Generating from Knowledge Bases

Claire Gardent¹ and Laura Perez-Beltrachini ² (1) CNRS/LORIA, (2) Université Henri Poincaré

SAKT 2012 November 2012, Bolzano

Outline



2 Lexicalisation





Grammar Based Surface Realisation

Surface realisation (SR) maps an abstract linguistic representation to a sentence

Grammar Based Approaches

- Use a grammar which associates NL expressions with both a syntactic tree and a semantic representation
- If the grammar is reversible, it can be used either for parsing (from string to meaning) or for surface realisation (from meaning to string)

 Outline
 Grammar Based Surface Realisation
 Lexicalisation
 Composing text

 NLG with Feature-Based Lexicalised Tree Adjoining
 Grammar —equipped with semantics (SemTAG)
 Grammar —equipped with semantics (SemTAG)



name(j,john), run(a,j), often(a) John often runs

Lexicalisation

Grammars and Algorithms

Grammars

- SemFRAG: a semantics FB-LTAG for French (core fragment)
- SemXTAG: A reconstruction of the UPenn XTAG grammar for English(partially) extended with a unification based semantics

Algorithms

- Genl and RTGen: Generate from flat semantics
- DRTgen: generate from unordered, lemmatised dependency tree (Surface Realisation Task)

Applications

- Generating grammar exercices for language learner (ALLEGRO EU Project)
- Generating from an ontology describing a virtual world (IFLEG 3D game for language learning)
- Generating the Penn Treebank sentences (SR Task)
- Generating class descriptions (AURA Project)

exicalisatio

Composing text

Discussion

Grammar Based NLG and Conceptual Authoring

Conceptual Authoring ([Hallett et al., 2007])

- Text and Meaning Representations (MR) are linked
- Editing the text modifies the corresponding MR and vice versa

SemTAG naturally supports conceptual authoring

- It systematically relates text, syntax and semantics
- The relation between text and semantics is mediated by unification variable and is therefore order independent or bidirectional

Lexicalisatio

Discussion

SemTAG and Conceptual Authoring: Quelo Example

I am looking for a man. One of his parents is a singer. $_0$ man $_1$ hasParent $_2$ singer $_3$



Lexicalisation

General approach to NLG from KB

Lexicalisation

- Choosing words and syntactic structures e.g. *the car owned by Mary* vs *Mary's car*
- rely on the fact that natural language is used to name symbols of the underlying formalism to construct the underlying *content lexicon* [Mellish and Sun, 2006]
 e.g. MAILMAN concept onto the noun *mailman*

lexical entries (inferred automatically) containing lemma, POS and word forms

• Formal language constructions map to syntactic constructions [Kaljurand and Fuchs, 2007, Power, 2010]

grammars or templates to build syntactic structure

\sim			
()	+1	ine	
\sim			

Lexicalisation

The case of Quelo templates

- concept and relation ++++ templates (ref. lexical entries)
- Iexicalisation (and SR):
 - (i) pick up templates
 - (ii) instantiate templates:

based on template type (noun-based, adjective-based or verb-based) build abstract syntactic structure

(iii) realise text according to abstract syntactic structure



In SemTAG lexicon associates lemmas with syntax and semantics (concepts and relations)

- concept and relation +++> SemTAG elementary trees
- Iexicalisation (and SR):
 - (i, ii) choose elementary trees
 - (iii) text is the yield of the tree (words in the leaves)

Outline

Lexicalisation

Discussion

The case of Quelo templates

0 man 1 hasFather 2 singer 3



Lexicon cat:n;singular:father;plural:-;countable:+;gender:m Template Map hasFather: Noun-based relation template

Instantiate template syntactic element: The N_{father}(Number) of NP_{the man} should be NP_{a singer}

The case of Quelo with SemTAG

0 man 1 hasFather 2 singer 3

Lexicon lemma: father TAG tree: RelNoun semantics: 7 hasFather 2



RelNoun:1 hasFather 2

Generating text

- Generation from KB: (axiom by axiom or combining them) generation takes place once before text is presented
- Conceptual Authoring: text is generated (and modified) dynamically and user takes part in content and ordering decisions

deletion, replacement and addition of text (modification of the underlying conceptual structure)



I am looking for a masters program).	Its coordinator	should be	a coordinator.	The masters	program	should include	a module).
A			A				⋧ ▲



Addition is naturally modeled by Adjunction in TAG

```
0 man 1 hasFather 2 singer 3
+actor
```





Adding a concept to a node

$_0$ man $_1$ hasFather $_2$ singer $_3$

- (1) a. I'm looking for a man. His father should be a singer.
 - b. I'm looking for a man whose father should be a singer.
 - c. I'm looking for a man. The father of the man should be a singer.

Adding a concept to a node

0 man 1 hasFather 2 singer 3

- (3) a. I'm looking for a man. His father should be a singer.
 - b. I'm looking for a man whose father should be a singer.
 - c. I'm looking for a man. The father of the man should be a singer.

```
_0 man _1 hasFather _2 singer _3 +actor
```

Adding a concept to a node

0 man 1 hasFather 2 singer 3

- (5) a. I'm looking for a man. His father should be a singer.
 - b. I'm looking for a man whose father should be a singer.
 - c. I'm looking for a man. The father of the man should be a singer.

$_0$ man $_1$ hasFather $_2$ singer $_3$

+actor

- (6) a. I'm looking for a man who is an actor. His father should be a singer.
 - b. I'm looking for a man who is an actor whose father should be a singer.
 - c. I'm looking for a man who is an actor. The father of the man should be a singer.

```
0 man 1 hasFather 2 singer 3
0 man 1 marriedTo 4 actress 5
```



Adding a relation to a node

0 man 1 hasFather 2 singer 3

- (7) a. I'm looking for a man. His father is a singer.
 - b. I'm looking for a man whose father is a singer.
 - c. I'm looking for a man. The father of the man is a singer.

Adding a relation to a node

$_0$ man $_1$ hasFather $_2$ singer $_3$

- (9) a. I'm looking for a man. His father is a singer.
 - b. I'm looking for a man whose father is a singer.
 - c. I'm looking for a man. The father of the man is a singer.

0 man 1 hasFather 2 singer 3

- 0 man 1 marriedTo 4 actress 5
- (10) a. I'm looking for a man who is married to an actress. His father is a singer.
 - b. X I'm looking for a man who is married to an actress whose father is a singer.
 - c. I'm looking for a man who is married to an actress. The father of the man is a singer.

Quelo NLG component with SemTAG

Open issues

- Develop the grammar
 - XMG framework ([Crabbé et al., 2012]) or grammar induction
- Associate concepts and relations with TAG trees
 - adapt Quelo techniques to associate them with templates
- Selection among elementary trees (non-determinism)
 - preserve options (tabulating)
 - select "appropriate" constructions using some heuristic
- Incremental generation algorithm
 - guided by input order
 - incremental grammar ([Demberg and Keller, 2008])
- Evaluation
 - comparison with Quelo (templates version)

Outline	Grammar Based Surface Realisation	Lexicalisation	Composing text	Discussion

Thank you!

Lexicalisation

References I

Crabbé, B., Duchier, D., Gardent, C., Leroux, J., and Parmentier, Y. (2012). XMG, eXtendible Meta Grammar. Computational Linguistics
Demberg, V. and Keller, F. (2008).
A Psycholinguistically Motivated Version of TAG. In 9th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+9 2008).
Hallett, C., Scott, D., and Power, R. (2007).
Composing questions through conceptual authoring. Comput. Linguist., 33(1):105–133.
Kaljurand, K. and Fuchs, N. E. (2007).
Verbalizing owl in attempto controlled english. In In Proceedings of OWLED07.
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. (2010).

Complexity assumptions in ontology verbalisation.

In Proceedings of the ACL 2010 Conference Short Papers, ACLShort '10, pages 132–136, Stroudsburg, PA, USA. Association for Computational Linguistics.