

# SimpleNLG-EnFr

*Adding French to SimpleNLG*

Pierre-Luc Vaudry   
Université de Montréal **rali**

# Goals

- Generate grammatically correct French
- Make as few changes as possible
  - API
  - Internal working
  - Existing JUnit tests
- Enable bilingual generation
- Facilitate further language addition

# French syntax challenges

- Noun phrase
  - Default adjective placement
  - Determiner and adjective agreement
- Verb phrase and clause
  - Negation (with or without negation auxiliary)
  - Clitic complement pronouns placement
  - Past participle agreement

*Je **n'ai pas** écrit cette lettre à ta soeur.  
Personne **ne** la lui a écrite.*

# French morphology

- Plural and feminine forms

*chat, chats; chatte, chattes; cheval, chevaux  
heureux, heureuse, heureuses*

- Verb tenses

*boire, tu buvais, tu bois, tu boiras  
bois, buvant, bu*

- Personal pronoun forms

*3<sup>rd</sup> pers. sg.: il, elle, le, la, lui, se; lui, elle, soi*

# French morphophonology and punctuation

- Morphophonology

- Prepositions : *à, de + le → au, du*
- Elision : *le + homme → l'homme*
- Liaison : *le + beau + homme → le bel homme*

- Punctuation

- Coordination :  
*Pierre a parlé **et** Jean lui a répondu.*  
*Pierre a parlé, **mais** Jean ne lui a pas répondu.*
- Frontmodifier : *Heureusement, il le savait.*

# Changes in SimpleNLG structure

- Each NLGElement knows its language
  - Language field added to lexicon
  - Lexicon field added to WordElement
- Syntax, morphology and orthography modules divided between :
  - NLGElement and subclasses
  - Generic abstract helper classes
  - English specific helper classes
- New processing level : morphophonology

# Statistics

- 22 out of 60 existing classes modified
  - 69 methods modified
  - 57 methods added
- 30 new classes
- 10 new packages
- 29 new features

# Building the French lexicon

- Function words entered manually
- Content words :
  - From Dubois Buyse orthographic scale
    - <http://o.bacquet.free.fr/db2.htm>
  - Inflected forms from Morphalou 2.0
  - Syntactical features and some inflected forms added manually with Grevisse
- Total : 3871 entries



# Achievements

- Bilingual English/French realization
- Covers most of the grammar in *Le Français fondamental (1<sup>er</sup> degré)*
- 3871 entry French default XML lexicon
- Detailed French manual
- API almost identical
- Internal working somewhat modified
- Original JUnit tests still passed successfully

# Future work

- Complete coverage of *Le français fondamental (1<sup>er</sup> degré)*
  - Interrogative sentences and relative clauses
  - Conditional and subjunctive moods
  - Number nouns, numeral and cardinal adjectives
- XML realizer and other changes in v4.3
- Specialized French lexicons
- Other languages...