Grupo
SemanTec
SEMÂNTICA E TECNOLOGIA

FRAMENET
BRASIL
COPA 2014

UNIVERSIDADE
FEDERAL DE JUIZ DE FORA
ufjf

UNISINOS

Grupo SemanTec
SEMÂNTICA E TECNOLOGIA

FRAMENET BRASIL
COPA 2014

FrameCup 2014: A Frame-based Domain-Specific Multilingual Electronic Dictionary

Rove Chishman, Tiago Torrent, Anderson Bertoldi, Margarida Salomão, Ely Matos, Cassiane Ogliari

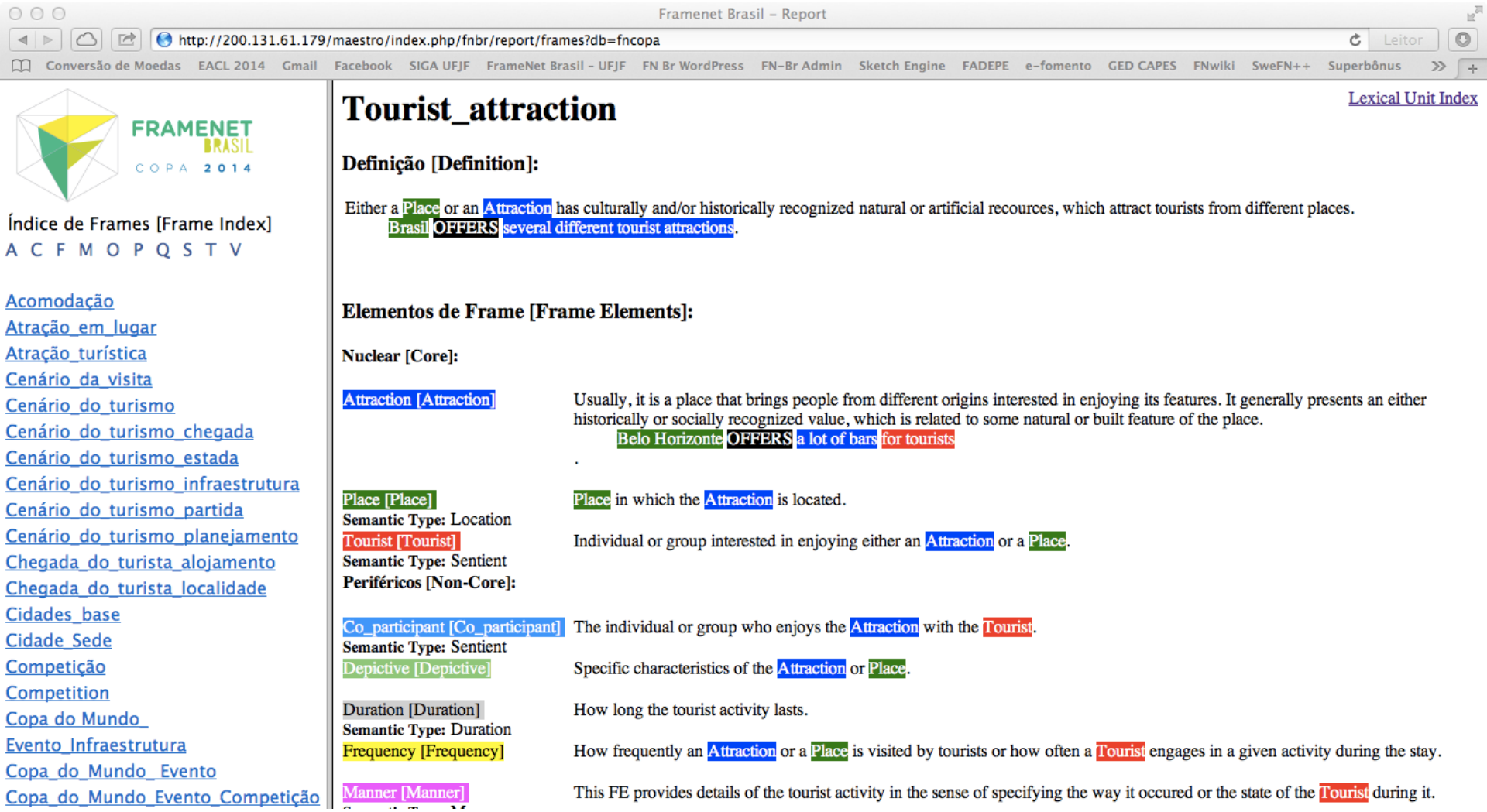
Introduction

- Exploiting the infrastructure of FrameNet (Rupenhoffer et al. 2010) and building on the methodology of the Kicktionary (Schmidt 2006, 2007), FrameCup 2014 is developing a multilingual frame-based electronic resource covering the vocabulary of soccer and tourism for Brazilian Portuguese, Spanish, and English.

Current Status

- Tourism_scenario
  - 20 frames and counting
- World_Cup_scenario
  - 19 frames and counting
- Soccer_scenario
  - 10 frames and counting
- More than 70 Lexical Units annotated

Example Frame



Research Questions

- Given that a Lexical Unit is a pairing of one form to one meaning (i. e. one frame), how to account for those lexemes that originate more than one LU in the FrameCup Dictionary?
  - How to account for polysemy?

The case of *marcar*

1. Mas [o árbitro$_{REFEREE}$] **marcou** [pênalti$_{COMPENSATION}$].
   - *But the referee awarded a penalty-kick {to some team}.*
     - Referee_decision frame
2. [O time $_{MARKER}$] **marcava** intensamente {e saía com velocidade.}
   - *The team kept their opponents intensively marked {and had a fast attack}.*
     - Mark frame
3. [Dentinho$_{SCORER}$] **marcou** [o gol de número 9.$_{GOAL}$]
   - *Dentinho scored the 9th goal.*
     - Score_goal frame

The case of *chegar*

1. [Os turistas $_{\text{TOURIST}}$] que **chegam** [ao aeroporto$_{\text{PLACE}}$]...
   – *Tourists arriving at the airport...*
      • Tourism_scenario_arriving frame

2. [O Brasil$_{\text{TEAM}}$] **chegará** [à semifinal$_{\text{PHASE}}$]?
   – *Will the Brazilian team make it to the semifinals?*
      • Playoffs_play frame

Frame Desambiguation

- Refers to the task of chosing, among the frames evoked by the various LUs of a lexeme, the best fit for a given context
  - Bases:
    - Word Sense Desambiguation
    - Conditional Random Fields
    - FrameNet Frame-to-Frame Relations

Frame Desambiguation

O turista que chega ao aeroporto..

- Text-matching techniques identify LUs in the sentence.

x [Target A] xxx [Target B] xx [Target C]...

- Frames evoked by each LU are retrieved

A= {F_a1, F_a2, …, F_aN}; B= {F_b1, … F_bN}; C={…

- Each possible combination of frames forms a cluster

Frame Desambiguation

- Since it is very unlikely that two or more LUs evoke the same frames in one sentence, and since frames in FrameNet are related to each other via Frame-to-Frame relations, cluster formation is extended so as to include those frames
- Similarity / Proximity between frames is assessed within each cluster

Frame Desambiguation

- Cluster assessement can be carried out taking into consideration 1, 2 or 3 levels in the FrameCup Database Hierarchy
- Frame-to-frame relations receive different weights
  - same frame = 1.0
  - Perspective = 1.0
  - Inheritance = 0.9
  - Subframe = 0.7
  - Using = 0.3

Frame Desambiguation

FRAMENET BRASIL

Banco de Dados: -- Selecione --

Login:

Search Tests

Lexema	Lexema	Lexema	Níveis (1,2 ou 3)
turista	chegar	aeroporto	2

Search1 | Search2 | Search3

Search1: Considera apenas cada frame evocado pelas LUs

Search2: Considera os super/sub frames de cada frame evocado pelas LUs (em 1 nivel)

Search3 (não implementado ainda): Considera os frames referenciados pelos Elementos de Frame de cada frame evocado pelas LUs

---

Níveis (1,2, ou 3): Nível do relacionamento a ser considerado - o peso é decrescido em 10% a cada nível, a partir do 2. nível

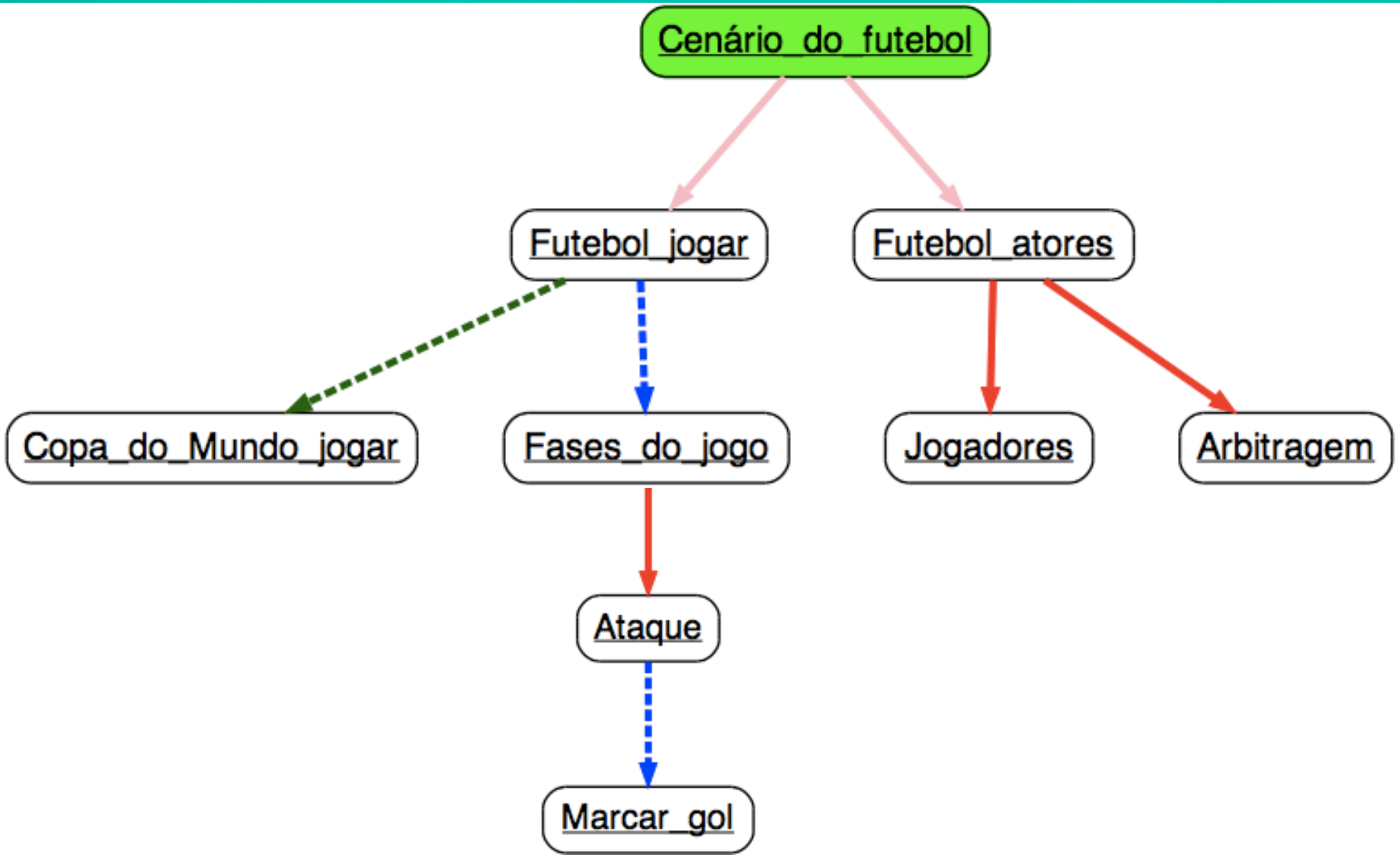
---

Rank (excluding zeroes) - 2 cluster(s)	1 Result(s)
[1] : Turismo_por_turista: Cenário_do_turismo_chegada: Transporte	Cenário_do_turismo_chegada
[0] : Turismo_por_turista: Mata_Mata_Jogar: Transporte	

Results

Sentence	Clusters	Playoffs_play	Tourist_arrival
Brasil chegar semifinal *Brasil make it semifinal*	Teams; Playoffs_play; Semifinals Teams; Tourist_arrival; Semifinals	**1.0**	0.0
Turista chegar aeroporto *Tourist arrive airport*	Touring; Playoffs_play; Transit Touring; Tourist_arrival; Transit	0.0	**1.0**
Sentence	**Clusters**	**Referee_decision**	**Score_goal**
Juiz marcar pênalti *Referee award penalty*	Referee; Score_goal; Score_goal Referee; Referee_decision; Score_goal	**1.0**	0.0
Atacante marcar gol *Striker score goal*	Players; Score_goal; Score_goal Players; Referee_decision; Score_goal	0.225	**0.774**

Results


Cenário_do_futebol
Futebol_jogar
Futebol_atores
Copa_do_Mundo_jogar
Fases_do_jogo
Jogadores
Arbitragem
Ataque
Marcar_gol

Contributions

- Incorporates contextual information to improve results for users interested not in all the possible meanings of a given word, but, instead, in the meaning of a given word relative to the context in which it appears
- Provides one more test case for the usefulness of the FrameNet hierachy for modeling human knowledge
- Creates an electronic dictionary which accounts for polysemy based not exclusively in the lexicographer's intuition

Further Developments

- Implement the Frame_Element-to-Frame relations, so as to improve the accuracy of cluster assessment
- Implement Valence Pattern analysis, so as to account for Perspective relations properly
  - Compare:
  - The coach substituted Pelé vs. Pelé substituted Garrincha
- Reevaluate weight values based on experimentation with training corpora

Further Developments

Questions?