

SemTAG: The LORIA toolbox for TAG-based Parsing & Generation

Eric Kow and Yannick Parmentier
INRIA/LORIA
Universite Henri Poincare
615, rue du Jardin Botanique
F-54 600 Villers-Les-Nancy
{kow,parmenti}@loria.fr

Claire Gardent
CNRS/LORIA
615, rue du Jardin Botanique
F-54600 Villers-Les-Nancy
gardent@loria.fr

PARSING

The SemTAG toolbox includes 2 parsing systems:

LLP2

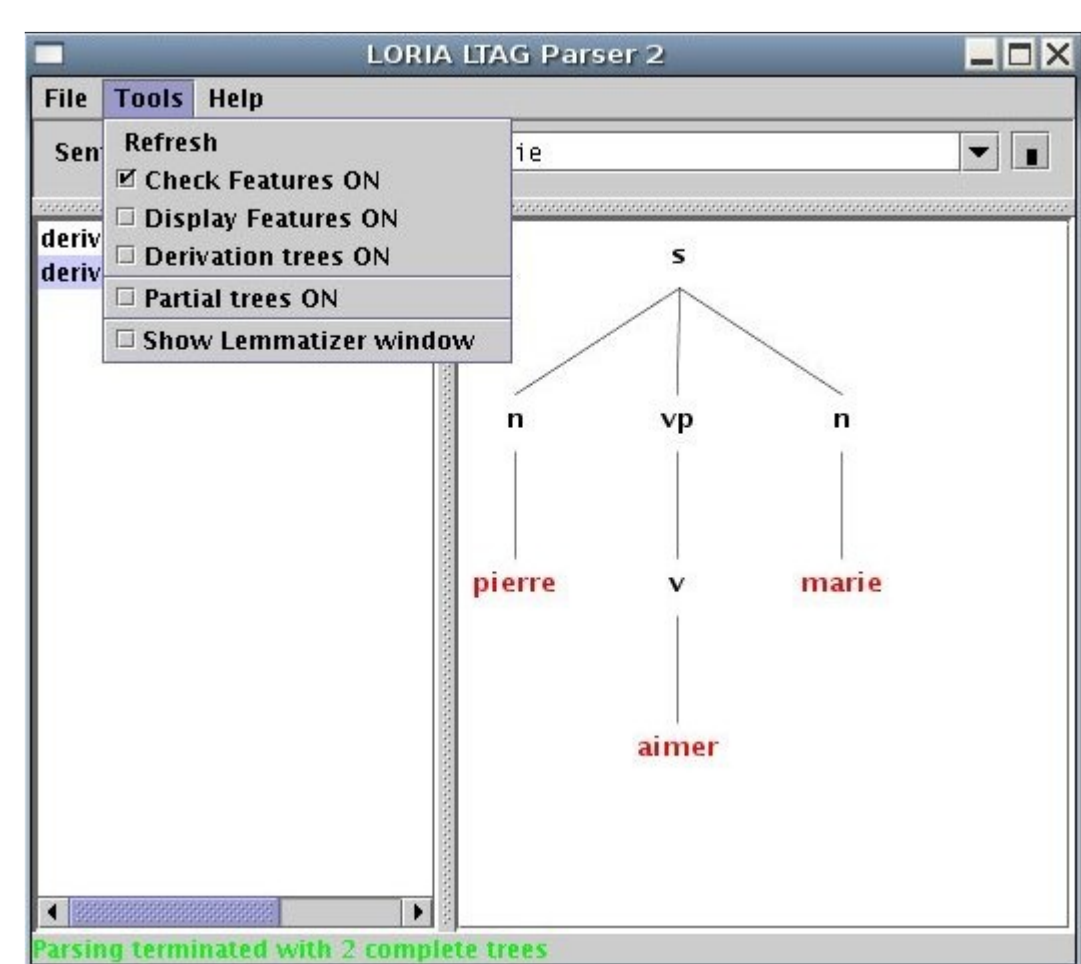
- parser using a bottom-up algorithm (Lopez, 99) and providing a friendly GUI with many statistical information (well suited for teaching projects).

DyALog

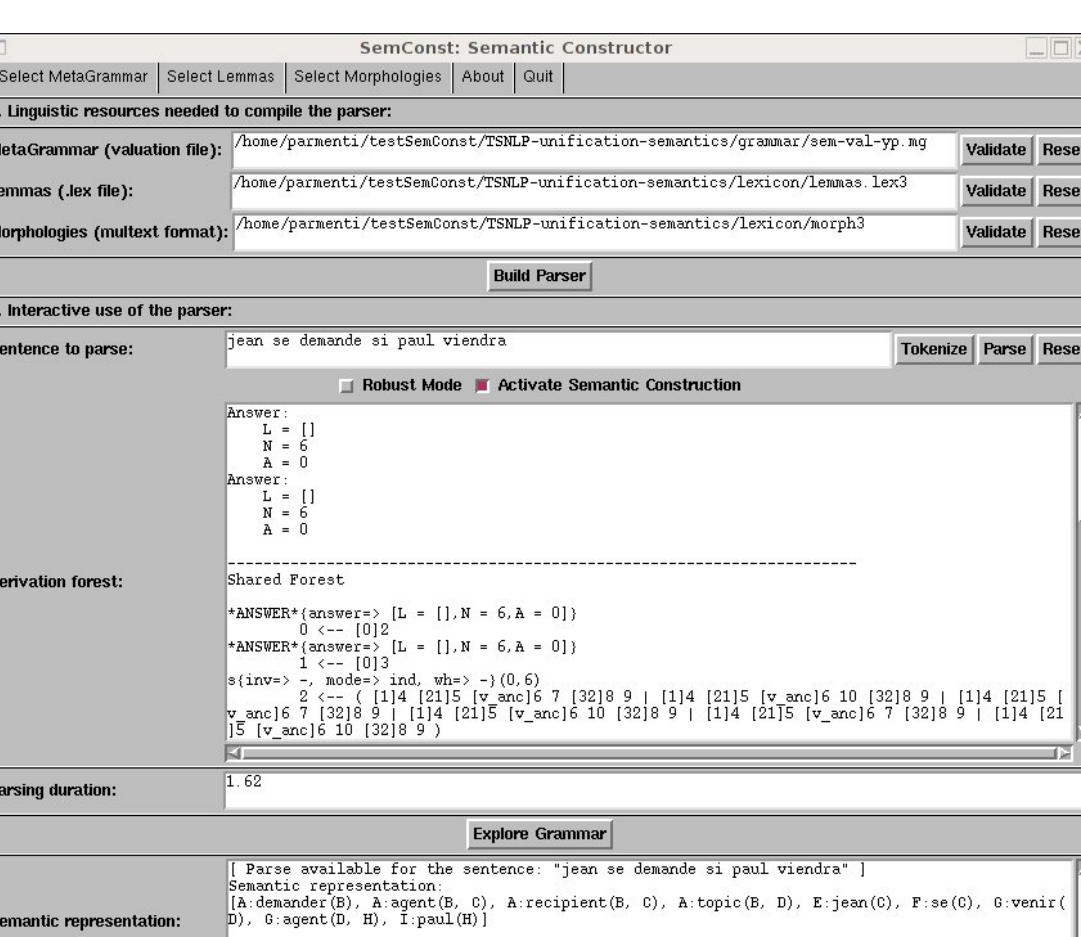
- optimised parsing system based on tabulation and automata techniques (De La Clergerie, 05).

SemConst

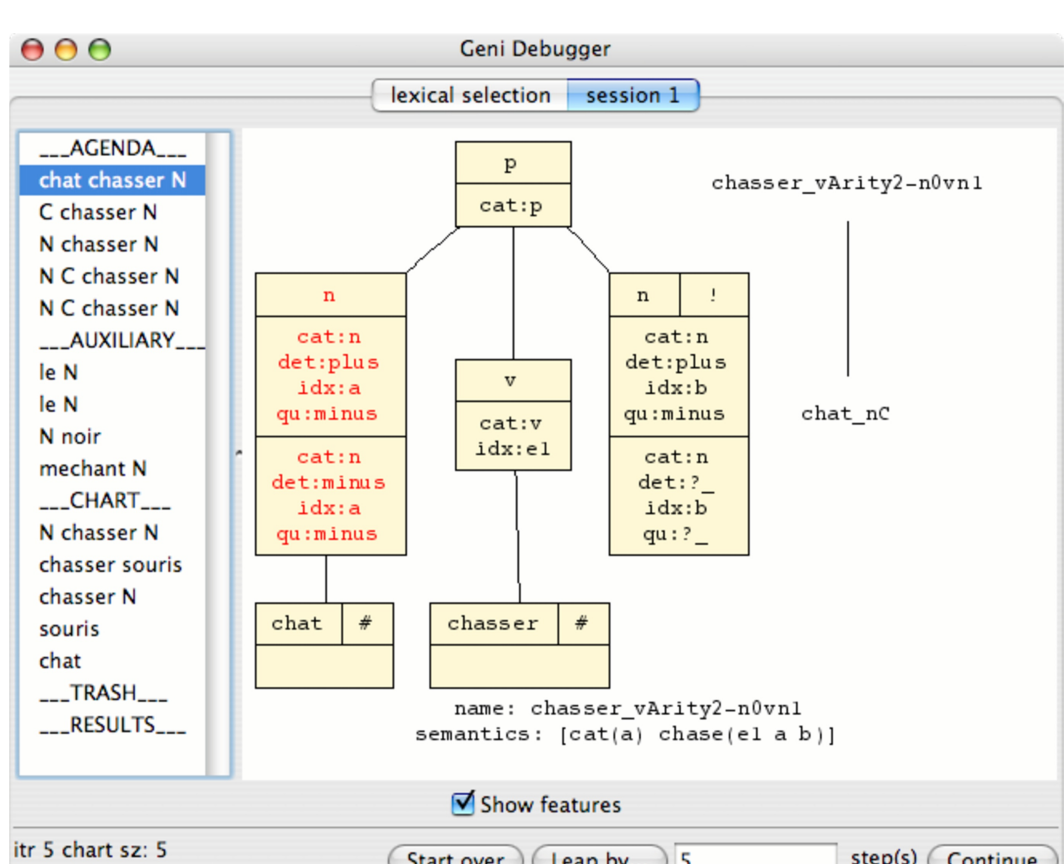
- a semantic construction module using the XMG and DyALog systems to perform semantic construction following the ideas of (Gardent and Kallmeyer, 03) and (Gardent and Parmentier, 05).



Loria LTAG Parser 2



SEMantic CONSTRUCTOR



GenI debugger

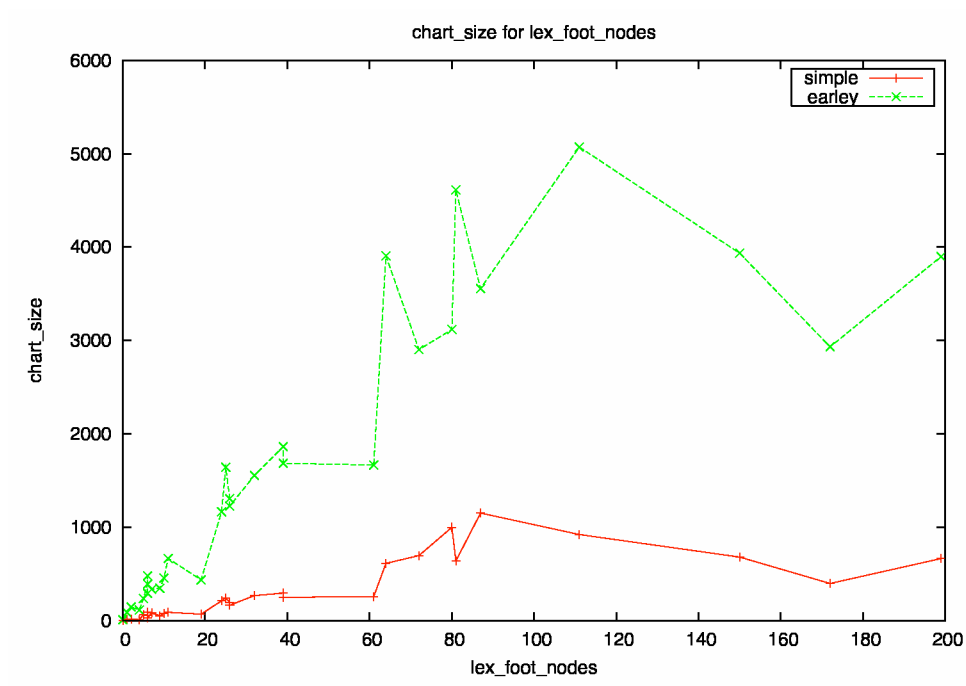
GENERATION

GenI (Gardent and Kow, 05)

- surface realiser (flat semantics + TAG = sentences)
- written in Haskell, 8 000 lines of code
- graphical debugger
- two algorithms (simple one, Earley) some optimisations (see other poster)

gtester

- test harness for GenI
- automated regression testing
- benchmarking for performance



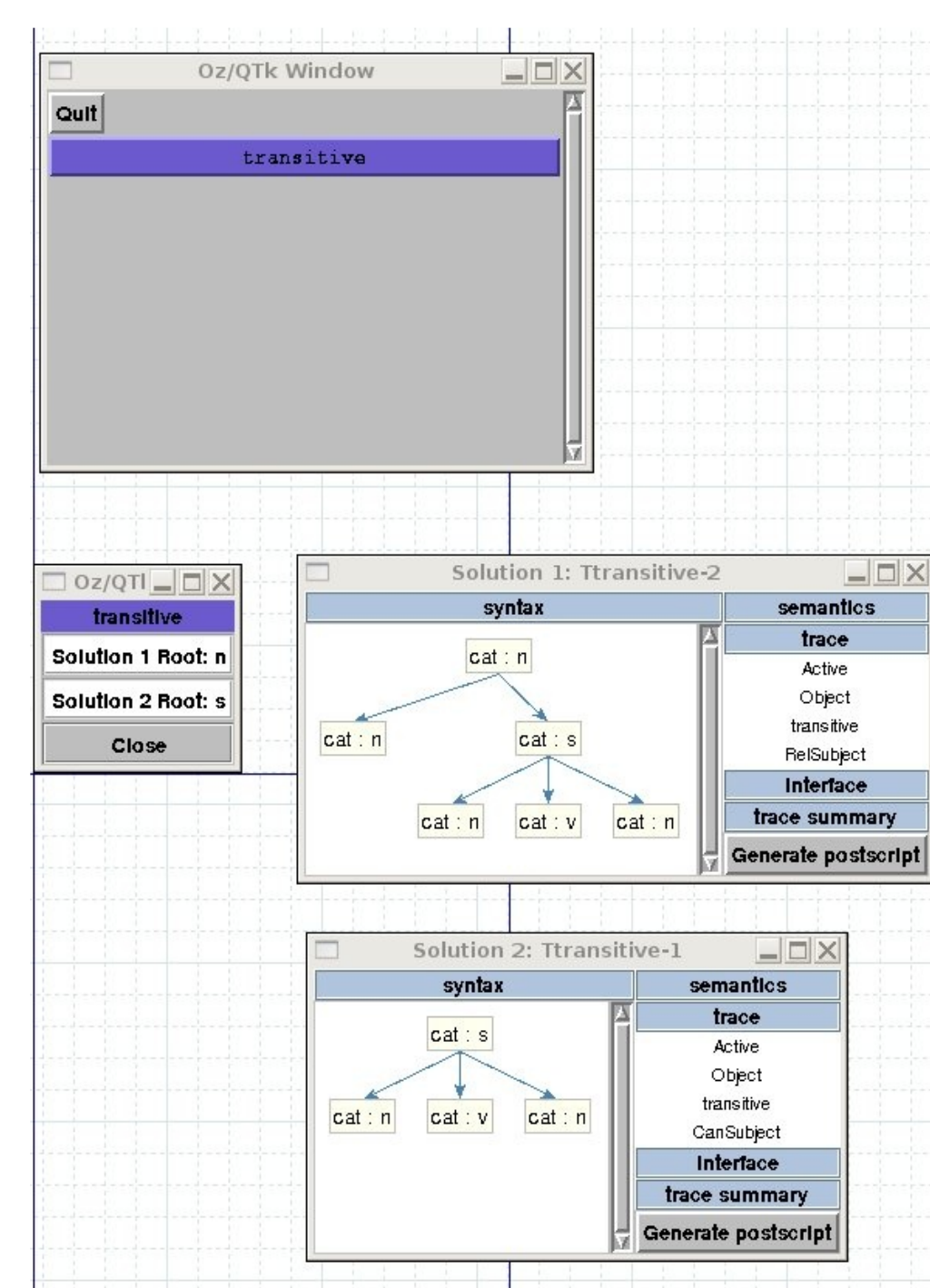
gtester benchmarking

test	expected	simple	carley
t1	il le accepter	pass	pass
t32	il nous accepter	pass	pass
t83	le ingénieur le lui apprendre	pass	FAIL
t114	le ingénieur nous le présenter	pass	pass
t145	le ingénieur vous le apprendre	pass	pass
t180	vous venir	pass	pass

gtester regression testing

(META) GRAMMAR DEVELOPMENT

- Formalism used to describe the grammar: extensible MetaGrammar (XMG) (Duchier et al., 04)
- High degree of factorisation:
 - Classes containing tree fragments (Tree Logic) are
 - (1) organised within an inheritance hierarchy and
 - (2) combined either conjunctively or disjunctively.
 - sophisticated treatment of identifiers scope (management of name spaces using import/export declarations).
- Integration of a semantic dimension: possibility to associate each tree with a flat semantic formula.



XMG

- This formalism and the corresponding implementation (MetaGrammar compiler) have been used to develop a core TAG for French (Crabbé, 05). The resulting grammar has been extended to include semantics (Gardent, 06).

- The XMG system is implemented using efficient techniques borrowed from Logic Programming (e.g. Warren's Abstract Machine). A TAG having 6,000 trees is compiled in 15 min with a P4 2.6 Ghz and 1 Gb RAM.

BENEFITS

- Shared resources for different tasks (parsing, generation)
- Advanced grammar debugging (using both parser and generator)
- Useful for building linguistic resources (Redwoods-esque treebanks?)
- New "combined" algorithms (interleaved parsing/generation)

GET SEMTAG (OPEN SOURCE)

<http://trac.loria.fr/~semtag>