

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

Rove Luiza de Oliveira Chishman<sup>1,3</sup>, Maria Margarida Martins Salomão<sup>2</sup>, Anderson Bertoldi<sup>1,4,5</sup>, Tiago Timponi Torrent<sup>2</sup>, Ely Edison da Silva Matos<sup>2</sup>, Cassiane Ogliari<sup>1</sup>

<sup>1</sup>UNISINOS, <sup>2</sup>Federal University of Juiz de Fora, <sup>3</sup>CNPq, <sup>4</sup>CAPES, <sup>5</sup>FAPERGS

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. The present paper reports on the treatment of polysemy in FrameCup 2014, and focuses on the Brazilian Portuguese verb *marcar*. More specifically, this work introduces the scoring of frame elements (FEs) and lexical units (LUs) to incorporate contextual information, also providing users with a probabilistically more likely sense of polysemous words.

To illustrate, consider the polysemy of the Brazilian Portuguese verb *marcar*, for which we have determined (at least) the following three senses: it refers to an activity that a referee performs, and evokes the **Referee\_decision** frame, as in (1); it identifies an activity that the teams' defense performs, evoking the **Mark** frame, as in (2); and it denotes an activity that the scorer performs, which evokes the **Score\_goal** frame, as in (3).

(1) Mas [o árbitro Paulo César Oliveira REFeree] **marcou** [pênaltiCOMPENSATION].  
but the referee Paulo César Oliveira award-PAST.3SG penalty  
*But the referee awarded a penalty-kick {to some team}.*

(2) [O time MARKER] **marcava** intensamente {e saía com velocidade}.  
the team mark-PAST.3SG intensively  
*The team kept their opponents intensively marked {and had a fast attack}.*

(3) [DentinhoSCORER] **marcou** [o gol de número 9. GOAL]  
Dentinho score-PAST.1SG the goal of number 9  
*Dentinho scored the 9th goal.*

FrameCup 2014 annotates corpus data for frames and frame elements and scores each FE and each LU based on that annotation to facilitate (a) searching for lexical units that evoke the frames for soccer and tourism and (b) finding suitable translation equivalents for polysemous LUs based on the context provided in a user query.

Thus, when users search for *marcar*, the dictionary can present the three possibilities, defined in terms of the three different frames presented above. Alternatively, if users enhance the query by searching for example sentences with the verb *marcar*, the dictionary can provide the most probable meaning of the polysemous verb in the given context. To achieve the latter, FrameCup 2014 associates two additional pieces of information with FEs in the database. First, semantic types for FEs have the same computational representation developed for frames, which enables the scoring system to check whether there is a relation between the words in the query in the frame level, and, if so, in regards to which of the frames evoked by the target LU such relation is stronger. Next, each FE is scored on a scale from 0 to 1, according to its relevance in the scene that the frame characterizes. According to such methodology, core FEs are scored higher than peripheral FEs, which, by their turn, are scored higher than extra-thematic FEs. Also, the syntactic properties of FEs are taken into consideration, thus, FEs used as external arguments or objects tend to be scored higher than those used as dependents. For example, in the **Score\_Goal** frame, the Scorer is evaluated as 1 because, in addition to being a core FE, it is usually the external argument of the LU that evokes the frame. The system performs the same kind of evaluation for each lemma in the dictionary user's query. The query answering system analyzes the whole database and calculates which frame (or frames) best fit the query.

This work contributes to the larger enterprise of developing frame-based multilingual electronic resources, and offers a means of incorporating 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.

**Keywords:** Frame Semantics; Polysemy; FrameNet; Domain-Specific Dictionary; Context-Based Query