Natural Language Generation from Knowledge Bases

Laura Perez-Beltrachini
Université Henri Poincaré/ LORIA

joint work with Claire Gardent (CNRS/LORIA)

Outline of the talk

Natural Language Generation

Task definition [Reiter and Dale, 2000]

- Goal: produce understandable texts in human language
- Input:
  - communicative intention, and
  - information source
    - data-to-text: non linguistic data representation. e.g. knowledge base, data base, measuring devices logs
    - text-to-text: natural language text. e.g. newspaper

Natural Language Generation

Architecture: modules and tasks [Reiter and Dale, 2000]

- Document planning
  - Content selection: what information should be communicated in the text?
  - Text planning: imposing ordering and structure over the information to be conveyed (discourse/ rhetorical relations e.g. contrast,elaboration)
- Microplanning
  - Lexicalisation: choosing the content words in order to express the content selected (*the car owned by Mary* vs *Mary's car*)
  - Referring expression generation: how to produce a description of an entity that enables the hearer to identify the entity in a given context
  - Aggregation:
    - mapping grouped structures onto linguistic structures and textual elements (sentences/paragraphs)
    - sentence aggregation
- Realisation: transform an abstract representation into a syntactically and morphologically correct text

Natural Language Generation

Architecture: modules and tasks [Reiter and Dale, 2000]

- Document planning
  - Content selection: what information should be communicated in the text?
  - Text planning: imposing ordering and structure over the information to be conveyed (discourse/ rhetorical relations e.g. contrast,elaboration)
- Microplanning
  - Lexicalisation: choosing the content words in order to express the content selected (*the car owned by Mary* vs *Mary's car*)
  - Referring expression generation: how to produce a description of an entity that enables the hearer to identify the entity in a given context
  - Aggregation:
    - mapping grouped structures onto linguistic structures and textual elements (sentences/paragraphs)
    - sentence aggregation
- Realisation: transform an abstract representation into a syntactically and morphologically correct text

Natural Language Generation

Architecture: modules and tasks [Reiter and Dale, 2000]

- Document planning
  - Content selection: what information should be communicated in the text?
  - Text planning: imposing ordering and structure over the information to be conveyed (discourse/ rhetorical relations e.g. contrast,elaboration)
- Microplanning
  - Lexicalisation: choosing the content words in order to express the content selected (*the car owned by Mary* vs *Mary's car*)
  - Referring expression generation: how to produce a description of an entity that enables the hearer to identify the entity in a given context
  - Aggregation:
    - mapping grouped structures onto linguistic structures and textual elements (sentences/paragraphs)
    - sentence aggregation
- Realisation: transform an abstract representation into a syntactically and morphologically correct text

Natural Language Generation from Knowledge Bases

Though for different application contexts

Automatic generation of natural language from knowledge representations such as OWL is known as "verbalisation".

- ontology verbalisation (ACE OWL Verbaliser, SWAT, OntoVerbal)
- generation of descriptions (NaturalOWL) and summaries (MIAKT, FootOWL)
- query formulation based on ontology navigation (Quelo)
- ontology authoring (Power's Authoring Tool)

[Kaljur and Fuchs, 2007],[Liang et al., 2011b], [Galanis and Androutsopoulos, 2007], [Bontcheva, 2005],

[Agha et al., 2011],[E. Franconi and Trevisan, 2010], [Power, 2009]

Natural Language Generation from Knowledge Bases

Though for different application contexts

Automatic generation of natural language from knowledge representations such as OWL is known as "verbalisation".

- ontology verbalisation (ACE OWL Verbaliser, SWAT, OntoVerbal)
- generation of descriptions (NaturalOWL) and summaries (MIAKT, FootOWL)
- query formulation based on ontology navigation (Quelo)
- ontology authoring (Power's Authoring Tool)

[Kaljur and Fuchs, 2007],[Liang et al., 2011b], [Galanis and Androutsopoulos, 2007], [Bontcheva, 2005], [Agha et al., 2011],[E. Franconi and Trevisan, 2010], [Power, 2009]

But particularly for knowledge engineering purposes

Reasons for using NLG technology

- facilitate communication between knowledge engineers and domain experts
- ease the edition of knowledge bases for users not familiar with the underlying (logic-based) formalism
- generate documentation of the ontology

Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. [Rector et al., 2004]
Owl pizzas: Practical experience of teaching OWL-DL: Common errors and common patterns.

Liang, Shao Fen and Scott, Donia and Stevens, Robert and Rector, Alan [Liang et al., 2011a]
Unlocking medical ontologies for non-ontology experts.

Ontology verbalisation

General underlying approach

- Most NLG applications
  - seek for high quality text in a specific domain
  - use specialized text-planners, grammars and lexicons
- Ontology verbalisation
  - aims for texts that are useful and understandable (not necessarily high quality) and domain independent methods

Ontology verbalisation

General underlying approach

- rely on the fact that natural language is used to name symbols of the underlying formalism to construct the underlying content lexicon (e.g. *MAILMAN* concept onto the noun *mailman*)
  - concepts map onto nouns or adjectives
  - roles map onto transitive verbs (inverse role maps to the passive form)
- map underlying semantic operators and constructors onto natural language constructs

Chris Mellish and Xiantang Sun [Mellish and Sun, 2006]
The Semantic Web as a linguistic resource: opportunities for Natural Language Generation.

OWL verbalisers

Depending on the design goals verbalisers

- focus on reversibility. Take a as target language a Controlled Natural Language (CNL) designed to be unambiguously translated to the underlying (logic-based) formalism.
  - e.g. ACE, SOS
- focus on presentation as text: SWAT, OntoVerbal

Verbalisers using CNLs focusing on reversibility

- provide a natural language realisation of the underlying representation in a target CNL
  - the CNL rules out ambiguity –lexical, structural, referential, semantic – and in some cases complexity
  - because of the target language their realisation is reversible
  - the use of a CNL for text generation might result in a stilted NL

Verbalisers using a larger set of NL constructions for text generation

- Provide a natural language realisation of the knowledge base but in this case the goal is to provide a fluent text describing ontology elements
- Require further stages of NLG to produce structured, coherent and fluent text
  - eg. text planning –discourse, aggregation referring expression generation

Verbalising OWL in ACE

Verbalisation goal [Kaljur and Fuchs, 2007]

- ACE is a subset of English where ambiguity is ruled out and that can be converted into FOL
- Verbalization of OWL in ACE should be reversible. i.e. one can convert the ACE representation back into OWL so that no loss in meaning occurs
- Set of construction rules and set of interpretation rules
  - the construction rules define its syntax

Verbalising OWL in ACE

Interpretation rules

- *of*-constructions
  A man sees a dog of {Mary and John }
- adjuncts
  A customer { enters a card with a code}
- quantification
  Every customer types { a code}
- relative clause coordination
  - A relative clause can modify the object via *that* or the subject via *and/or that* of a previous clause.
- anaphoric references
  - accessibility of antecedents
  - *There is a blue ball. There is a red ball. John sees the ball [=the red ball].*

Verbalising OWL in ACE

Verbalization approach

- Mapping terms to words (automatic lexicon creation):
  - individuals → proper nouns; classes → nouns
  - properties → Tr. verbs in active voice; inverse properties → Tr. verbs verbs in passive voice
  - class expressions → ACE complex noun phrases
    - relative clauses and coordination: *'and that'*, *'or that'*, *'that is/does not'* and *'that'*.
- Mapping axioms to sentences (defined through DCG grammar):
  - except *ClassAssertion*-axioms, other axioms map onto *Every*-sentences with pattern `[NounPhrase VerbPhrase]`

Verbalising OWL in ACE

Verbalization approach (Cont.)

- 1-sentence 1-axiom (the selected content to be verbalised is the whole ontology)
- for readability in some cases:
  - ordering within the sentence (simple class descriptions mentioned first)
  - referring expression generation use of pronouns, e.g. *Everything hasConsorts itself* .

Verbalising OWL in ACE

Limitations

OWL parenthesis vs ACE interpretation rules

- Syntactically complex class descriptions, e.g. $(\exists R_1(\exists R_2 C_1)) \sqcap (\exists R_3(\exists R_4 C_2))$ cannot be verbalized :

(1) Class: haulage_truck_driver
EquivalentTo:
person and (drives some (truck and equipped_with some Outside_lock)) and (works_for some (branch and part_of some haulage_company))

Every haulage truck driver is a person that drives a truck that is equipped_with an outside_lock and that works_for a branch that is part_of a haulage_company. Every...

Generating text from OWL: The SWAT system

SWAT Alphabetical English glossary from class, individual and property definitions

Verbalisation goal

- Generate text rather than sentences
- Instead of verbalising axioms one by one
  - applies some structuring rules and aggregation (applicable to any ontology) to provide coherent descriptions

Generating text from OWL: The SWAT system

Verbalisation approach

- lexicon: inferring lexical entries for atomic entities from IRI identifiers (special attention on POS and plural forms)
- content selection: grouping related axioms for each entity (class, individual or property)
- text planning: organizing selected content into glossary tags
- aggregation: grouping axioms sharing a common pattern (to obtain more fluent descriptions)
- realization: sentence generation is accomplished by a generic grammar with rules for each logical patterns in owl

Generating text from OWL: The SWAT system

Verbalisation approach

- lexicon: inferring lexical entries for atomic entities from IRI identifiers (special attention on POS and plural forms)
- content selection: grouping related axioms for each entity (class, individual or property)
- text planning: organizing selected content into glossary tags
- aggregation: grouping axioms sharing a common pattern (to obtain more fluent descriptions)
- realization: sentence generation is accomplished by a generic grammar with rules for each logical patterns in owl

Generating text from OWL: The SWAT system

Text planning for OWL entities of type class [Williams et al., 2011]

Structure for description of entities of type class *TopicClass*

- **Examples.** $A \sqsubseteq$ *TopicClass* , $B \sqsubseteq$ *TopicClass*, a : *TopicClass*
- **Description.** *TopicClass* $\sqsubseteq$ D, R range *TopicClass* , R domain *TopicClass*,
- **Definition.** *TopicClass* EquivalentTo D
- **Distinctions.** *TopicClass* Disjoint B
- **Typology.** *TopicClass* $\sqsubseteq$ A

where *TopicClass*,A,B are atomic classes; D is a complex class description; and R is an object property

Generating text from OWL: The SWAT system

Text planning for OWL entities of type class: Example people+petsB.owl

ANIMAL (class)
Description. An animal eats a thing.
If X has as pet Y then Y is an animal.
If X eats Y then X is an animal.
Examples. The following are animals: tigers, sheep and people, and so on (6 items in total).

BROADSHEET (class)
Typology. A broadsheet is a newspaper.
Examples. The Times and The Guardian are broadsheets.
A quality broadsheet is a broadsheet.
Distinctions No broadsheet is a tabloid.

ANIMAL LOVER (class)
Definition. An animal lover is defined as a person that has as pet at least three things.

Generating text from OWL: The SWAT system

Grouping axioms [Williams and Power, 2010]

Group *SubClassOf* axioms with the same pattern
Example: pattern $A \sqsubseteq B$, aggregation at the left:

```
bicycle SubClassOf vehicle
van SubClassOf vehicle
coach SubClassOf vehicle
truck SubClassOf vehicle
```

Aggregation: subClassOf([class(bicycle),class(van),class(coach), class(truck)],class(vehicle))

VEHICLE (class)
Examples The following are vehicles: vans, trucks and coaches, and so on (5 items).
Aggregation can help minimize redundancy and repetitions

Both ACE and SWAT

- Use a DCG grammar to parse OWL axioms and class expressions to produce text.

Querying with QUELO

An intelligent query interface based on ontology navigation using Natural Language Generation

- The user defines a query by editing an initial sentence generated by Quelo.

  I am looking for a thing.

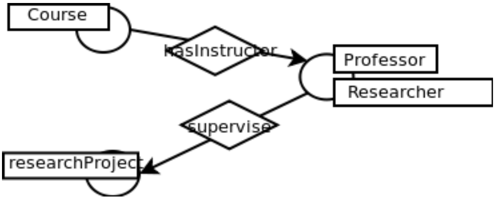
- Refinement process of the query.
  - replace text: substitute the focused node by a super or sub concept or an equivalent concept,
  - insert text: add compatible concept or relation,
  - delete text: delete query operations
- The suggestions for these operations are provided by a reasoning service over the query and ontology
  - the query logic component provides the content for the operations
  - the NLG component is in charge of its verbalization
- Text highlighting gives visual hints of the underlying query tree.

Conjunctive query verbalisation: Quelo

Verbalisation goal

Generate a description of the concept(s) to be retrieved

$$Course \sqcap \exists has\_instructor.(Professor \sqcap Researcher \sqcap \exists supervise.Project)$$



Quelo's query model is a data structure linking tree-shaped conjunctive DL queries and text.

I am looking for a course. Its instructor should be a professor. The professor should be a researcher and supervise a research project.

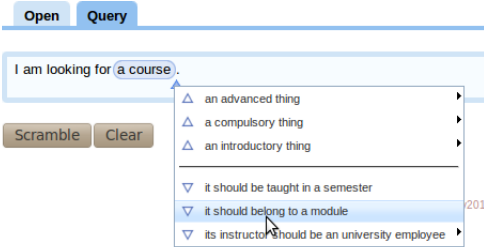
Conjunctive query verbalisation: Quelo

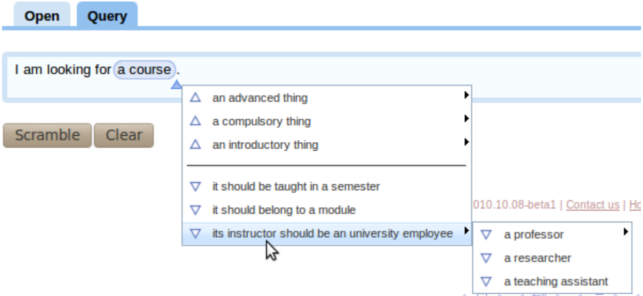
Verbalisation approach

- content selection: interaction between user choices and the query logic component providing the content for selection
- text planning: The underlying query tree determines the order
- lexicalisation and surface realisation:
  - Automatically generated lexicon
  - Nodes are mapped onto NPs
  - Edges are mapped onto clause templates
  - Lexicalisation varies depending on whether the concept is noun- or adjective-based and whether a relation is verb- (sell or noun- (hasInstructor) based

Conjunctive query verbalisation: Quelo

Verbalisation approach





Conjunctive query verbalisation: Quelo

Verbalisation approach (Cont.)

Improving readability: aggregation and anaphoric references

- Analyze the sequence of clauses obtained from a Quelo query

  Joins adjacent clauses with the same subject and the same voice (active, passive) into a single sentence
  The firm should manufacture goods.
  The firm should provide services.
  The firm should manufacture goods and provide services.

- First reference NP and subsequent references: usage of the semantic information for anaphoric references

  I am looking for a master program. It should include a module. The module should include a course.

Experiment with flat semantics and SemXTAG grammar

Ontology verbalisation

$\text{Factory} \sqsubseteq \exists\ \text{has\_part}.(\exists \text{has\_purpose}.\text{Manufacturing} \sqcap \text{Building})$

(2) a. ACE. *For every factory its part is a building whose purpose is a manufacturing.*

b. SOS. *Every factory has a building as a part that has a manufacturing as a purpose.*

Lexicalisation choices? Use transitive verbs instead of nominalisations? [Schwitter et al., 2008]

Every factory comprises a building that manufactures something.

Experiment with flat semantics and SemXTAG grammar

Ontology verbalisation

dog $\sqsubseteq$ animal

(3)

a. ACE. *Every dog is an animal .*

b. SWAT. *A dog is an animal.*

c. FOL. $\forall x.(dog(x) \rightarrow animal(x))$

$\Downarrow \quad \Downarrow \quad \Downarrow$

d. FS. $\{l_0 : forall(x, l_1, l_3), l_1 : dog(x), l_3 : animal(y)\}$

e. GenI. *dogs are animals.*

Verbalization closely related to the underlaying formula (*all* and *every*) or take the view point that we talk about generic concepts and then use indefinites or plurals?

Experiment with flat semantics and SemXTAG grammar

Ontology verbalisation (Cont.)

dog $\sqsubseteq$ $\exists$eat.bone

(4) a. ACE. *Every dog eatses a bone.*

b. SWAT. *A dog eats a bone.*

c. FOL. $\forall x : [dog(x) \rightarrow \exists y : [eat(x, y) \wedge bone(y)]]$

d. Everything that is a dog eats a bone.
If there is something that is a dog then it eats a bone.

$\Downarrow \quad \Downarrow \quad \Downarrow$

e. FS. $\{l_0 : forall(x, l_1, l_3), l_1 : dog(x), l_4 : exists(y, l_2, l_3), l_2 : bone(y), l_3 : eat(e, x, y)\}$

f. GenI. *dogs eat bones.*

Experiment with flat semantics and SemXTAG grammar

Mapping Quelo queries onto flat semantics

(5) *I am looking for a masters program. It should include a module.*

a. MasterProgram ⊓ ∃includes.Module

b. $MasterProgram(x) \wedge \exists y.(includes(x,y) \wedge Module(y))$

c. $\{l_0 : john(j), l_1 : look\_for(e1, j, x), l_7 : exists(x, l_2, l_1), l_2 : masterProgram(x), l_3 : exists(y, l_4, l_5), l_4 : module(y), l_5 : includes(e2, x, y)\}$

d. *john looks for a masterProgram that includes a module.*
john looks for a masterProgram which includes a module.

e. $\{l_0 : john(j), l_1 : look\_for(e1, j, x), l_7 : exists(x, l_2, l_1), l_2 : masterProgram(x), l_3 : exists(y, l_4, l_1), l_4 : module(y), l_5 : includes(x, y)\}$

f. *john looks for a masterProgram with a module.*

aggregation choice –one sentence with relative clause? lexicalization choice – preposition with?

Experiment with flat semantics and SemXTAG grammar

Mapping Quelo queries onto flat semantics

(6) *I am looking for a masters program. It should include a module.*

a. MasterProgram $\sqcap$ $\exists$includes.Module

b. $\mathit{MasterProgram}(x) \wedge \exists y.(\mathit{includes}(x, y) \wedge \mathit{Module}(y))$

c. $\{l_0 : \mathit{john}(j), l_1 : \mathit{look\_for}(e1, j, x), l_7 : \mathit{exists}(x, l_2, l_1), l_2 : \mathit{masterProgram}(x), l_3 : \mathit{exists}(y, l_4, l_5), l_4 : \mathit{module}(y), l_5 : \mathit{includes}(e2, x, y)\}$

d. *john looks for a masterProgram that includes a module.*
john looks for a masterProgram which includes a module.

e. $\{l_0 : \mathit{john}(j), l_1 : \mathit{look\_for}(e1, j, x), l_7 : \mathit{exists}(x, l_2, l_1), l_2 : \mathit{masterProgram}(x), l_3 : \mathit{exists}(y, l_4, l_1), l_4 : \mathit{module}(y), l_5 : \mathit{includes}(x, y)\}$

f. *john looks for a masterProgram with a module.*

aggregation choice –one sentence with relative clause? lexicalization choice – preposition with?

Discussion

NLG from knowledge bases summary

- NLG task were KB are the input source
- according to the application goals different NLG task requirements
- trade off between having a language very close to the underlying formalism or applying more sophisticated NLG methods to get a more fluent text.

Thank you!

References I

Agha, N. B., Casamayor, G., Wanner, L., and Hernández, S. L. (2011).
Footbowl: Using a generic ontology of football competition for planning match summaries.
In *8th Extended Semantic Web Conference (ESWC2011)*.

Bontcheva, K. (2005).
Generating tailored textual summaries from ontologies.
In *ESWC*, pages 531–545.

E. Franconi, P. G. and Trevisan, M. (2010).
An intelligent query interface based on ontology navigation.
In *Workshop on Visual Interfaces to the Social and Semantic Web*, Hong Kong.

Galanis, D. and Androutsopoulos, I. (2007).
Generating multilingual descriptions from linguistically annotated owl ontologies: the natural owl system.
pages 143–146, Schloss Dagstuhl, Germany.

Kaljur, K. and Fuchs, N. E. (2007).
Verbalizing owl in attempto controlled english.
In *In Proceedings of OWLED07*.

Liang, S. F., Scott, D., Stevens, R., and Rector, A. (2011a).
Unlocking medical ontologies for non-ontology experts.
In *Proceedings of BioNLP 2011 Workshop*, BioNLP '11, pages 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.

Liang, S. F., Stevens, R., Scott, D., and Rector, A. (2011b).
Automatic verbalisation of snomed classes using ontoverbal.
In *Proceedings of the 13th conference on Artificial intelligence in medicine*, AIME'11, pages 338–342, Berlin, Heidelberg. Springer-Verlag.

References II

Mellish, C. and Sun, X. (2006).
The semantic web as a linguistic resource: opportunities for natural language generation. knowledge based systems.
In *Presented at the Twenty-sixth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*.

Power, R. (2009).
Towards a generation-based semantic web authoring tool.
In *ENLG*.

Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., and Wroe, C. (2004).
Owl pizzas: Practical experience of teaching OWL-DL: Common errors and common patterns.
In *In Proc. of EKAW 2004*, pages 63–81. Springer.

Reiter, E. and Dale, R. (2000).
Building natural language generation systems.

Schwitter, R., Kaljur, K., Cregan, A., Dolbear, C., and Hart, G. (2008).
A comparison of three controlled natural languages for owl 1.1.
In *In 4th OWL Experiences and Directions Workshop (OWLED 2008 DC*.

Williams, S. and Power, R. (2010).
Grouping axioms for more coherent ontology descriptions.
In *Proceedings of the 6th International Natural Language Generation Conference*, INLG '10, pages 197–201, Stroudsburg, PA, USA. Association for Computational Linguistics.

Williams, S., Third, A., and Power, R. (2011).
Levels of organisation in ontology verbalisation.
In *Proceedings of the 13th Workshop on Natural Language Generation*, ENLG '11.