



BRIDGES AND TEXTUAL LOGICS

(PROJECT UPDATE)

Valeria de Paiva

October 2011

Fundacao Getulio Vargas (FGV)

“Fundação Getulio Vargas (FGV) is a Brazilian higher education institution founded in December 20, 1944. It offers regular courses of Economics, Business Administration, Law, Social Sciences and Information technology management. Its original goal was to train people for the country's public- and private-sector management.[...] It is considered by Foreign Policy magazine to be a top-5 "policymaker think-tank" worldwide.”
(Wikipedia)



CPDOC

**Centro de Pesquisa e
Documentação de História
Contemporânea do Brasil**
(*Contemporary Brazilian
History Research and
Documentation Center*)

Originally:

**To house personal archives
of public figures**

**Develop historic research
using privileged archive**

Documentation & Research



CPDOC

Now:

**Graduate program on
History, Political Science
and Cultural Artifacts
(since 1974)**

**School of Social Sciences
and History of the FGV
(since 2005)**

**Since 2010 EMApp (School
of Applied Mathematics)**



FGV School of Applied Math

- Just minted...
- Experts on image processing, signal/sound processing
- Not much on textual processing
- Possibility of impact...

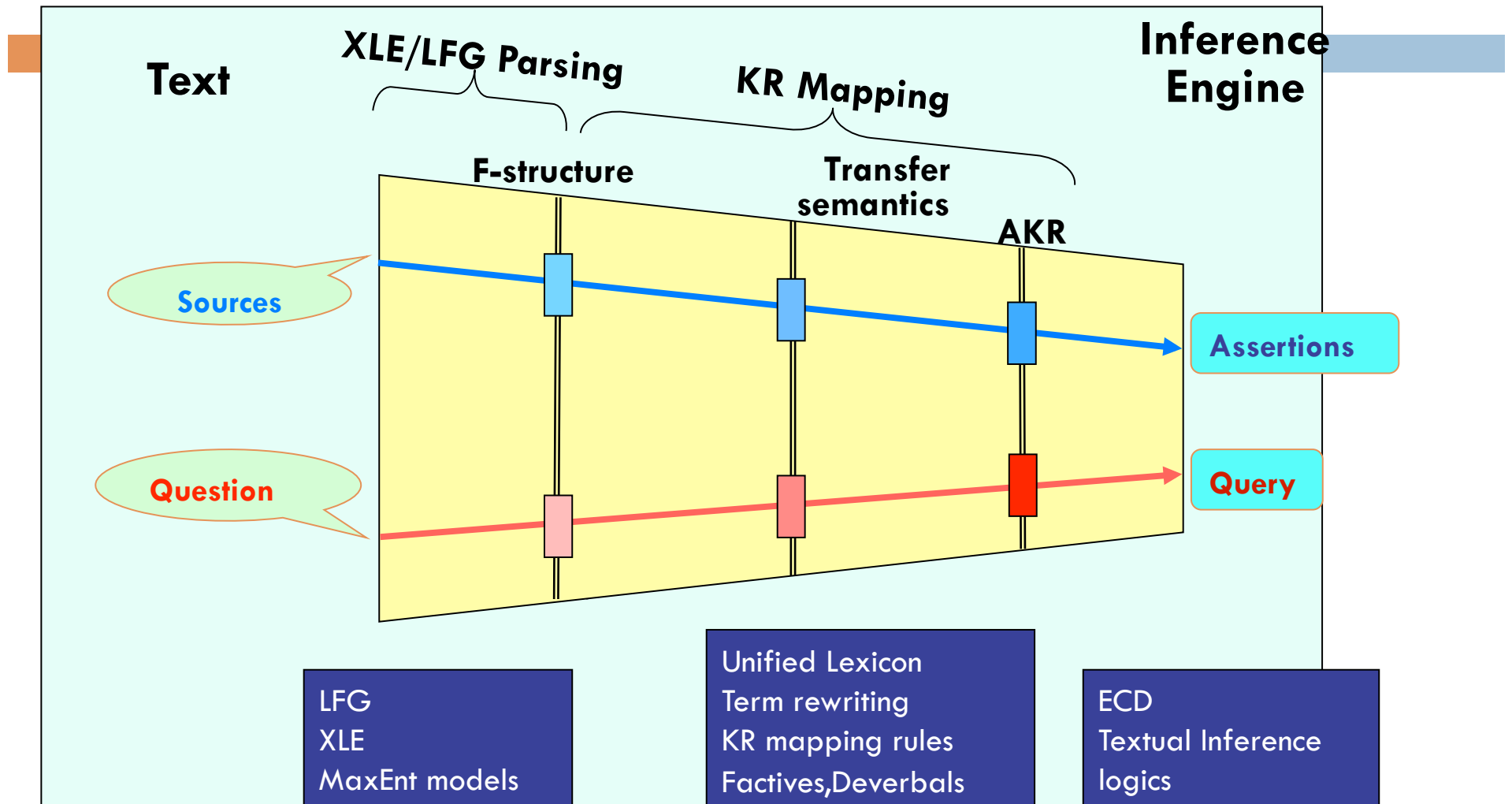


Portuguese NLP/PARSEM?

- Alexandre Rademaker, Ph D, PUC-Rio, Mar2010
- Work on Automating Description Logics
- A Bridge-like system for Portuguese?



Bridge or Just Inference?



How?

Portuguese NLP? Nah...

- Lots of Homework: Read some...
- Um panorama do Núcleo Interinstitucional de Linguística Computacional às vésperas de sua maioria Maria das Graças V. Nunes, Sandra M. Aluisio, Thiago A. S. Pardo NILC – ICMC – Universidade de São Paulo São Carlos – SP, Brasil Junho de 2010
- Pardo, T.A.S.; Gasperin, C.V.; Caseli, H.M.; Nunes. M.G.V. (2010). Computational Linguistics in Brazil: An Overview. In the *Proceedings of the NAACL-HLT Young Investigators Workshop on Computational Approaches to Languages of the Americas*, pp. 1-7. June 1-6, Los Angeles, CA/USA. [Pdf](#)
- **SBC special interest group in NLP, 2007**

Challenges



- Lack of large and robust language resources
- Lack of formal models for linguistic description and analysis of Portuguese
- Difficulty in attracting students and researchers
- Lack of multidisciplinary collaboration
- CL/NLP marginalization in both Computer Science and Linguistics.
- Poor interaction between universities and industry
- Insufficient funding.

Timeline (so far...)

- 
- Rademaker @SRI Feb-May 2009
 - ▣ PhD Mar 2010 Description Logics.
 - ▣ Pre-proposal: Jul2010(3p) PARC license?
 - De Paiva @FGV Sept 2010 course
 - ▣ announcement PARSEM Oct 2010
 - ▣ NO GO Nov 2010, Funcap perhaps?
 - Feb 2011: What if? May 1? (license...)
 - ▣ May project approved for 2011
 - ▣ July, 6 funds released: A baby & job later
 - ▣ Nov 21-25 Workshop...

Plans Now...



- Working on wordnet WN-BR using Global WordNet Association suggested approach (manuscript submitted)
- Hoping to 'project' MENTA to Portuguese, hand-clean and improve it, semi-automatically
- Gerard de Melo mini-course in Nov...
- YAGO/YAGO2/MENTA/UWN...

Plans Now...

- SUMO/SIGMA instead of/as well as Unified Lexicon?..
- <http://www.ontologyportal.org/>
- SUMO-BR started (Rearden Commerce pitstop day)
- NOMLEX-BR started...
- Linguistic work for flight + accommodation in the most beautiful city in the world?...

The image features a decorative layout with a grid of colored rectangles. A large, solid light blue rectangle occupies the top right portion of the frame. To its left is a vertical orange rectangle with a horizontal line pattern. Below the orange rectangle is a solid orange rectangle. To the right of this solid orange rectangle is a horizontal blue rectangle containing the text "Thanks!". Below the blue rectangle is another horizontal orange rectangle with a horizontal line pattern. The bottom right corner is a large orange rectangle with a horizontal line pattern.

Thanks!