

“The Similarity Jury: Combining expert judgements on geographic concepts”



Andrea Ballatore

Michela Bertolotto

University College Dublin

David C. Wilson

University of North Carolina, Charlotte

Strategic Research in
Advanced Geo-Technologies
stratag.ie



Science Foundation Ireland
sfi.ie



Why similarity?



When grouping individuals, we see
“a complicated network of similarities
overlapping and criss-crossing”
(Wittgenstein, 1956)



“The ability to perceive similarities and
analogies is one of the most fundamental
aspects of human cognition”
(Vosniadou & Ortony, 1989)

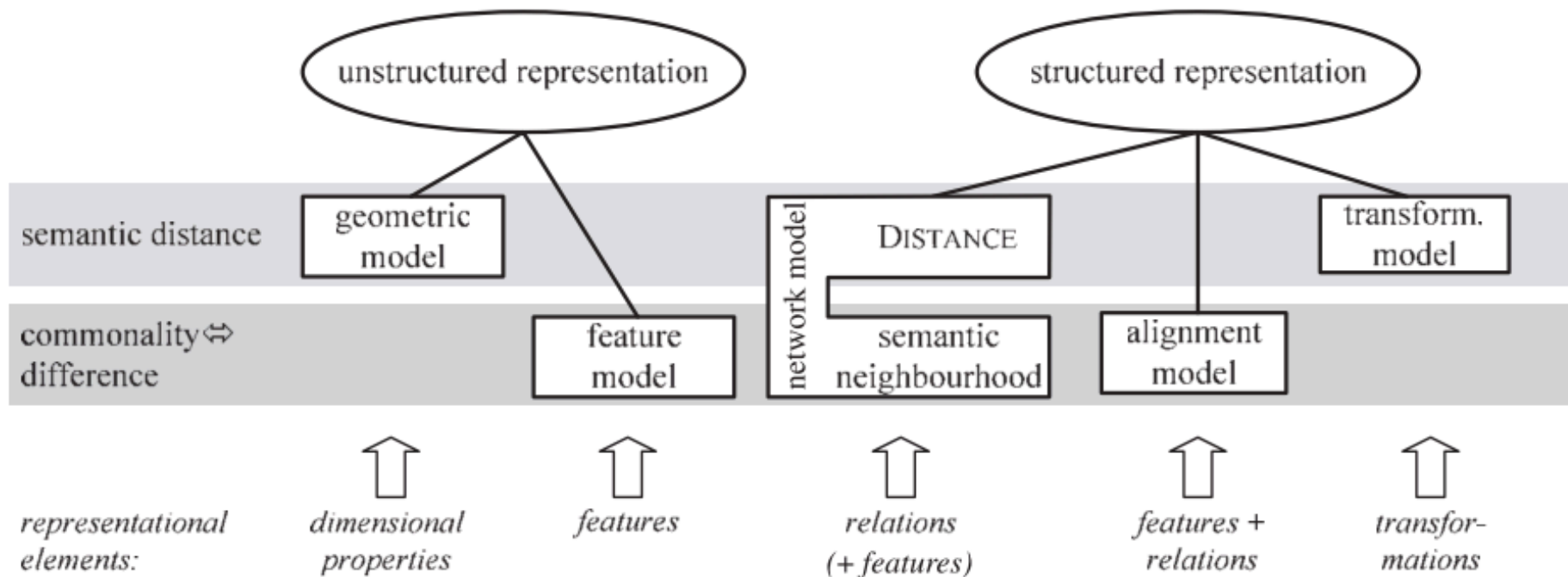


Figure 9 Semantic similarity measures are based on different notions of similarity

(Schwering 2008)

Geo-similarity measures

- Variety of approaches:
MDSM (Rodriguez); **SIM-DL** (Janowicz)
- Impact of **context** (Keßler 2009)
- Different performance in different formalisms and datasets
- Cognitive plausibility, imitation of human behaviour (Janowicz et al. 2008)
- **Unstable ground truth**

Golden calf?



Expert disagreement

- Uncertain, complex domains, trade-offs
- Epistemic, cultural, cognitive, ideological bias
- “The history of scholarship is a record of disagreements” (Hughes, 1936)
- Medicine, psychology, economics
- ‘Best’ expert? Gold standards?



© Original Artist

Reproduction rights obtainable from
www.CartoonStock.com

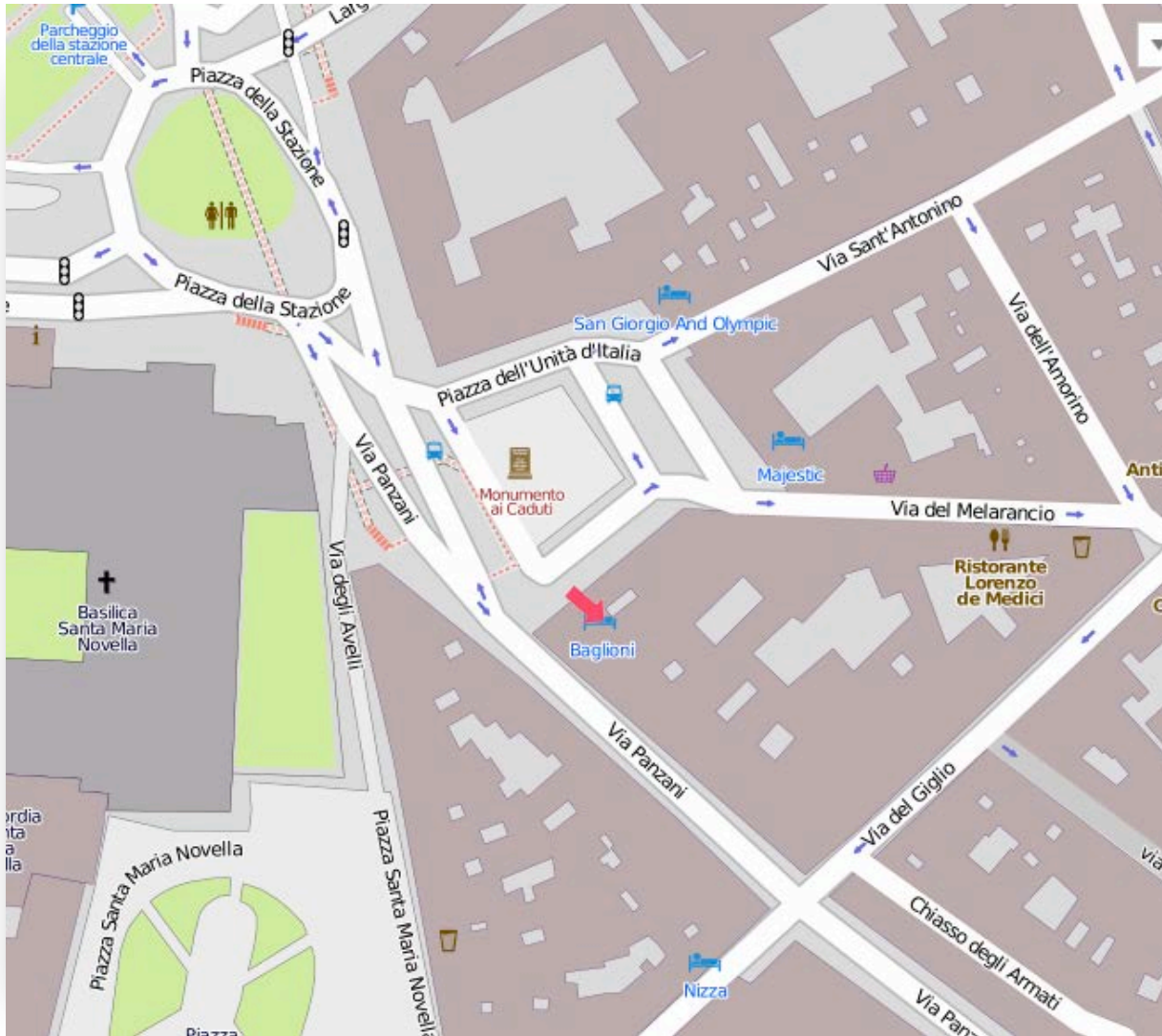


Overcoming expert disagreement



- Combine expert judgements into a **representative average**
- Analogy of the jury of experts

Similarity measure = human expert



OpenStreetMap

*"The Free Wiki
World Map"*



Semantic similarity of OSM concepts

- OSM semantics: **semi-structured folksonomy**
- Open set of tags (concepts)
- Concept definitions on OSM Wiki website:
<http://wiki.openstreetmap.org>
- VGI: ambiguity, noise, semantic gap
- $sim(tagA, tagB) = \text{real number in } [0,1]$



8 WordNet-based semantic similarity

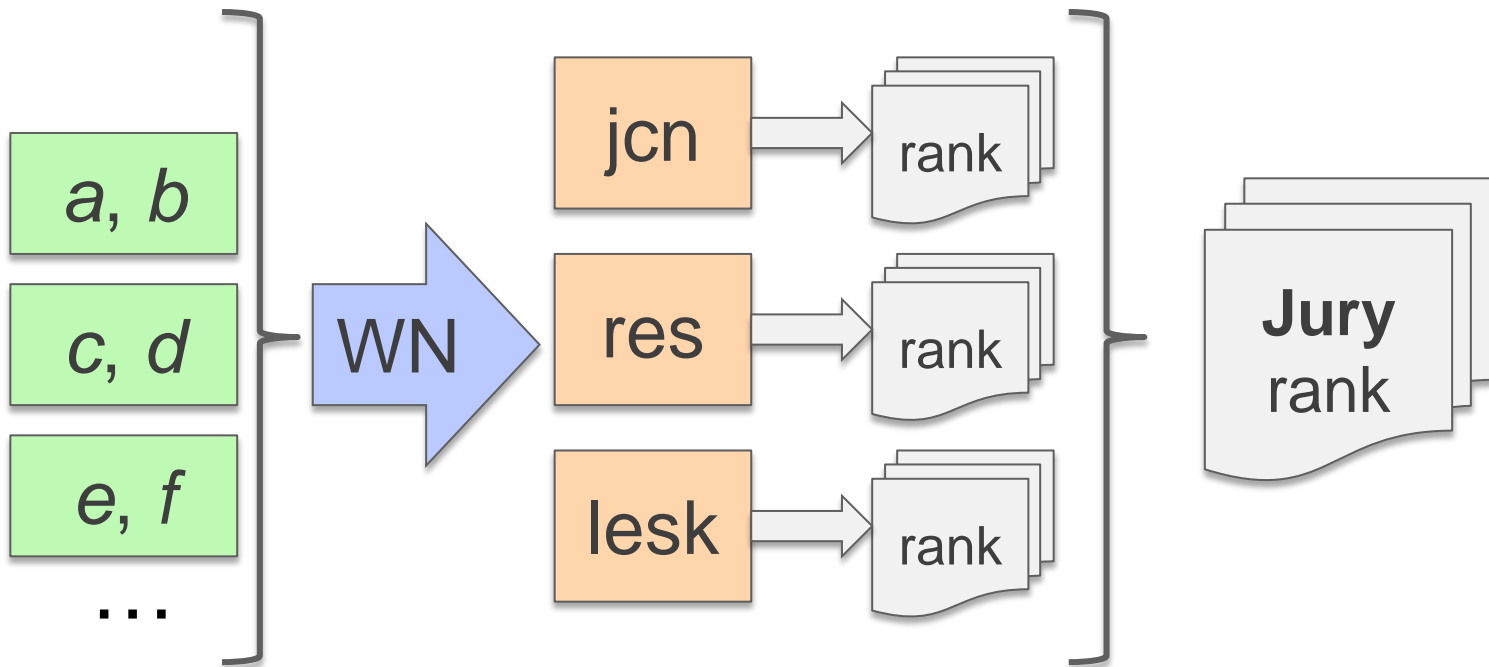
Name	Authors	Description	SPath	Gloss	InfoC
path	Rada et al. [21]	Edge count in the semantic network	✓		
lch	Leacock and Chodorow [15]	Edge count scaled by depth	✓		
res	Resnik [23]	Information content of <i>lcs</i>	✓		✓
jcn	Jiang and Conrath [14]	Information content of <i>lcs</i> and terms	✓		✓
lin	Lin [16]	Ratio of information content of <i>lcs</i> and terms	✓		✓
wup	Wu and Palmer [26]	Edge count between <i>lcs</i> and terms	✓		
lesk	Banerjee and Pedersen [1]	Extended gloss overlap		✓	
vector	Patwardhan and Pedersen [19]	Second order co-occurrence vectors		✓	

Similarity jury



- Juries of 2,3, and 4 members
- **Average of similarity rankings** vs individual members
- **Cognitive plausibility** against human-generated dataset

Semantic similarity of OSM terms



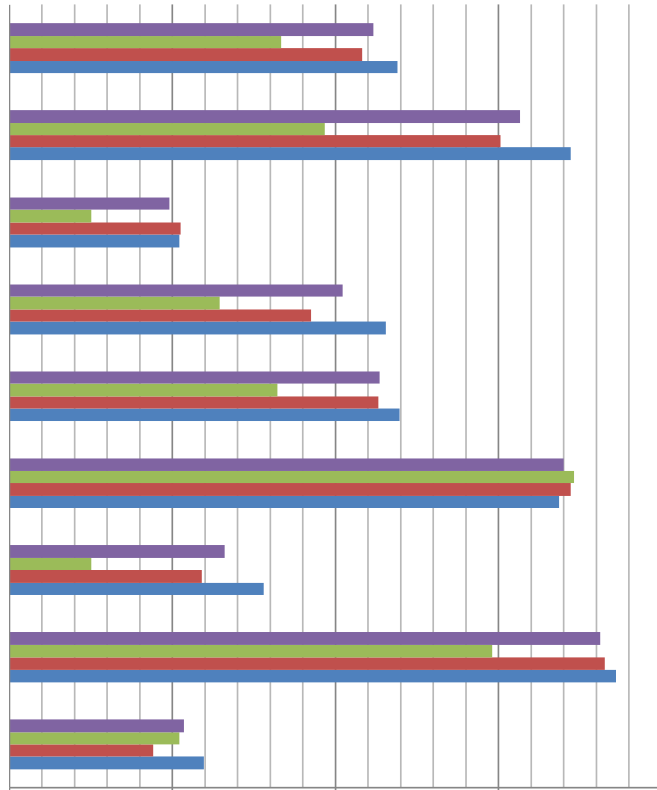
- Pairs of concepts sorted by descending similarity
- E.g. <river,stream>, <river,lake> ... <river,restaurant>

Similarity jury



- **Total success:** jury $>$ all of its members
e.g. $\text{corr}(J(\text{res}, \text{jcn}), H) > \text{corr}(\text{res}, H) \ \& \ \text{corr}(\text{jcn}, H)$
- **Partial success:** jury $>$ individual measure
e.g. $\text{corr}(J(\text{res}, \text{jcn}), H) > \text{corr}(\text{res}, H)$

Evaluation



Results

- The jury is generally more cognitively plausible than individual measures (mn **partial success ratio = 72%**)
- The jury is generally less cognitively plausible than the best of its members (mn **total success ratio = 35%**)
- The jury is higher than the mean of members (**93%**)
- The jury can outperform even the most plausible measures

Conclusions

- Policy when no gold standard is available: **rely on a jury** rather than on an arbitrary measure
- “A group of experts tends to perform better than the average solitary expert, but the best individual in the group often outperforms the group as a whole.” (Cooke and Goossens, 2004)
- Generalisable to non-geographic domains?
- Experts-Should-Converge hypothesis (Shanteau 2001) for geo-semantic similarity?

OSM Semantic Network



The **OpenStreetMap Semantic Network** is an RDF graph extracted from the OSM Wiki website, encoded as a [SKOS vocabulary](#). It contains terms utilised in OSM (mainly tags and keys), and several semantic relations between them. This semantic network is extracted by an open source crawler developed in Java/Groovy, the OSM Wiki Crawler.

The network can be used to compute the **semantic similarity** of OSM tags, and can be explored through a [web interface](#) similar to DBpedia.

Contents [\[show\]](#)

Dataset

The dataset is available in the following ways:

- Linked Open Data (LOD): <http://spatial.ucd.ie/od/osn>
- Static RDF file: osm_semantic_network.skos.rdf
- SPARQL end point: <http://spatial.ucd.ie/od/sparql>
- RDF/NT dumps (archive): <http://github.com/ucd-spatial/OsmSemanticNetwork>



OSM Tag: **waterway=canal**

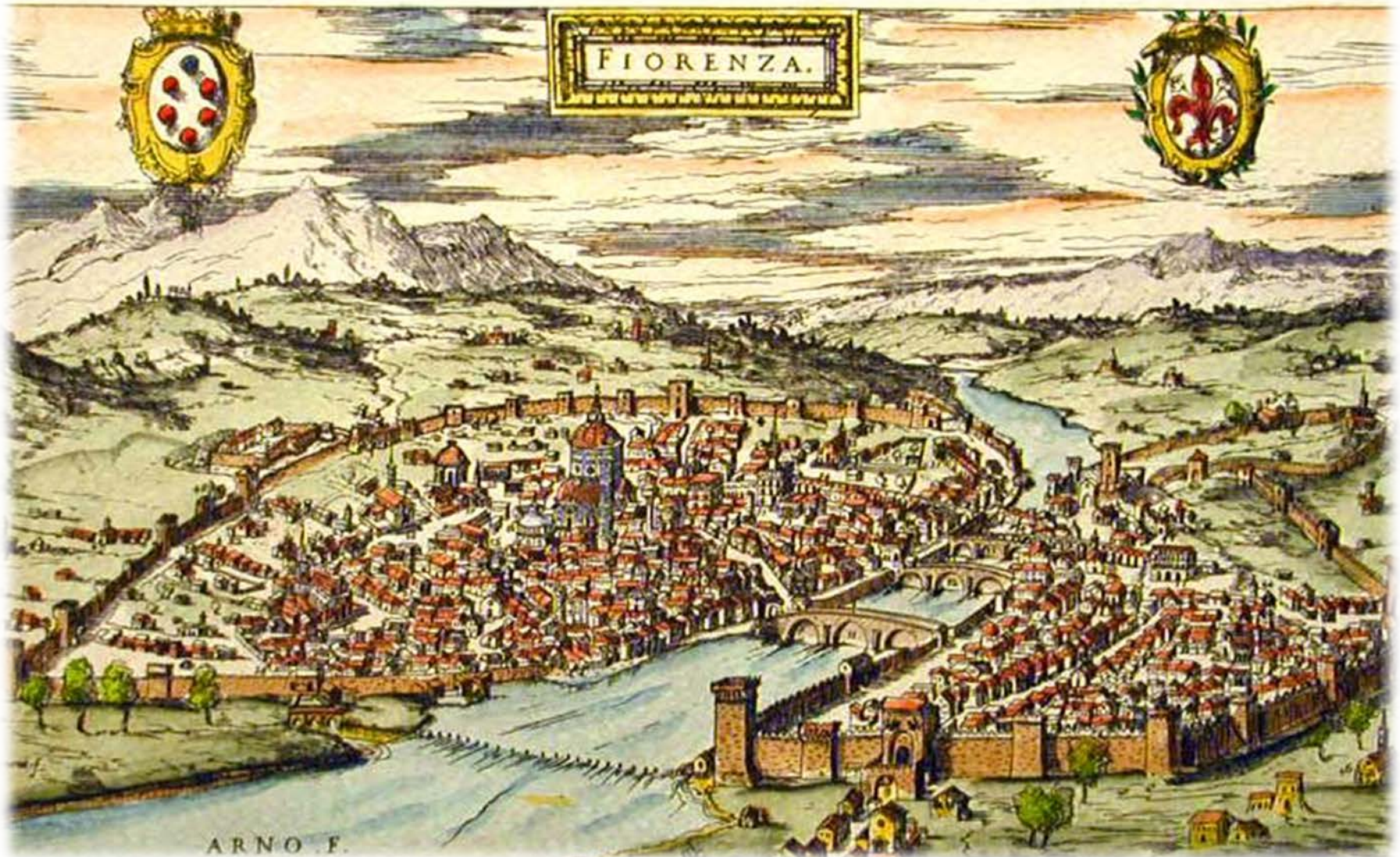
<http://spatial.ucd.ie/lod/osn/term/k:waterway/v:canal>

waterway = canal at OSM Semantic Network

<http://spatial.ucd.ie/lod/osn/term/k:waterway/v:canal>



Property	Value
skos:altLabel	<ul style="list-style-type: none">▪ (waterway) canal (en)▪ waterway#canal (en)▪ waterway, canal (en)▪ waterway=canal (en)
skos:broader	<ul style="list-style-type: none">▪ osn:term/k:waterway
skos:definition	<ul style="list-style-type: none">▪ An artificial open waterway used for transportation, waterpower, or irrigation. An artificial open waterway used for transportation, waterpower, or irrigation. (en)▪ Uma hidrovía artificial aberto usado para transporte, energia hidráulica, ou irrigação. (pt-br)▪ Un cours d'eau artificiel utilisé pour le transport, l'irrigation ou l'hydroélectricité. (fr)▪ Канал для судоходства, орошения или гидроэнергетики. (ru)▪ 輸送、用水、灌漑などに使われる、人工の（暗渠でない）水路です。 (ja)
skos:exactMatch	<ul style="list-style-type: none">▪ lgv:Canal



andrea.ballatore@ucd.ie

