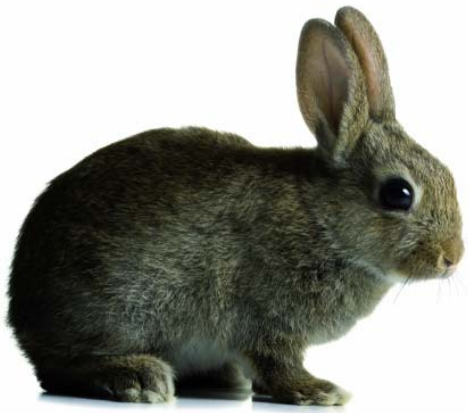
Rabbit: Developing a controlled natural language for authoring ontologies

Glen Hart, Martina Johnson and Catherine Dolbear

Overview

- Motivation for Controlled Natural Language (CNL) research at Ordnance Survey
- Rabbit CNL
  - Design considerations
  - Human subject testing



Ordnance Survey – who we are

- National Mapping Agency of Great Britain
- Data vendor: one of the largest geospatial databases in the world
- Customers use GIS systems & spatially enabled databases to process data
- Building ontologies to describe our data & make reuse and integration easier

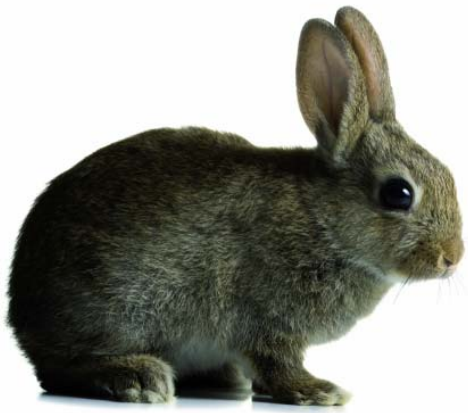
Current issues with ontology authoring

- Domain experts find ontology authoring difficult in terms of:
  - Understanding meaning of OWL-DL language
  - Good practice modelling techniques

- Controlled natural language required for:
    - Domain experts to author ontologies
    - Domain experts to validate ontologies for reuse
    - Documentation for OWL version.

Overview

- Motivation for Controlled Natural Language (CNL) research at Ordnance Survey
- Rabbit CNL
  - Design considerations
  - Human subject testing



CNLs in the literature

- Other CNLs available: CLOnE, ACE, PENG (SOS)
- OWL WG Task Force to develop a controlled natural language syntax
- Rabbit developed out of modelling ontologies, rather than driven by linguistics or Description Logics

Building ontologies

Rabbit

OWL

Conceptual Ontology

Knowledge represented in a form understandable to people

Logical Ontology

Knowledge represented in a form manipulable by computers

Every River Stretch is part of a River → River_Stretch ⊏ direct_part_of ∃ River →

```
<owl:ObjectProperty rdf:ID="directPartOf">
    <rdf:type rdf:resource="&owl;FunctionalProp
</owl:ObjectProperty>
<owl:Class rdf:ID="River"/>
<owl:Class rdf:ID="RiverStretch">
            <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#directP
      <owl:someValuesFrom rdf:resource="#R
    </owl:Restriction>
  </rdfs:subClassOf>
```

Rabbit – design principles

- Allow the domain expert, supported by a knowledge engineer and software, to express their knowledge:
  - Easily and simply
  - In as much detail as necessary.
- Formal grammar so that the subset which can be expressed as OWL can be systematically translated.
- Understandable by domain experts with little or no knowledge of Rabbit.
- Independent of any specific domain.

Rabbit – design

- Simple, ‘natural’ expression
- Short statements preferred
- Balance between natural expression and ambiguity
- Avoid extra input for the author!
- Driver is support of domain experts’ expression, not representing DL

Rabbit – translating to OWL

Structure	Rabbit	OWL
Concept Definition	Bank (River) is a concept.	Class: Bank_River rdf:label “Bank”
Instance Declaration	England is a Country.	Individual: england Types: Country
Subsumption	Every Bourne is a kind of Stream.	Bourne subClassOf Stream
Existential Quantifier	Every River flows into a Sea.	River -> flowsInto Some Sea
Universal Quantifier	Every Channel is connected to a Pool, Lake or nothing	Channel -> isConnectedTo Only ((Pool or Lake) and not (Pool and Lake))
Defined Class definition	A Source is defined as: Every Source is a kind of Spring or Wetland; Every Source feeds a River or a Stream.	Source = Spring or Wetland and feeds Some (River or Stream)
Disjoints	Field and Lake are mutually exclusive	Field DisjointWith: Lake

Rabbit – translating to OWL

Structure	Rabbit	OWL
Qualified Cardinality Restriction	Every Braided River Stretch flows in at least 2 Channels.	BraidedRiverStretch -> flowsIn min 2 Channel
Union	Every Floodplain is adjacent to 1 or more of a River or a Stream.	Floodplain -> adjacentTo Some (River or Stream)
Covering Axiom	Every Mill Stream only flows in a Mill Race	MillStream -> flowsIn some MillRace and flowsIn only MillRace
General Concept Inclusion Axiom	Everything that has a part, that contains some Water, will also contain some Water	hasPart Some (contains Some Water) -> contains Some Water
Concept roles/ functions/ purposes	Every School has purpose Education of Children.	School -> hasPurpose some (Education and of some Child)

Rabbit – reuse of ontologies

- Experimental structures needed for research on merging ontologies

  “Use references:

  OSPlaces from http://www.ordnancesurvey.co.uk/ontology/Places.rbt.”

- Referencing with extension – can add additional axioms to the foreign concept
- Referencing with restriction – all axioms from foreign ontology’s concept ignored

  “Reference Pub as Secondary Concept from OSPlaces”.

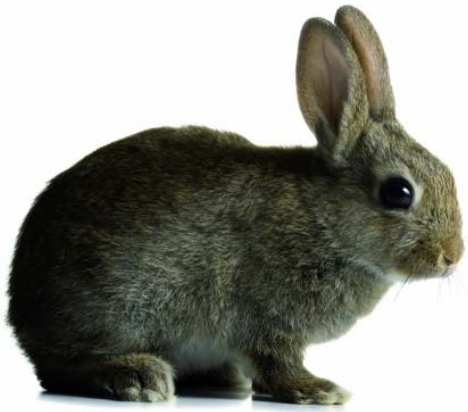
ROO : a tool for authoring Rabbit

- Joint project with the University of Leeds to produce a Protégé plugin
  - Follows the OS ontology authoring method
  - Authors ontologies in Rabbit
  - Direct translation to OWL using GATE

- Poster at this conference

Overview

- Motivation for Controlled Natural Language (CNL) research at Ordnance Survey
- Rabbit CNL
  - Design considerations
  - Human subject testing

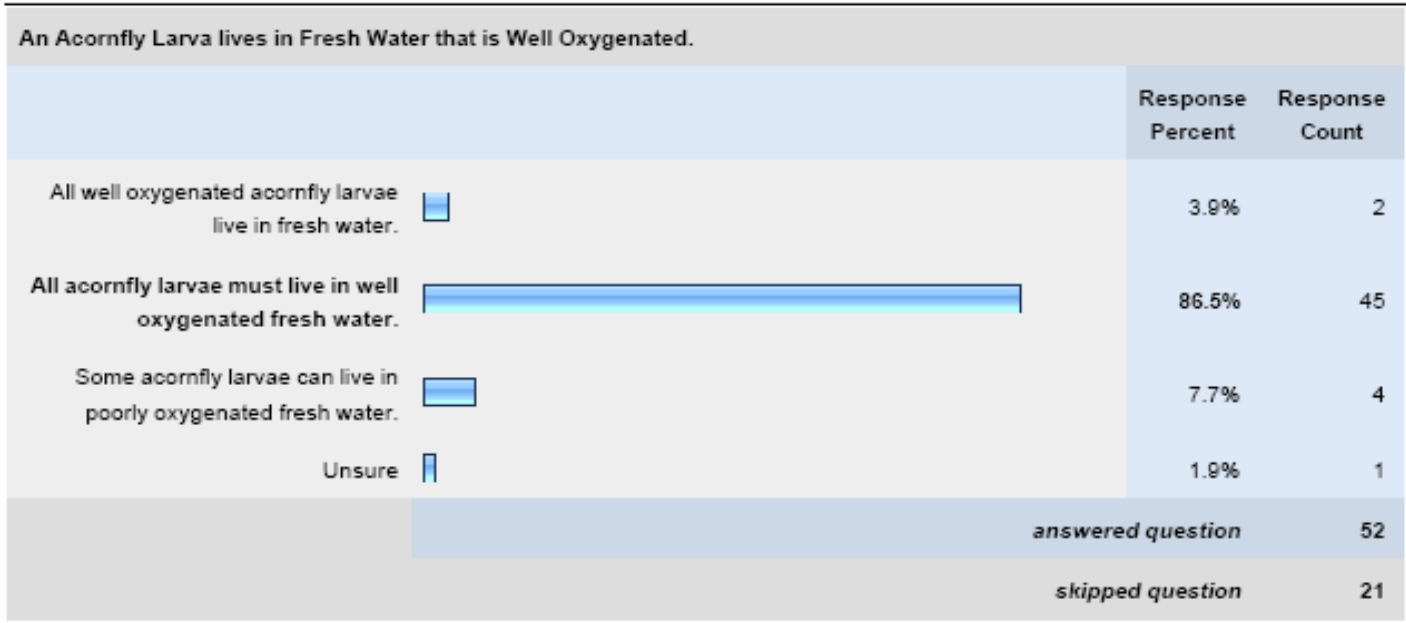


Human subject testing

- Phase 1: Understanding of Rabbit
  - Modifications to the language based on results
- Phase 2: Understanding of Rabbit – testing language improvements
  - Relationships problematic for users
  - Rabbit more understandable than Manchester OWL Syntax
- Phase 3: Authoring Rabbit
  - Designing the experimental set-up
  - Pilot testing

Human subject testing - phase 1

- Web-based questionnaire tested with 221 geography students
- Fictitious domain – users had no background knowledge
- Comprehension of sentence structure tested by multiple choice

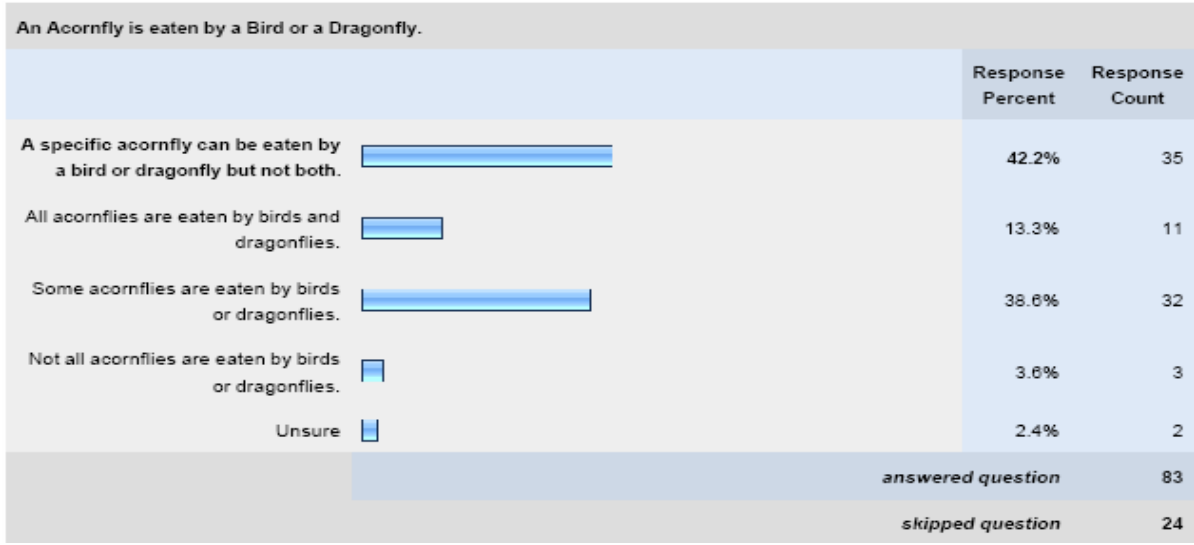


An Acornfly Larva lives in Fresh Water that is Well Oxygenated.

	Response Percent	Response Count
All well oxygenated acornfly larvae live in fresh water.	3.9%	2
All acornfly larvae must live in well oxygenated fresh water.	86.5%	45
Some acornfly larvae can live in poorly oxygenated fresh water.	7.7%	4
Unsure	1.9%	1
answered question		52
skipped question		21

Human subject testing - phase 1

- Testing design tried to isolate one structure at a time
- Over 75% correct was counted as acceptable
  - Cardinality (“at least”, “more than”, “exactly”)
  - Inclusive OR (“1 or more of”)
  - Existential and Universal quantifiers



An Acornfly is eaten by a Bird or a Dragonfly.

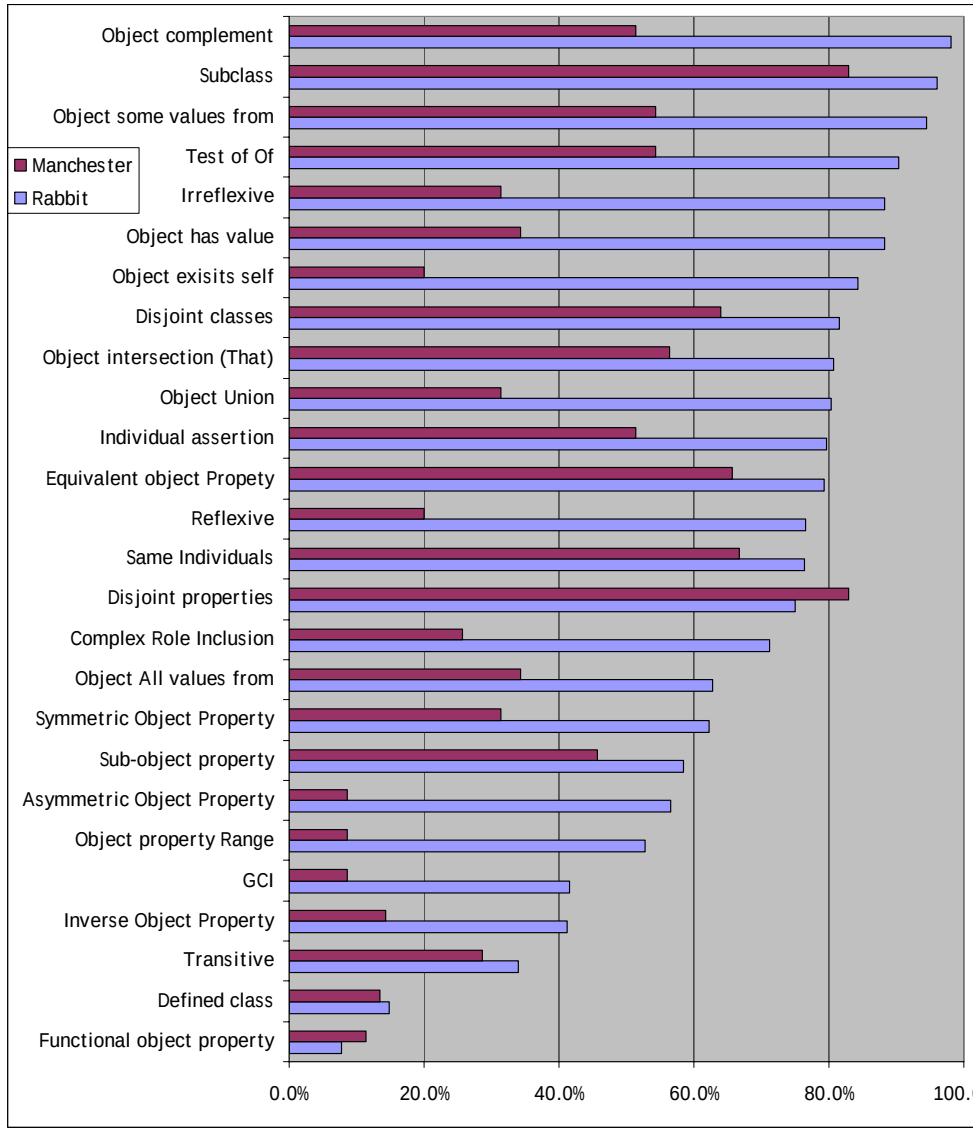
	Response Percent	Response Count
A specific acornfly can be eaten by a bird or dragonfly but not both.	42.2%	35
All acornflies are eaten by birds and dragonflies.	13.3%	11
Some acornflies are eaten by birds or dragonflies.	38.6%	32
Not all acornflies are eaten by birds or dragonflies.	3.6%	3
Unsure	2.4%	2
answered question		83
skipped question		24

Modifications to Rabbit based on phase 1 results

- *"is an instance of"* is difficult - use *"is a"* instead.
- Originally used *"A Halucinoptera is a kind of Insect"*
  - Applicability to all Haluncinoptera not understood
  - Changed to *"Every Halucinoptera is a kind of Insect"*
- Relationship characteristics (transitive, symmetric etc) very difficult to understand – use variables?

Phase 2 Results

- Comprehension of both concept and relationship sentences statistically significantly better for Rabbit compared with Manchester OWL syntax (univariate Anova)
- Rabbit concept sentences more comprehensible than relationship sentences.
- Relationship characteristics (transitive, inverse, functional etc) likely to always be a problem.
- Defined classes not understood



Phase 3: authoring

- Participants given:
  - 15 minute overview of the language
  - Table of Rabbit structures
- Text provided to control what participants “know” - 1 hour available to convert that knowledge to Rabbit
- Pilot testing with 3 participants produced encouraging results
- Average of 75% of sentences correctly authored
- Errors often due to systematic misunderstanding (e.g. omitting Every)

Conclusions

- If we truly want the web to be semantic, we need to encode those semantics accurately
- This can only be achieved with the active involvement of domain experts.
- A controlled natural language assists domain experts to:
  - Express their knowledge
  - Validate others' ontologies
  - Model knowledge well

Thank you for your attention

http://www.ordnancesurvey.co.uk/ontology