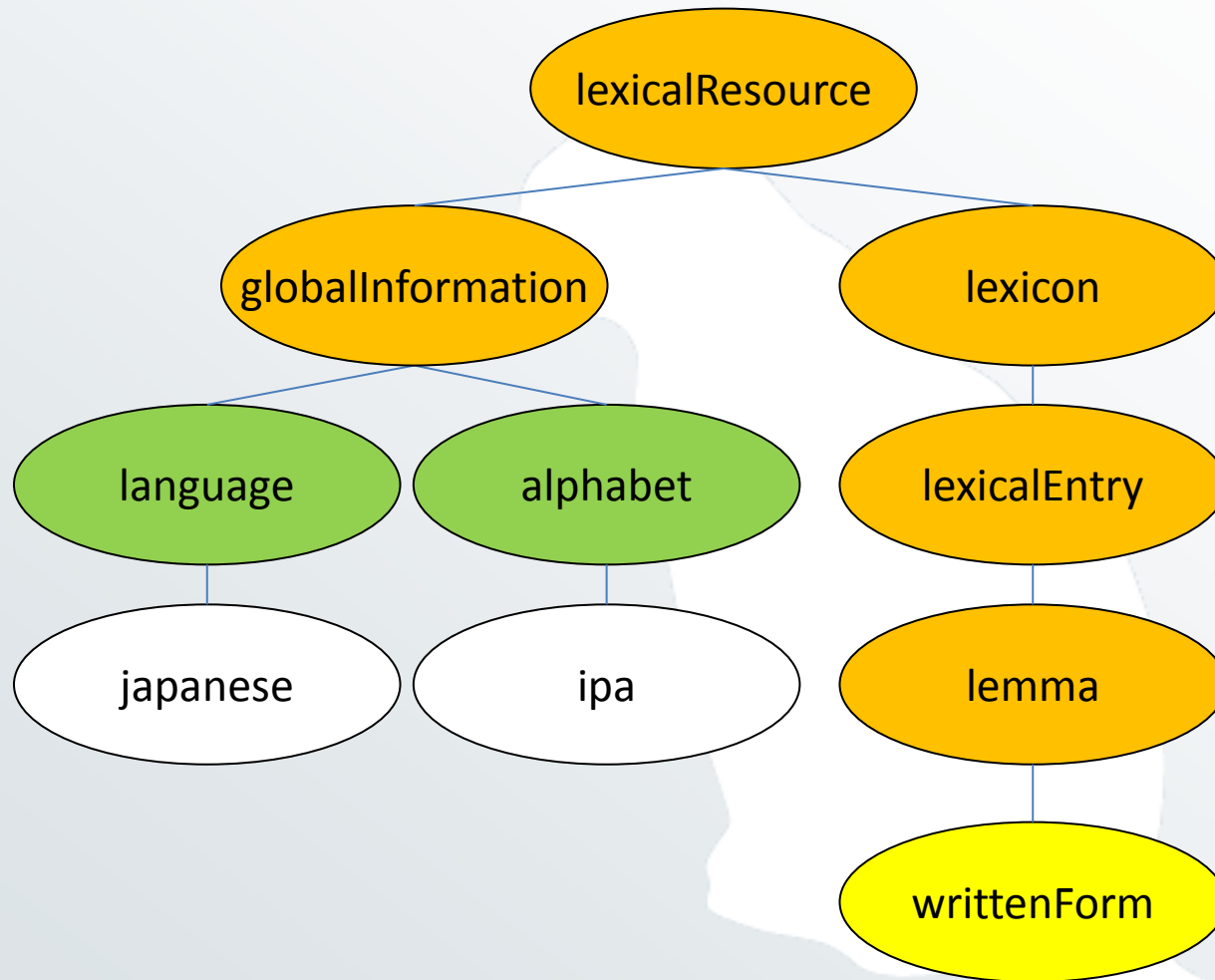
*string*

Constraint: .+@.+

simple:

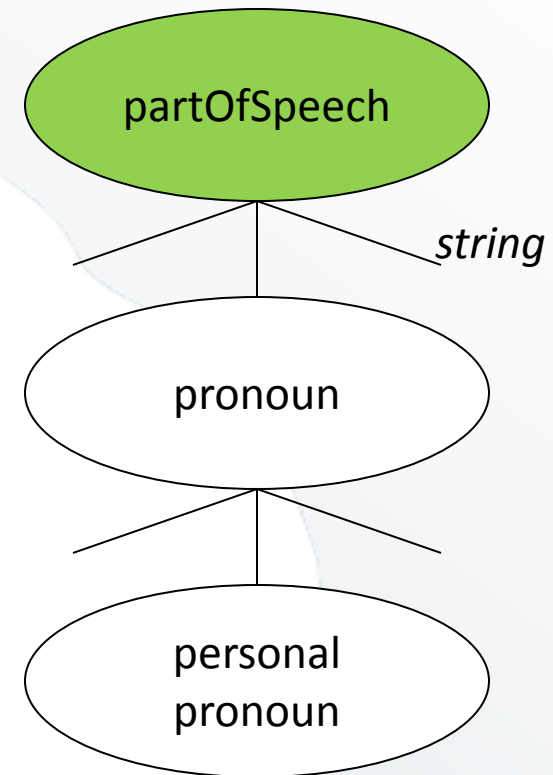
# Data category types

container:



# Data category relationships

- Value domain membership
- Subsumption relationships between simple data categories (legacy)
- Relationships between complex/container data categories are not stored in ISOcat ...
- ... but can be stored in RELcat, a Relation Registry





# No relationships in ISOcat?

- Rationale:
  - Relation types and modeling strategies for a given data category may differ from application to application;
  - Motivation to agree on relation and modeling strategies will be stronger at individual application level;
  - Integration of multiple relation structures in DCR itself could lead to endless ontological clutter.

# RELISH and ISOcat

# Harmonization of Terminology between GOLD and the ISO Data Category Registry in the RELISH project

- The ILIT team created a suitable transformation of the concepts of the GOLD ontology into data categories.
- The team in Frankfurt (FRA) has created an MDF data category selection.
- These data category selections (GOLD and MDF) have been made public.
- By mapping on the ISOcat data categories, a harmonization of the used semantic categories should be achieved.
- A chart with interrelations between the MDF, GOLD and standard ISOcat data categories was created; types of relations between the categories were established to be implemented in the Relation Registry and in mapping the categories in the process of lexicon import into LEXUS as well as into RELISH-LIFT interchange formats.

# Problems encountered by an attempt of mapping

- MDF categories are mostly complex ones. They contain a language property (value):
  - v - vernacular
  - e – English
  - n – national
  - r – regional
- GOLD and „standard“ ISOcats do not contain a language property
- GOLD and „standard“ ISOcats are thought of more as concepts, not as their concrete realizations

# Types of relations between the MDF, GOLD and standard ISOcat data categories

- sameAs: MDF-ISOcat “gloss” (DC-3711) is sameAs the ISOcat “gloss” <http://www.isocat.org/datcat/DC-244>
- almostSameAs: MDF-ISOcat “gloss national” (DC-3711) is almostSameAs the ISOcat “gloss” <http://www.isocat.org/datcat/DC-244>
- partOf: The MDF-ISOcat “Gloss (national)” <http://www.isocat.org/datcat/DC-3711> combines 1) the ISOcat “gloss” <http://www.isocat.org/datcat/DC-244>, 2) the MDF-ISOcat „national language” <http://www.isocat.org/datcat/DC-3702>
- They have, thus the relation partOf the MDF-ISOcat “Gloss (national)”

# Does the name of the category help to find a match?

- The names can be completely different, but the same phenomena are meant:
  - The MDF „Bibliography“ (DC-3687) is sameAs the standard ISOcat “external reference” (DC-1975)
  - The MDF “Borrowed word (loan)” (DC-3688) is sameAs the standard ISOcat “source language” (DC-2494)

# Does the name of the category help to find a match?

- The names can coincide, but different phenomena are meant:
  - The MDF ISO category “citation form (vernacular)”, defined as “a form for representing a lexeme” (DC-3716) should not be confused with the GOLD category “citation”.

The GOLD category is defined as “The action of citing or quoting any words or written passage from a publication that allows others to locate and identify the original source. Typical details include the title, author's name, the journal title (for articles), publication date and page numbers used in research.”

The MDF ISO category is almostSameAs the ISOcat “sort key” (DC-469), defined as “a form for sorting a printed dictionary”.

# Various relations between „standard“ categories in the ISOcat Data Registry

- **No generic** term in the standard set of ISO categories:
  - In the ISOcats registry, there is no generic term “date”, but a number of more specific “date” terms:
    - “origination date”:
      - <http://www.isocat.org/datcat/DC-166>
    - “check date”:
      - <http://www.isocat.org/datcat/DC-126>
    - “creation date”
      - <http://www.isocat.org/datcat/DC-2251>
    - “importation date”:
      - <http://www.isocat.org/datcat/DC-265>
    - “modification date”:
      - <http://www.isocat.org/datcat/DC-365>
    - etc.
  - The MDF category “date” (DC-365) marks both the date of the data creation and the date of the data modification.



## Various relations between „standard“categories in the ISO cat Data Registry

- **Polysemy** of the categories:

MDF-ISOcat “gloss” (DC-3707) is almostSameAs  
the ISOcat “gloss”

<http://www.isocat.org/datcat/DC-244>

only in its first meaning: “In TEI: A phrase or  
word used to provide a gloss or definition for  
some other word or phrase. In 1951: Any  
editorial comment.”

# Various relations between „standard“ categories in the ISOcat Data Registry

- **Synonymy** of the categories:
  - The ISOcat “etymology”  
<http://www.isocat.org/datcat/DC-221>  
is **sameAs** to the ISOcat “etymological root”  
<http://www.isocat.org/datcat/DC-1987>
  - The ISOcat „comment“  
<http://www.isocat.org/datcat/DC-1846>  
is **sameAs** to the ISOcat “note”  
<http://www.isocat.org/datcat/DC-382>

# RELcat a Relation Registry



# Relation Registry

- Stores typed relationships between
  - data categories: ISOcat, Dublin Core, ...
  - concepts: GOLD, ...
- Sets of relationships
  - are owned by individuals or groups
    - no standardization effort
  - can be combined
    - might need conflict resolution
  - can use their own vocabulary
    - hookup into a core taxonomy of relation types
- Exploitation:
  - Various levels of semantic search
    - CMDI metadata search [CMDRSB](#)

# RELcat

- An implementation of a Relation Registry
- Based on a RDF quad store, SPARQL (and RDFS Plus)
- Status: read-only backend in alpha stage
  - Upload of relations by admin
  - Available sets
    - Metadata: CMDI and Dublin Core
      - <http://lux13.mpi.nl/isocat/relcat/set/cmdl>
      - <http://lux13.mpi.nl/isocat/relcat/set/dc>
    - RELISH: RELISH and GOLD
      - <http://lux13.mpi.nl/isocat/relcat/set/relish>
      - <http://lux13.mpi.nl/isocat/relcat/set/gold>
  - Representations: .rdf .trig .svg .png ...
  - Queries on multiple sets are supported
  - Misses: UI, reasoner, handling of same-as cliques, ...

# Relationship types

- rel:related

- rel:sameAs (symmetric and transitive)
- rel:almostSameAs (symmetric)
- rel:narrower (inverse of rel:broader)
  - rel:superClassOf (inverse of rel:subClassOf)
- rel:broader (inverse of rel:narrower)
  - rel:subClassOf (inverse of rel:subClassOf)
  - rel:partOf
    - rel:directPartOf
    - rel:indirectPartOf

- rel:sameAs
  - owl:sameAs
  - owl:equivalentClass
  - owl:equivalentProperty
  - ...

<http://lux13.mpi.nl/isocat/relcat/relations>

- Generic algorithms can use the upper part of the taxonomy and combine arbitrary graphs
  - <http://lux13.mpi.nl/isocat/query/combine?set=relish&set=gold>
- Specific vocabularies (SKOS, OWL) should put (some) predicates in the proper place in the taxonomy
  - Specific algorithms (OWL reasoners) can use the original set or combinations of sets using the same vocabulary

# A fragment of the relation chart

Gloss (E)	1) Yes in the standard <u>ISOcats</u> or GOLD  2) Yes in the <u>MDF-ISOcats</u>	Is equal to  Is equal to	<u>The ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>  The <u>MDF-ISOcat "Gloss (English)"</u> <a href="http://www.isocat.org/datcat/DC-3707">http://www.isocat.org/datcat/DC-3707</a>	Only in the first meaning of "gloss" in <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>
Gloss (n)	1) Yes in the standard <u>ISOcats</u> or GOLD or <b>No</b> in the standard <u>ISOcats</u> or GOLD  2) Yes in the <u>MDF-ISOcats</u>	Roughly corresponds to  Combines  Is equal to	<u>The ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>  1) The <u>ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a> 2) the <u>MDF-ISOcat "national language"</u> <a href="http://www.isocat.org/datcat/DC-3702">http://www.isocat.org/datcat/DC-3702</a>  The <u>MDF-ISOcat "Gloss (national)"</u> <a href="http://www.isocat.org/datcat/DC-3711">http://www.isocat.org/datcat/DC-3711</a>	Only in the first meaning of "gloss" in <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>
Gloss (r)	1) Yes in the standard <u>ISOcats</u> or GOLD or <b>No</b> in the standard <u>ISOcats</u> or GOLD  2) Yes in the <u>MDF-ISOcats</u>	Roughly corresponds to  Combines  Is equal to	<u>The ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>  1) The <u>ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a> 2) the <u>MDF-ISOcat "regional language"</u> <a href="http://www.isocat.org/datcat/DC-3703">http://www.isocat.org/datcat/DC-3703</a>  The <u>MDF-ISOcat "Gloss (regional)"</u> <a href="http://www.isocat.org/datcat/DC-3712">http://www.isocat.org/datcat/DC-3712</a>	Only in the first meaning of "gloss" in <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>
Gloss (v)	1) Yes in the standard <u>ISOcats</u> or GOLD or <b>No</b> in the standard <u>ISOcats</u> or GOLD  2) Yes in the <u>MDF-ISOcats</u>	Roughly corresponds to  Combines  Is equal to	<u>The ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>  1) The <u>ISOcat "gloss"</u> <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a> 2) the <u>MDF-ISOcat "vernacular language"</u> <a href="http://www.isocat.org/datcat/DC-3706">http://www.isocat.org/datcat/DC-3706</a> The <u>MDF-ISOcat "Gloss (vernacular)"</u> <a href="http://www.isocat.org/datcat/DC-3713">http://www.isocat.org/datcat/DC-3713</a>	Only in the first meaning of "gloss" in <a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>

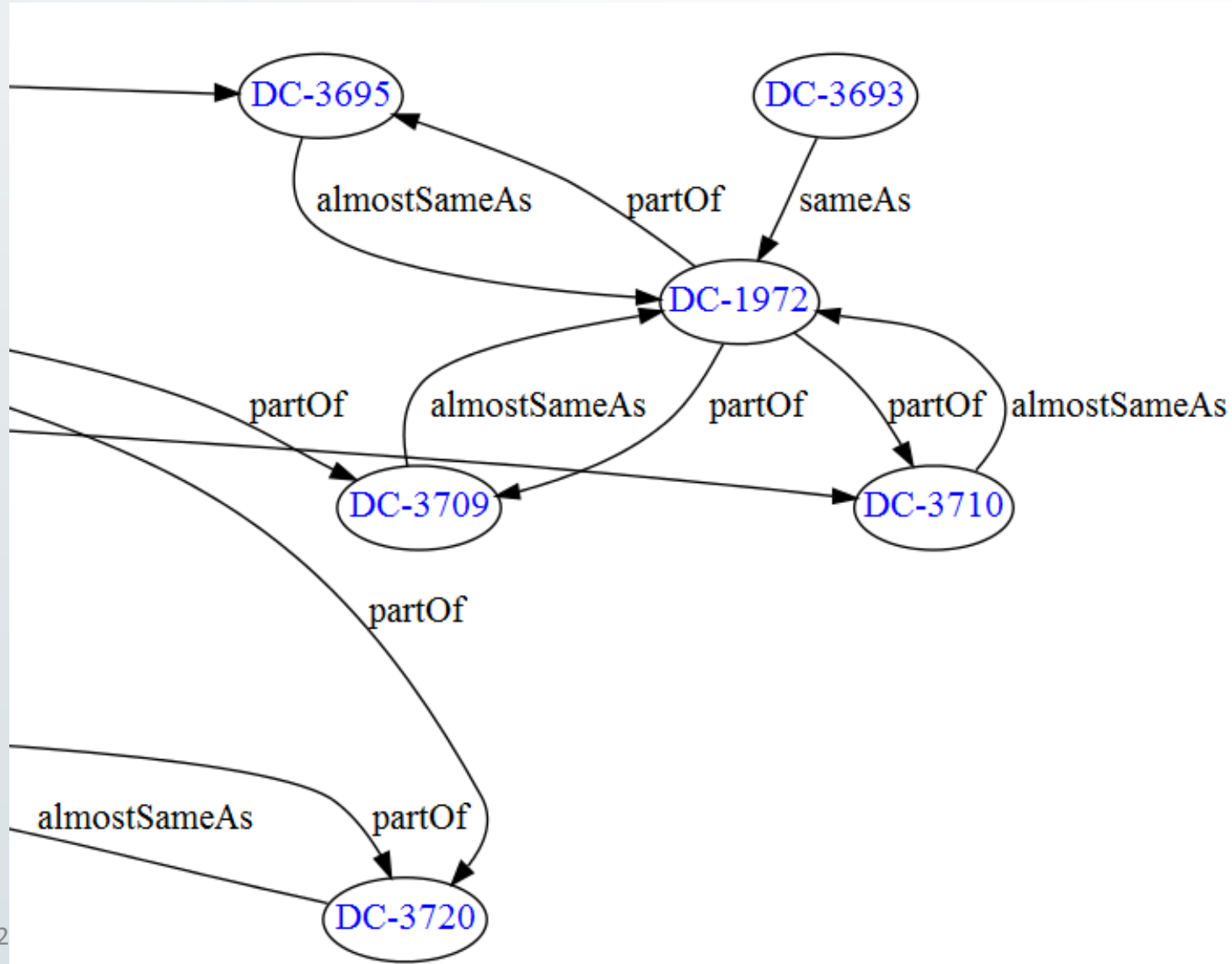


# Types of semantic relations

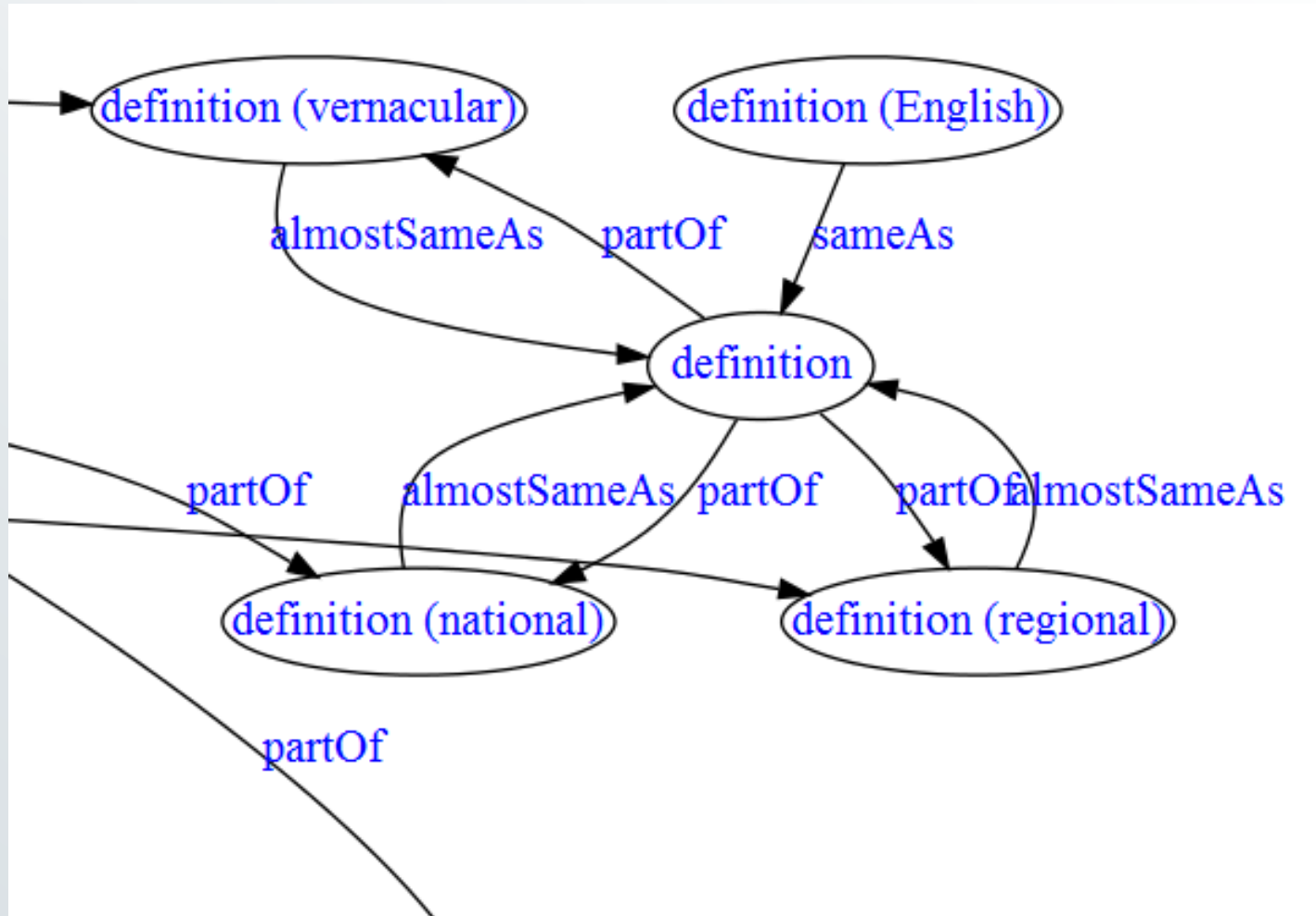
20	<a href="http://www.isocat.org/datcat/DC-3702">http://www.isocat.org/datcat/DC-3702</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3691">http://www.isocat.org/datcat/DC-3691</a>
21	<a href="http://www.isocat.org/datcat/DC-3692">http://www.isocat.org/datcat/DC-3692</a>	related	<a href="http://www.isocat.org/datcat/DC-164">http://www.isocat.org/datcat/DC-164</a>
22	<a href="http://www.isocat.org/datcat/DC-3692">http://www.isocat.org/datcat/DC-3692</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>
23	<a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3692">http://www.isocat.org/datcat/DC-3692</a>
24	<a href="http://www.isocat.org/datcat/DC-216">http://www.isocat.org/datcat/DC-216</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3692">http://www.isocat.org/datcat/DC-3692</a>
25	<a href="http://www.isocat.org/datcat/DC-3702">http://www.isocat.org/datcat/DC-3702</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3692">http://www.isocat.org/datcat/DC-3692</a>
26	<a href="http://www.isocat.org/datcat/DC-3693">http://www.isocat.org/datcat/DC-3693</a>	sameAs	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>
27	<a href="http://www.isocat.org/datcat/DC-3709">http://www.isocat.org/datcat/DC-3709</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>
28	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3709">http://www.isocat.org/datcat/DC-3709</a>
29	<a href="http://www.isocat.org/datcat/DC-3702">http://www.isocat.org/datcat/DC-3702</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3709">http://www.isocat.org/datcat/DC-3709</a>
30	<a href="http://www.isocat.org/datcat/DC-3710">http://www.isocat.org/datcat/DC-3710</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>
31	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3710">http://www.isocat.org/datcat/DC-3710</a>
32	<a href="http://www.isocat.org/datcat/DC-3703">http://www.isocat.org/datcat/DC-3703</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3710">http://www.isocat.org/datcat/DC-3710</a>
33	<a href="http://www.isocat.org/datcat/DC-3694">http://www.isocat.org/datcat/DC-3694</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-365">http://www.isocat.org/datcat/DC-365</a>
34	<a href="http://www.isocat.org/datcat/DC-3695">http://www.isocat.org/datcat/DC-3695</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>
35	<a href="http://www.isocat.org/datcat/DC-1972">http://www.isocat.org/datcat/DC-1972</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3695">http://www.isocat.org/datcat/DC-3695</a>
36	<a href="http://www.isocat.org/datcat/DC-3706">http://www.isocat.org/datcat/DC-3706</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3695">http://www.isocat.org/datcat/DC-3695</a>
37	<a href="http://www.isocat.org/datcat/DC-3696">http://www.isocat.org/datcat/DC-3696</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1846">http://www.isocat.org/datcat/DC-1846</a>
38	<a href="http://www.isocat.org/datcat/DC-1846">http://www.isocat.org/datcat/DC-1846</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3696">http://www.isocat.org/datcat/DC-3696</a>
39	<a href="http://www.isocat.org/datcat/DC-221">http://www.isocat.org/datcat/DC-221</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3696">http://www.isocat.org/datcat/DC-3696</a>
40	<a href="http://www.isocat.org/datcat/DC-3697">http://www.isocat.org/datcat/DC-3697</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1846">http://www.isocat.org/datcat/DC-1846</a>
41	<a href="http://www.isocat.org/datcat/DC-3698">http://www.isocat.org/datcat/DC-3698</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-244">http://www.isocat.org/datcat/DC-244</a>
42	<a href="http://www.isocat.org/datcat/DC-1846">http://www.isocat.org/datcat/DC-1846</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3698">http://www.isocat.org/datcat/DC-3698</a>
43	<a href="http://www.isocat.org/datcat/DC-3702">http://www.isocat.org/datcat/DC-3702</a>	partOf	<a href="http://www.isocat.org/datcat/DC-3698">http://www.isocat.org/datcat/DC-3698</a>
44	<a href="http://www.isocat.org/datcat/DC-3700">http://www.isocat.org/datcat/DC-3700</a>	almostSameAs	<a href="http://www.isocat.org/datcat/DC-1846">http://www.isocat.org/datcat/DC-1846</a>



# A Fragment of the RELcat



# A Fragment of the RELcat



# Thank you for your attention!

Visit and use: [www.isocat.org](http://www.isocat.org)  
Track and debug: [lux13.mpi.nl/relcat/](http://lux13.mpi.nl/relcat/)

Questions or feedback?  
[nevskaya@em.uni-frankfurt.de](mailto:nevskaya@em.uni-frankfurt.de)  
[Menzo.Windhouver@mpi.nl](mailto:Menzo.Windhouver@mpi.nl)