Software and the digital humanities (N. Derschowitz & W. Sack)
Institut d'Études Avancées Paris 4 novembre 2015

GRAIL

A CATEGORIAL PARSER FOR SYNTAX AND SEMANTICS

Discourse on software vs software on discourse

A logician's approach: discourse on software on discourse

CHRISTIAN RETORÉ UNIVERSITÉ DE MONTPELLIER & LIRMM
(WITH **RICHARD MOOT** CNRS LABRI & LIRMM)

UNIVERSITÉ DE MONTPELLIER

FROM TEXT TO LOGIC

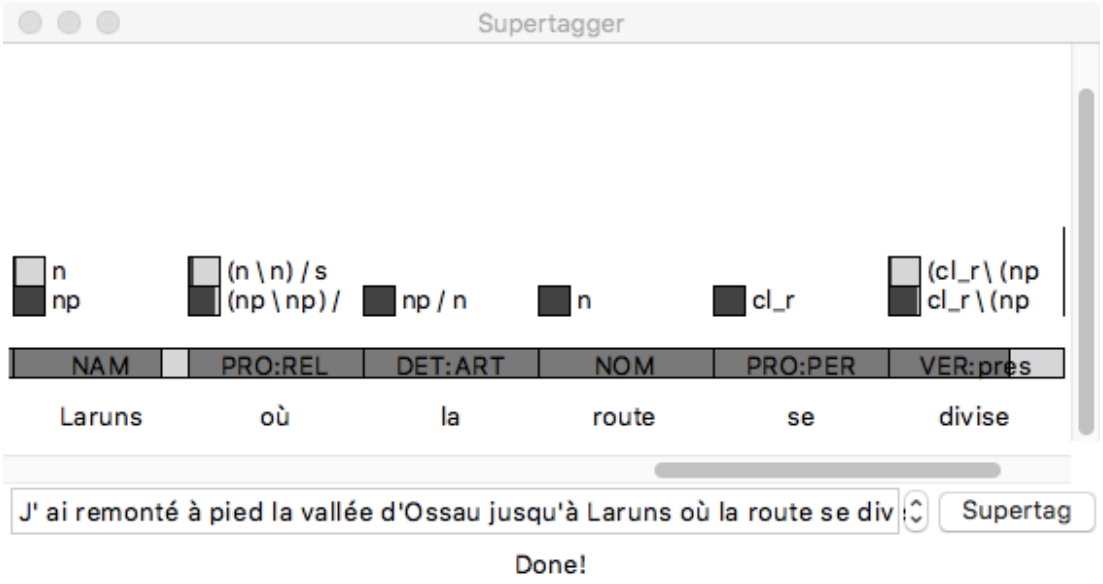
- An example:
  - Children will have a pizza.
  - For all child x there exists a pizza ordered for x
  - There exists a pizza p ordered for all children.
- **what is asserted** ≠ what it speaks about
  - Ok it's windy but that's not an ouragan.
  - Was Geach a student of Wittgenstein?
    En 1941, IL épousa la philosophe Elizabeth Anscombe, grâce à LAQUELLE IL entra en contact avec Ludwig Wittgenstein. Bien qu'IL N'ait JAMAIS suivi l'enseignement académique de CE DERNIER, cependant IL EN éprouva fortement l'influence.

WHAT DOES SEMANTICS LOOKS LIKE

$$\left\langle \left\{ \begin{array}{|l|}\hline x_0\ y_0 \\ \hline \textbf{nommé}(y_0, \textbf{Ossau}) \\ \textbf{d'}(x_0, y_0) \\ \textbf{vallée}(x_0) \\ \textbf{lieu}(x_0) \\ \hline \end{array} , \begin{array}{|l|}\hline z_0 \\ \hline \textbf{nommé}(z_0, \textbf{Laruns}) \\ \textbf{lieu}(z_0) \\ \hline \end{array} , \begin{array}{|l|}\hline x_1 \\ \hline \textbf{route}(x_1) \\ \hline \end{array} \right\} , \begin{array}{|l|}\hline y_1\ z_1\ x_2\ y_2 \\ \hline \textbf{à}(y_1, z_1) \\ \textbf{pied}(z_1) \\ \textbf{orateur}(x_2) \\ \textbf{remonter}(y_1, x_2, x_0) \\ \quad \begin{array}{|l|}\hline z_2 \\ \hline \textbf{temps}(y_1) < \textbf{temps}(z_2) \\ \textbf{temps}(z_2) \circ \textbf{maintenant} \\ \hline \end{array} \\ \textbf{lieu}(z_0, y_2) \\ \textbf{se\_diviser}(y_2, x_1) \\ \textbf{temps}(y_2) \circ \textbf{maintenant} \\ \textbf{jusqu'}(y_1, z_0) \\ \hline \end{array} \right\rangle$$

SYNTACTIC ANALYSIS

- Categorial Grammar
- Rules are language independent
- The lexicon makes the difference
- Word -> categories ~ logical formulae
- Parsing as deduction



$$\dfrac{\dfrac{\text{jusqu'}}{(s_{main}\backslash_1 s_{main})/pp_{\grave{a}}}\,[Lex] \quad \dfrac{\dfrac{\grave{a}}{pp_{\grave{a}}/np}\,[Lex] \quad \dfrac{\dfrac{\text{Laruns}}{np}\,[Lex] \quad \dfrac{\dfrac{\text{où}}{(np\backslash np)/s_{main}}\,[Lex] \quad \dfrac{\dfrac{\dfrac{\text{la}}{np/n}\,[Lex] \quad \dfrac{\text{route}}{n}\,[Lex]}{\text{la} \circ \text{route} \vdash np}\,[/E] \quad \dfrac{\dfrac{\text{se}}{cl_r}\,[Lex] \quad \dfrac{\text{divise}}{cl_r\backslash(np\backslash s_{main})}\,[Lex]}{\text{se} \circ \text{divise} \vdash np\backslash s_{main}}\,[\backslash E]}{(\text{la} \circ \text{route}) \circ (\text{se} \circ \text{divise}) \vdash s_{main}}\,[/E]}{\text{où} \circ ((\text{la} \circ \text{route}) \circ (\text{se} \circ \text{divise})) \vdash np\backslash np}\,[\backslash E]}{\text{Laruns} \circ (\text{où} \circ ((\text{la} \circ \text{route}) \circ (\text{se} \circ \text{divise}))) \vdash np}\,[/E]}{\text{à} \circ (\text{Laruns} \circ (\text{où} \circ ((\text{la} \circ \text{route}) \circ (\text{se} \circ \text{divise})))) \vdash pp_{\grave{a}}}\,[/E]}{\text{jusqu'} \circ (\text{à} \circ (\text{Laruns} \circ (\text{où} \circ ((\text{la} \circ \text{route}) \circ (\text{se} \circ \text{divise}))))) \vdash s_{main}\backslash_1 s_{main}}\,[\backslash E]$$

SYNTAX ➤ SEMANTICS

- Grammatical categories ~ logical structure
  - Common noun : property
  - Verb phrase: property
  - Definite noun phrase: individual

COMPOSITIONAL SEMANTICS

- Frege:
  the meaning of a compound expression is a function of the meanings of its parts
- Montague: AND of the syntactic structure
- Semantic lambda terms in the lexicon combined according to syntax …
- reduction (computation)
  -> logical formula (~meaning?)

LEXICAL SEMANTICS WORD MEANING IN CONTEXT

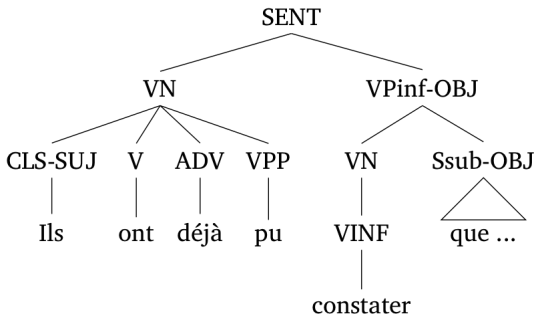
- Word meaning in context (polysemy)
  - The bank if muddy.
  - The bank resisted the crisis.
- Types & sorts to force coherence:
  - “bark” applies to animals (usally dogs)
  - * The chair barked
  - The secretary barked upon students that registered lately.
- Copredication
  - * On Sunday, Barcelona won and voted.
  - Barcelona voted and will host the Olympiads.

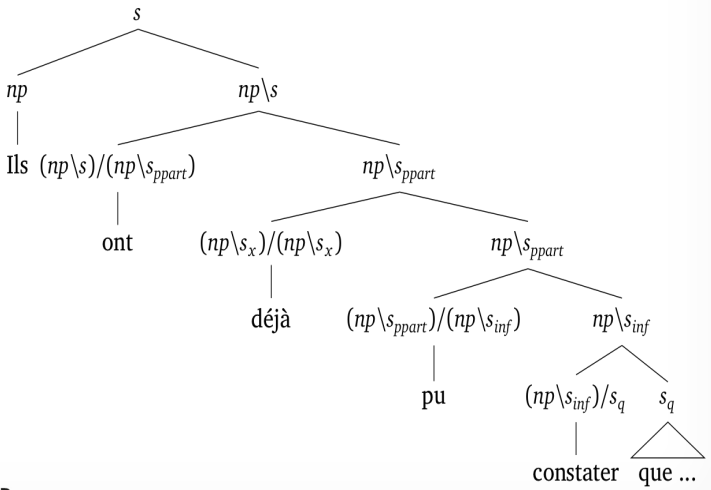
OTHER ASPECTS OF LINGUISTIC CONTEXT

- Reference and anaphora
  - Anaphora resolution (not implemented)
  - Carlos' dog thinks he does not like him. he≠him
  - If a farmer owns a donkey then he beats it. (acessibility is implemented)
- Presupposition (implemented)
  - A thinks/knows B came. => B came? no/yes

ACQUISITION OF THE LINGUISTIC RESOURCES

- Syntax (ok)
- Compositonal semantics (little)
- Lexical semantics (+/-)
- Discursive structures (little)





GRAIL FOR FRENCH

- Initially developed for spoken Dutch
- Acquired from annotated corpora

  constituents and dependencies

  corpus Paris 7 -> tigra annotations --> categories
- Up to 80 categories per word
- Supertagging
- Parsing

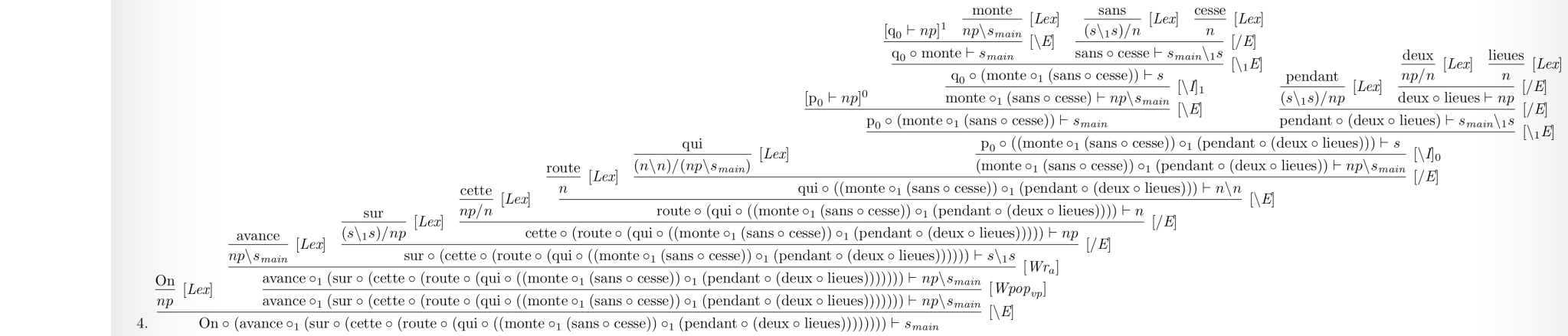
A HISTORICAL & REGIONAL CORPUS

- Travel stories in the Pyrenees
- Set of XIX century texts
- 576 334 words

- A difficult question: can we reconstruct the itineraries followed by the narrator?

CASE STUDY: FICTIVE MOTION AND GRAIL

- cette route qui monte sans cesse
- For all x if x follows the road then x keeps going up



GRAIL: NUMBERS

- Swi Prolog (interf. Tcl/Tk) acquisition+parsing 30.000 lines
- Supertagging: C software by Clark & Curran
- Corpus French Treebank 423 152 words
- Lexicon
  - 28753 distinct words
  - Up to 80 formulae per word (et, être, prepositions…)
- Semantic lexicon:
  - 350 default rules
  - 500 specific entries (coordination, quantifiers)

APPLICATION OF LOGICAL ANALYSIS

- Analysis of argumentative dialog
- Text entailment
  - One paragraph -> logical formulae Γ
  - One sentence -> a logical formula C
  - Is C a consequence of Γ ?
  - (Coq Coquand, **Huet, Dowek,…**)

RELEVANCE TO THE WORKSHOP : TRANSLATION

- Translation of text into logic (faithfulness, approximation)
  - CS translation :
    annotated corpora -> CG trees (data)
  - Sentence +supertagging -> parse tree
  - Parse tree + lexicon -> semantics

RELEVANCE TO THE WORKSHOP : TRANSLATION

- Le sens ne se produit jamais que de la traduction d'un discours en un autre. (Lacan)
- Translation of linguistic/cognitive theories into principles of computational analysis (artificial intelligence ?)
- Approximation:
  - Degrees of grammaticality, of assertion, etc.
  - « un peu » vs. « peu » (Anscombe & Ducrot)

RELEVANCE TO THE WORKSHOP

- Do there exist humanities-based approaches to software that acknowledge that software is sometimes not a tool, but rather a condition or a problem?
- Software implements/test linguistic/cognitive theories (here on syntax and semantics)

A PARTISAN PARSER

- A view of syntax, semantics, discourse theories:
  - **Language as a structure and as a vehicle for information**
  - vs. Language as opinion expression and speech acts
- Implicit claim (cf. Turing test or Chomsky): language faculty is a **computational** device