

Os desafios da anotação de um *corpus* da linguagem do futebol com *frames* semânticos

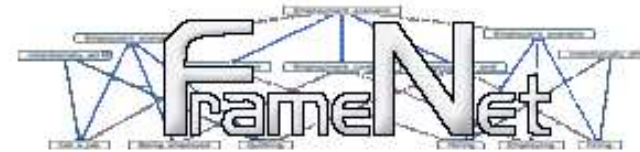
Rove Chishman
Anderson Bertoldi
João Padilha
Diego Spader

Programa de Pós-Graduação em Linguística Aplicada (UNISINOS)



Questões introdutórias:

- O objetivo deste trabalho foi avaliar a aplicabilidade das etiquetas semânticas desenvolvidas pelo projeto Kicktionary para a anotação de um *corpus* de linguagem do futebol.
- O Kicktionary é uma base de dados lexicais multilíngue da linguagem do futebol, em inglês, francês e alemão.
- No Kicktionary, as unidades mínimas de descrição são a **cena**, o **frame** e a **unidade lexical**.



Home

FrameNet Search

FrameNet Data Search for pass

Frame search results: Closest match is pass

Passing, Passing_off

Lexical unit search results: Closest match is pass

Lexical Unit	Frame	LU Status	Lexical Entry Report
Passover Eve.n	Commemorative	Needs_SCs	LE
pass away.v	Death	Finished_Initial	LE
pass off.v	Passing_off	Rules_Defined	LE
pass on.v	Dispersal	Created	LE
pass out.v	Fall_asleep	Created	LE



Kicktionary

The multilingual electronic dictionary of football language.

HOME DIC

Dictionary

- Lexical Units
- Scenes
- Concept Hierarchies
- Sound Examples
- Parallel Texts

Scenes

On the pitch	
<p>Shot</p> <p>12 frames 269 lexical units</p>	<p>Ball_And_Goal (11 LUs) Feign (6 LUs) Finish (5 LUs) Follow_Up (7 LUs) Goal_Kickoff (7 LUs) Hit (4 LUs) Shoot_At (5 LUs) Shot (123 LUs) Shot_Supports (25 LUs)</p>
<p>Pass</p> <p>11 frames 226 lexical units</p>	<p>Bad_Pass (2 LUs) Being_Free (14 LUs) Connect (5 LUs) Control (37 LUs) Flick_On (7 LUs) Intercept (11 LUs) Pass_Combination (21 LUs) Supply_Pass (12 LUs)</p>
<p>Goal</p> <p>10 frames 148 lexical units</p>	<p>Award_Goal (11 LUs) Celebrate_Goal (5 LUs) Concede_Goal (7 LUs) Convert_Chance (8 LUs) Goal (85 LUs) Own_Goal (4 LUs) Prepare_Goal (6 LUs) Score_Goal (7 LUs)</p>
<p>One_On_One</p> <p>7 frames 129 lexical units</p>	<p>Beat (49 LUs) Challenge (18 LUs) Deny (19 LUs) Lose_Ball (7 LUs) One_On_One (25 LUs) Take_On (5 LUs)</p>
<p>Foul</p> <p>12 frames 211 lexical units</p>	<p>Advantage (4 LUs) Concede_Compensation (4 LUs) Dissent (12 LUs) Foul (54 LUs) Give_Card (8 LUs) Offs Referee_Decision (12 LUs) Sanction (36 LUs) Set_Piece (45 LUs) Simulation (6 LUs) Win_Compensation (6 LUs)</p>



Dictionary

- Lexical Units
- Scenes
- Concept Hierarchies
- Sound Examples
- Parallel Texts

Pass Scene

Frames

- A pass [Pass](#) [Pass_Back](#)
- The path of the pass
- At the recipient's end [Mark](#) [Being_Free](#) [Control](#) [Connect](#) [Flick_On](#)
- Intercepting a pass [Intercept](#) [Bad_Pass](#)
- Others [Pass_Combination](#) [Supply_Pass](#)
- Related frames [Set_Piece](#) [Goal_Kickoff](#)

Description

A pass



Dictionary

- Lexical Units
- Scenes
- Concept Hierarchies
- Sound Examples
- Parallel Texts

pass (noun) 

Scene *Pass* Frame *Pass*

Examples

1. However, 12 minutes later Jan Polák, winning his fourth Czech cap as a replacement for injured captain Tomáš Galásek, cleared the empty net after [Rosický]_{PASSER} beat two players and supplied the decisive **pass** .
2. A minute past the half-hour, however, the home team did find an equaliser, as the Romanian defence failed to deal with M... who was left alone in front of Bogdan Lobont's goal to make it 1-1.
3. Midway through the half, [Georgiou]_{PASSER} was again the instigator, threading a perfectly weighted **pass** to Ioannis Okkas, but...
4. Moments after the restart Milan doubled their lead as [Rui Costa]_{PASSER} threaded a **pass** [between three defenders]_{PATH} [t... [1077219 / p8]
5. [Kuijt]_{PASSER} chipped a sublime **pass** [from the right]_{SOURCE} [to Kalou]_{RECIPIENT} who scored with a diving header into the bottom...

Semantic relations

Synonyms	Abspiel Anspiel Pass Zuspier pass ball balle passe
Hypernyms [Moving_Balls]	<i>Moving_Ball</i>

Frame elements

PASSER [Player] **PATH** [On_The_Field_Location] **RECIPIENT** [Player] **SOURCE** [On_The_Field_Location]

Support	LU	PASSER	PATH	RECIPIENT
<i>supplied</i>	pass	Rosický		
	pass	Krstev's		to Maz
<i>threading</i>	pass	Georgiou		
<i>threaded</i>	pass	Rui Costa	between three defenders	to Cre
<i>chipped</i>	pass	Kuijt		to Kal

Metodologia:

- (a) tradução das cenas, frames e unidades lexicais propostos no Kicktionary para o desenvolvimento de um manual de apoio à anotação;
- (b) criação de um *corpus* especializado;
- (c) segmentação do *corpus* em sentenças;
- (d) análise e anotação das sentenças por pares de anotadores;
- (e) anotação em editor.

Corpus:

- o *corpus* é composto de 100 arquivos de texto, contendo descrições de partidas de futebol, totalizando 3.307 sentenças;
- os textos foram coletados de *sites* de clubes brasileiros em dias subsequentes às partidas de campeonatos, como a Copa do Brasil e a Copa Libertadores;
- a escolha por criar um *corpus* contendo descrições de partidas de futebol foi guiada pela necessidade de se encontrar um *corpus* representativo para anotação de *frames* do futebol.

Anotação:

- A partir de uma apreciação geral do *corpus*, percebeu-se a complexidade das sentenças a serem anotadas.
- As sentenças apresentavam muitos verbos suporte, verbos metafóricos e polissêmicos.
- Também se percebeu a grande recorrência das cenas Chute, Gol e Passe.
- Então, optou-se por anotar, primeiramente, apenas as três cenas mais recorrentes.

- Para a anotação das sentenças, primeiramente, foram identificados o evocador de *frame* na sentença e o *frame* evocado.
- Em seguida, os principais sintagmas da sentença foram anotados com elementos de *frame*:

Cena: Substituição/Frame: Substituir

Em seguida, [Ricardinho_{JOGADOR_SUBSTITUÍDO}] **foi substituído** [pelo
estreante Neto Berola_{SUBSTITUTO}]

Anotação da Cena Passe:

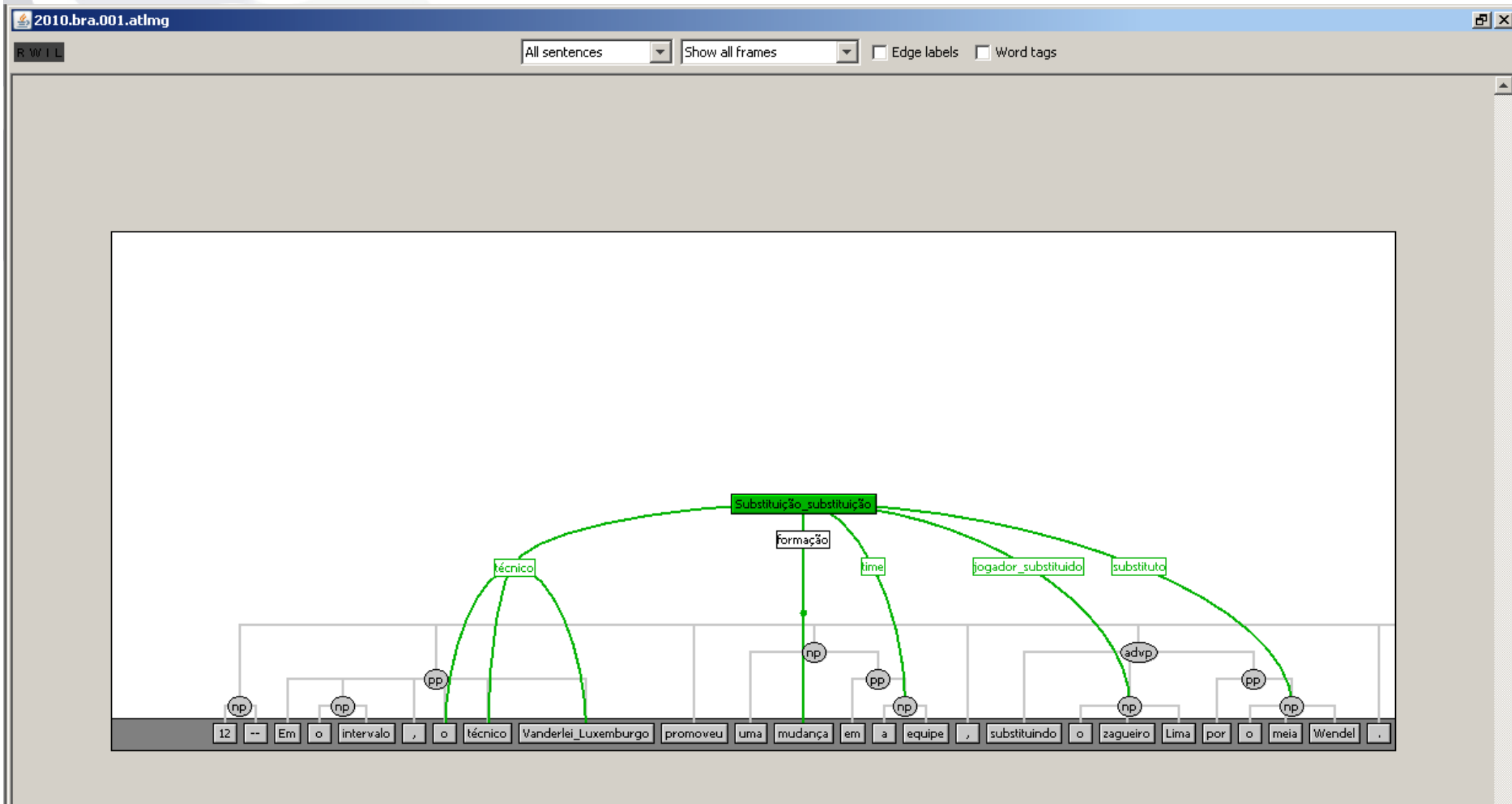
- Sentenças anotadas: 159
- Passe: 96
- Dominar: 48
- Interceptar: 39
- Combinação_de_passe: 16
- Estar_livre: 06
- Rolar: 05
- Fornecer_passe: 02
- Marcar: 0
- Conectar: 0
- Devolver_o_passe: 0

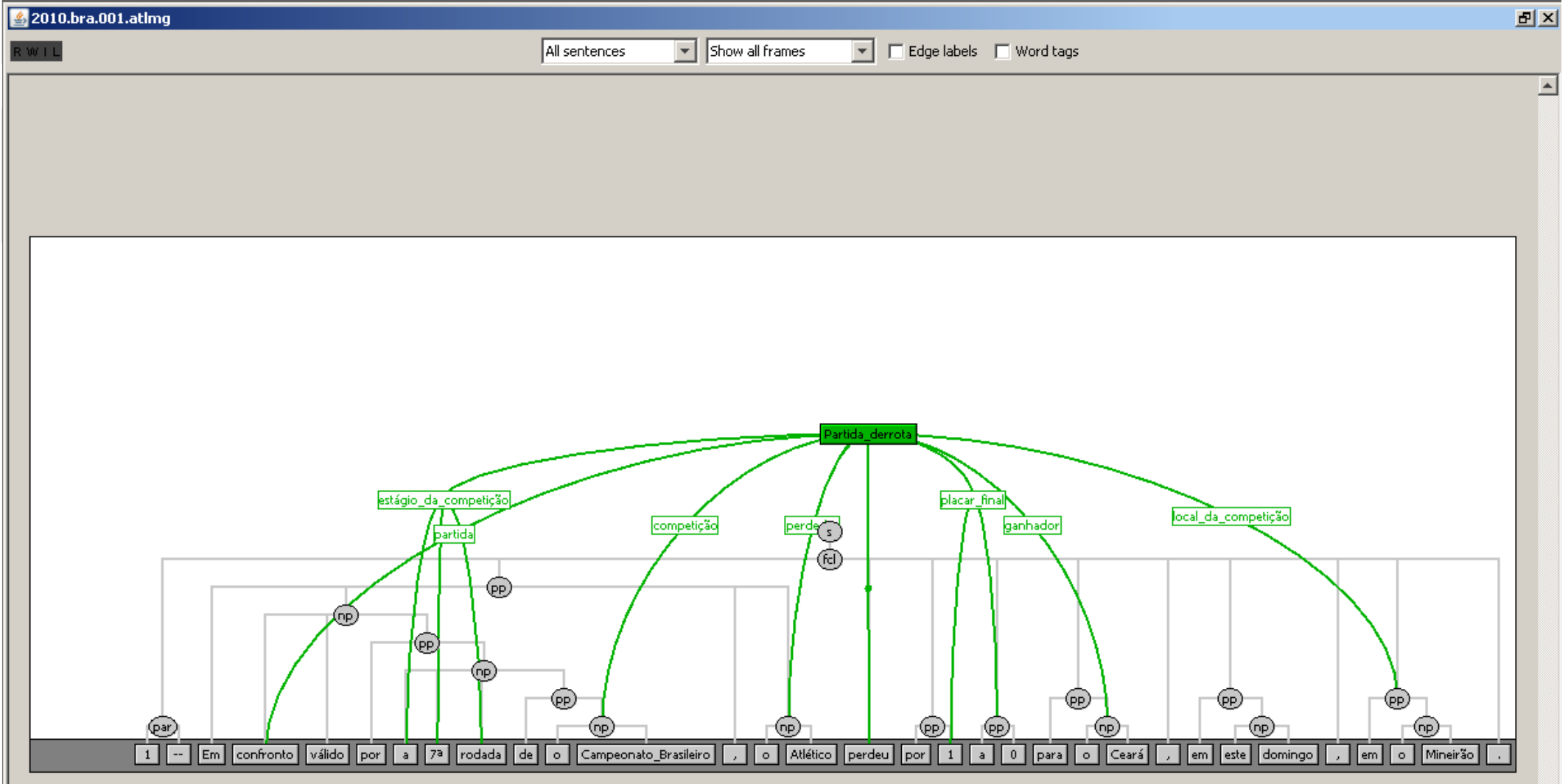
Anotação da Cena Chute:

- Sentenças anotadas: 203
- Chute: 147
- Defender: 72
- Finalizar: 48
- Intervenção: 17
- Rebote: 9
- Tipos_de_chute: 4
- Acertar: 2
- Chutar_a: 2
- Tiro_de_meta: 1

Anotação da Cena Gol:

- Sentenças anotadas: 156
- Conceder_gol: 1
- Comemorar_gol: 6
- Sofrer_gol: 3
- Gol: 139
- Vencer_goleiro: 5
- Gol_contra: 1
- Preparar_gol: 1
- Chance_de_gol: 0





Metáforas:

- As metáforas no *corpus* de linguagem do futebol são os casos que requerem especial atenção dos anotadores. Nesses casos, para se identificar o *frame* evocado, é necessário interpretar o significado metafórico da expressão linguística.

Cena: Passe/Frame: Interceptar

[Camacho_{INTERCEPTADOR}] **roubou** [grande bola_{BOLA}] [no meio_{LOCAL_DA_INTERVENÇÃO}] e deixou Diego Maurício livre para chutar colocado, mas a bola passou raspando a trave.

Cena: Passe/Frame: Interceptar

Mas a [zaga_{INTERCEPTADOR}] **cortou**.

Verbos-suporte:

- No caso dos verbos-suporte, o evocador de *frame* não é o verbo, como habitual, mas um substantivo.
- Nos exemplos abaixo, o item lexical *passe* incorpora o elemento de *frame* ao evocador de *frame*.

Cena: Passe/Frame: Fornecer_passe

[Ramirez JOGADOR_QUE_PASSA] fez [**o passe** PASSE] [para Batti6n JOGADOR_QUE_RECEBE].

Cena: Passe/Frame: Fornecer_passe

[O autor do primeiro gol JOGADOR_QUE_PASSA] deu [**passe milim6trico** PASSE] [para Adriano JOGADOR_QUE_RECEBE] apenas tocar na sa6da do goleiro chileno e fazer 2 a 0 aos 13.

Polissemia:

- A polissemia dos itens lexicais também representou um desafio para a anotação. Nos exemplos abaixo, pode-se ver diferentes sentidos para o verbo *tocar*.

Cena: Chute/ Frame: Chute

Aos sete minutos, [Tardelli_{JOGADOR_QUE_CHUTA}] recebeu ótimo passe de Muriqui e [**tocou** na saída do goleiro_{ALVO}] para fazer o terceiro gol: Galo 3 x 1.

Cena: Passe/ Frame: Passe

[O Furacão_{JOGADOR_QUE_PASSA}] tentou **tocar** mais a [bola_{BOLA}] para tirar a velocidade do Timão, mas não adiantou.

Metonímia:

Cena: Passe / Frame: Passe

Atuando de forma segura, [**o time atleticano** JOGADOR_QUE_PASSA] **tocava** bem [a bola BOLA] e praticamente não era ameaçado pelo Sport, que chegava apenas em alguns cruzamentos.

Cena: Passe / Frames: Passe e Interceptar

O Galo tentou reagir no **cruzamento** de [Muriqui JOGADOR_QUE_PASSA] [pela direita DIREÇÃO], **cortado** pela [**defesa do Fluminense** INTERCEPTADOR].

Cena: Passe / Frames: Passe e Interceptar

No ataque seguinte, o último da etapa inicial, [Tardelli jogador_que_passa] disparou [pela direita direção] e **crizou** para novo **corte** da [**zaga tricolor** interceptor].

Dificuldades encontradas:

- **Vários frames pertencentes a uma mesma cena:**

Cena: Passe / Frames: Dominar/Rolar/Passe/Interceptar

[Dentinho JOGADOR_QUE_RECEBE] **recebeu** pela esquerda e **rolou** para [Roberto Carlos JOGADOR_QUE_RECEBE], que foi [à linha de fundo DIREÇÃO] e **cruzou** forte; [a zaga uruguaia INTERCEPTADOR] **afastou** parcialmente (...).

- **Ambiguidade na evocação:**

Cena: Passe / Frame: Passe Cena: Falta / Frame: Jogada_de_bola_parada

Aos 39 minutos, [Coelho JOGADOR_QUE_PASSA] **cruzou** em cobrança de **falta** [pela direita DIREÇÃO] e a [bola BOLA] atravessou a grande área.

- **Mais de uma cena:**

Cena: Passe / Cena: Chute / Cena: Falta

No ataque seguinte, [Muriqui JOGADOR_QUE_RECEBE] **recebeu** a [bola BOLA] [na área LUGAR] e **completou** para o gol, mas o árbitro **marcou** falta no toque de mão involuntário do atacante.

Considerações finais:

- O trabalho aqui relatado, não se reduziu à tarefa técnica de anotar, mas permitiu a reflexão sobre a própria estrutura do Kicktionary, organizado por cenas e frames.
- Essa organização em cenas e frames, bem como a não disponibilização da base de dados do Kicktionary em xml, causaram dificuldades no uso da ferramenta de anotação SALTO.
- Apesar de ser um trabalho de anotação de *corpus* monolíngue, foram enfrentadas dificuldades relacionadas à multilinguagem, tais como padrões de lexicalização e equivalência de tradução.
- As marcas de oralidade e de informalidade do *corpus* brasileiro levaram a reflexões acerca da delimitação entre termos e “gírias” do futebol e a sua equivalência em inglês.

- Os desafios enfrentados na anotação de um *corpus* monolíngue de linguagem do futebol levam a reflexões acerca da criação de léxicos bilíngues para o domínio do futebol.
- Um trabalho de desenvolvimento de léxicos bilíngues necessita encontrar equivalência para expressões metafóricas e para as expressões com verbo-suporte, além de dar conta da polissemia das línguas.
- O Kicktionary_Br, projeto monolíngue de anotação de *corpus*, está se desdobrando no projeto Kicktionary_Br Copa 2014.
- O projeto Kicktionary_Br Copa 2014, a ser realizado em parceria com a UFJF, está desenvolvendo um recurso lexical bilíngue, em inglês e português, da linguagem do futebol, a ser utilizados por leigos durante a copa do mundo de 2014.

Contatos:

- Rove Chishman:
rove@unisinos.br
- Anderson Bertoldi:
andersonbertoldi@yahoo.com
- Grupo de Pesquisa SemanTec:
<http://projeto.unisinos.br/semantec/index.php/pt/>