

ELC - 2011 X Encontro de Linguística de Corpus

EBRALC - 2011 - V Escola Brasileira de Linguística Computacional



Contribuições metodológicas para o desenvolvimento da plataforma FrameNet Brasil: a descrição de algumas Unidades Lexicais dos frames FECHAMENTO e MOVIMENTO_CORPORAL

Gabriela da Silva Pires- UFJF

Maria Margarida Martins Salomão- UFJF (orientadora)

Faculdade de Letras

Programa de Pós-Graduação em Estudos Linguísticos

Novembro/2011

Objetivos desta apresentação no X ELC

- Evidenciar a importância de dados reais da língua para a construção de um dicionário *online* baseado na Semântica de Frames
- Destacar etapas do processo metodológico concernentes à utilização de *corpora* eletrônicos
- Destacar a relevância do trabalho no diálogo entre Lexicografia e Linguística Cognitiva/Semântica de Frames

Escopo do Trabalho- Em termos Metodológicos

→ Documentação detalhada dos procedimentos para a implantação do Projeto FrameNet Brasil como metodologia de descrição lexicográfica e como plataforma de apresentação dos dados que compõem esta descrição

Escopo do Trabalho-

Em termos Lexicográficos

→ Descrição lexicográfica de **sete** Unidades Lexicais do frame FECHAMENTO:

→ i) Quatro ULs monolexêmicas:

Desabotoar, Desarrolhar, Desatarraxar, Destampar

→ ii) Três ULs polilexêmicas:

Abrir_((tampa)), Levantar_((tampa)), Tirar_((tampa))

→ Descrição lexicográfica de **três** Unidades Lexicais polilexêmicas do frame MOVIMENTO_CORPORAL:

Abrir_((boca)), Abrir_((mão)), Abrir_((olho))

A FrameNet Americana: Projeto-Mãe

- O projeto FrameNet, desenvolvido no International Computer Science Institute (ICSI) da Universidade da Califórnia - USA), é um projeto de **lexicografia computacional** que extrai, de extensos *corpora* eletrônicos (cuja base é o BNC – *British National Corpus*), informação sobre as propriedades semântico-sintáticas de Unidades Lexicais (ULs) do inglês. Essas propriedades emergem de anotações feitas manualmente que, consolidadas por ferramentas computacionais, geram relatórios de resultados, acessáveis via Internet através do site www.framenet.icsi.berkeley.edu

Aplicações

- **Para estudantes:** é um dicionário que fornece os sentidos das palavras com base em exemplos anotados, mostrando o sentido e uso das palavras e/ou expressões
- **Para pesquisadores em Processamento de Linguagem Natural:** possui sentenças anotadas manualmente, o que fornece uma base de dados para treinamento em etiquetagem semântica, uso em extração de informação, tradução etc.
- **Para alunos e professores de Linguística:** serve como dicionário de valências, com evidências sobre as propriedades de combinação dos vocábulos descritos.
(informações adaptadas do site FrameNet em Berkeley)

Página inicial da FrameNet

http://framenet.icsi.berkeley.edu/

ELC 2011 - X Encontro de Li x Welcome to FrameNet | fnd x Welcome to FrameNet | fnd x

← → ↻ <https://framenet.icsi.berkeley.edu/fndrupal/> ☆

FrameNet

[ACCOUNT](#) [FRAMENET](#) [FRAMENET DATA](#) [BIBLIOGRAPHY](#) [FORUMS](#) [HOME](#)

RECENTLY CHANGED FRAMES

- Experiment
- Capital_stock
- Commerce_collect
- Records
- Sharing

USER LOGIN

Username: *

Password: *

[Create new account](#)
[Request new password](#)

Welcome to FrameNet

This is now the official website for the FrameNet Project. The old website, built using a CMS called "Mambo" is hereby retired. If you notice anything that was on the old website, but can't be found on the new site, please let us know, and we'll try to add it.
Thanks. -Collin Baker

13954 reads

RECENT NEWS

10/26/11 Outages due to Drupal server

We have been having some mysterious problems in which the website becomes unavailable, although the machine it's running on seems to be OK. The sysadmins are working on the problem, and hope to have a fix soon. In the meantime, please be patient, and try again after a few hours. Thanks.

09/25/11 Forums on the new FrameNet website

We are happy to have a "Forums" section on the new website, which should facilitate users exchanging ideas. All visitors to the site should now be able to view the posts in the forum. (Thanks to Adam Richman, of ICSI support staff!) However, we have been getting a bit of spam from anonymous postings, so you will now need to request an account in order to post anything yourself. Thanks.

05/06/11 Current data now on public website

The frames, FEs, LUs, and annotation browsable from this website are

22:05
09/11/2011

O Projeto FrameNet Brasil

- A FrameNet Brasil é um projeto de pesquisa desenvolvido na UFJF em associação com o ICSI (EUA), em cooperação com o Projeto FrameNet americano e está desenvolvendo um recurso lexical *on-line* para o Português, baseado na semântica de frames e sustentado por evidência de *corpus*.

Página inicial da FrameNet Brasil

<http://www.framenetbr.ufjf.br/>

The screenshot shows a web browser window with the URL www.framenetbr.ufjf.br/index.php?lang=en. The page features a dark red header with the text "FRAMENET BRASIL" in white. Below the header, the word "Home" is visible on the left and "Português" on the right. A "Main Menu" is highlighted with a red border, containing links to Home, Overview, Corpus, Data, FN-Br Related Projects, News, Publications, FAQ, Links, People, and Contact Us. The main content area includes three news items: "The FrameNet Brasil Project" (last updated 25 July 2011), "First Data Release" (last updated 19 July 2011), and "FrameNet BR Desktop Launched" (last updated 19 July 2011). A "Login" section is located on the left side of the main content area, with fields for Username and Password, a "Remember Me" checkbox, and a "Login" button. A link for "Forgot your password?" is also present. The Windows taskbar at the bottom shows the date as 09/11/2011 and the time as 22:11.

Home Português

Main Menu

- Home
- Overview
- Corpus
- Data
- FN-Br Related Projects
- News
- Publications
- FAQ
- Links
- People
- Contact Us

The FrameNet Brasil Project

The FrameNet Brasil Project (FN-Br) aims at constructing an on-line lexical database for Brazilian Portuguese, which is based on Frame Semantics and relies on corpus evidence. The project started from the contact between brazilian researchers from the [Federal University of Juiz de Fora \(UFJF\)](#) and the [Berkeley FrameNet Project](#) team, in 2007. By the end of that year, Professor Carlos Subirats, Principal of the [Spanish FrameNet Project](#), visited UFJF, where he conducted a workshop on his experience in implementing, for Spanish, a lexicographic resource analogous to that being developed to English.

Last Updated on Monday, 25 July 2011 20:32

[Read more... >>](#)

First Data Release

FN-Br has just launched its first data release. A total of 32 frames and 38 lexical units are now available for the general public through the link [Data](#) in the Main Menu.

Last Updated on Tuesday, 19 July 2011 21:09

FrameNet BR Desktop Launched

FrameNet Brasil has now its own tool for lexicographic annotation: FrameNet BR Desktop, a web-based software which handles the whole process of the project, from the creation/translation of a new frame in Portuguese to the release of the data, via the annotation of the sentences containing the target lexical units.

Last Updated on Tuesday, 19 July 2011 21:41

[Read more... >>](#)

Login

Username

Password

Remember Me

• [Forgot your password?](#)

Semântica de Frames

- Frame é um conjunto estruturado de experiências, crenças ou práticas que constituem um tipo de pré-requisito para a compreensão do sentido linguístico (Fillmore, 1975; 1977; 1982).
- Os significados são relativizados às cenas.

Definição de Desempregado no Dicionário Houaiss



Desempregado

Acepções

- adjetivo e substantivo masculino que ou aquele que se desempregou ou que não trabalha; descolocado, desocupado



A Escolha dos Frames e das Unidades Lexicais

- Escolha inspirada no estudo do Instituto Max Planck sobre os padrões de lexicalização dos esquemas conceptuais de “desconstituição de uma unidade física”, verificados numa amostra formada por dezessete línguas geneticamente dissimilares (conforme Majid *et al.* 2007).

SEPARAÇÃO FÍSICA

COM desconstituição da matéria
Processo irreversível
(eventos como *cortar* e *quebrar*)

SEM desconstituição da matéria
Processo reversível
(eventos como *abrir*)

Frames

Fechamento
João abriu a caixa.
A caixa abriu.

Movimento_corporal
João abriu os olhos.
Os olhos abriram.

Predição das partes resultantes
(eventos como *cortar*)

Não predição das partes resultantes
(eventos como *quebrar*)

Frame

Corte
João fatiou o pão.
**O pão fatiou.*

Frames

Causar_fragmentação
João quebrou o copo.

Cenário_de_fragmentação
O copo quebrou.

Unidades Lexicais selecionadas

Unidades Lexicais
selecionadas
(que evocam o frame
Fechamento)
Monolexêmicas

desabotoar.v

desarrolhar.v

desatarraxar.v

destampar.v

Polilexêmicas

abrir_((tampa)).v

levantar_((tampa)).v

tirar_((tampa)).v

Unidades Lexicais
selecionadas
(que evocam o frame
Movimento_corporal/ usam
o frame Fechamento)
Polilexêmicas

Abrir_((boca)).v

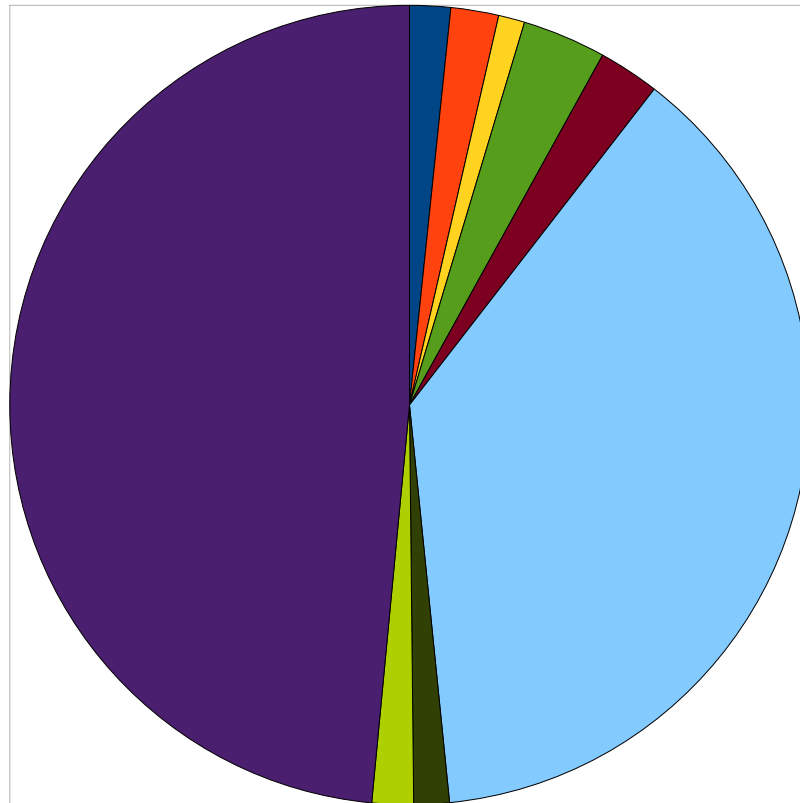
Abrir_((mão)).v

Abrir_((olho)).v

Procedimentos Metodológicos

- Busca dos lexemas em *corpora* eletrônicos
- **Corpus:** a base de dados atual da FrameNet Brasil é constituída pela combinação de cinco *corpora*, que representam usos do Português do Brasil. Contempla diversos gêneros e encontra-se em fase de expansão.
- Tratamento dos dados e Constituição de sub-*corpora*
- Anotação lexicográfica

Corpus FrameNet Brasil



- Oral
- Didático
- Jurídico
- Literário
- Téc. & Cien.
- Jornalístico
- Universitário
- Mensagem eletrônica
- Legendas de filmes

Relação dos *Corpora* do Projeto FrameNet Brasil

<http://www.linguateca.pt/acesso/index.php>

(i) **ANCIB:** corpus criado a partir de mensagens enviadas para a lista homônima da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (até Novembro de 2003) e para a lista abarreto-1, após essa data;

(ii) **ECI-EBR:** corpus criado pela ECI (European Corpus Initiative), baseado no Borba-Ramsey. É uma seleção de excertos de obras brasileiras, contendo pelo menos discurso literário, didático e oral cuidado (discursos políticos);

(iii) **NILC/São Carlos:** contém textos brasileiros do registro jornalístico (do qual se originou o CETENFolha), didático, epistolar e redações de alunos;¹⁷

Relação dos *Corpora* do Projeto FrameNet Brasil

<http://www.sketchengine.co.uk/>

(iv) **NURC-RJ**: corpus constituído por entrevistas gravadas nas décadas de 1970 e 1990, num total de 350 horas, com informantes de nível superior completo, nascidos no Rio de Janeiro e filhos de pais preferencialmente cariocas;

(v) **LF (Legendas de Filmes)**: corpus criado pelo Projeto FrameNet Brasil, sediado na Universidade Federal de Juiz de Fora, contém legendas de filmes em Português do Brasil cedidas pelo portal OpenSubtitles.org .

Fórmulas de Busca nos Corpora Eletrônicos

Unidade Lexical	Corpora lematizados	Corpora não lematizados
<i>Desabotoar</i>	[lema="desabotoar"]	"(?i)desaboto.*"
<i>Desarrolhar</i>	[lema="desarrolhar"]	"(?i)desarrolh.*"
<i>Desatarraxar</i>	[lema="desatarraxar"]	"(?i)desatarrax.*"
<i>Destampar</i>	[lema="destampar"]	"(?i)destamp.*"
<i>Abrir_((tampa))</i>	[lema="abrir"] []{0,20} "tamp.*" "tamp.*" []{0,20}[lema="abrir"]	"(?i)abr.*" []{0,20} "tamp.*" "tamp.*" []{0,20} "(?i)abr.*"
<i>Levantar_((tampa))</i>	[lema="levantar"][]{0,20} "tamp.*" "tamp.*"[]{0,20}[lema="levantar"]	"(?i)levant.*" []{0,20} "tamp.*" "tamp.*"[]{0,20} "(?i)levant.*"
<i>Tirar_((tampa))</i>	[lema="tirar"] []{0,20} "tamp.*" "tamp.*" []{0,20}[lema="tirar"]	"(?i)tir.*"[]{0,20} "tamp.*" "tamp.*" []{0,20} "(?i)tir.*"
<i>Abrir_((boca))</i>	[lema="abrir"] []{0,20} "boca.*" "boca.*" []{0,20}[lema="abrir"]	"(?i)abr.*" []{0,20} "boca.*" "boca.*" []{0,20} "(?i)abr.*"
<i>Abrir_((mão))</i>	[lema="abrir"] []{0,20} "mão.*" "mão.*" []{0,20}[lema="abrir"]	"(?i)abr.*" []{0,20} "mão.*" "mão.*" []{0,20} "(?i)abr.*"
<i>Abrir_((olho))</i>	[lema="abrir"] []{0,20} "olh.*" "olh.*" []{0,20}[lema="abrir"]	"(?i)abr.*"[]{0,20} "olh.*" "olh.*" []{0,20} "(?i)abr.*"

Tipos de Classificações para os diferentes usos dos lexemas

Grupo A: Utilizado para todos os subcorpora	
Número	Classificação
1	Sentido físico
2	Sentido figurativo
3	Adjetivo
4	Substantivo
5	Contexto ambíguo ou insuficiente
6	Outros

O item polilexêmico Abrir_((tampa))

Resultado da limpeza de *corpus*

Corpus	Sent. Físico	Figurativo	Adjetivo	Substantivo	Contexto insuf.	Outros	TOTAIS
ANCIB	-	-	-	-	-	-	-
ECI-EBR	3	-	-	-	1	5	9
NILC/São Carlos	11	-	3	-	-	10	24
NURC-RJ	1	-	-	-	-	-	1
LF	16	-	-	-	3	16	35
TOTAIS	31	-	3	-	4	31	69

Anotação das Unidades Lexicais

(Ruppenhoffer *et al*, 2006)

- **Valências das Unidades Lexicais:** Possibilidades combinatórias no âmbito semântico (Elemento de Frame) e no âmbito sintático (Função Gramatical e Tipo Sintagmático)
- Realização sintática de cada Elemento de Frame
- Padrões sintáticos desempenhados pelos Elementos de Frame centrais (que distinguem o frame)

Parâmetros de Anotação na FrameNet

Retirar a panela do fogo e sem DESTAMPAR deixar esfriar . [INI] [IND]

Camadas			sem	DESTAMPAR
EF	Agt= INI	Obj_cont= IND		Prend
TS				
FG				
Verbo				

Conceitos relacionados à Palavra-alvo

Lexema: comporta todas os sentidos de *Destampar*

Unidade Lexical: *Destampar*

Palavra-alvo: *Destampar*

Incorporação morfológica

Conceitos relacionados aos Elementos de Frame

Instanciações Nulas (omissões sintáticas dos EFs)

- **IND** - Instanciação Nula Definida
- **INI** - Instanciação Nula Indefinida
- **INC-** Instanciação Nula Construcional

O Frame Fechamento

Definição

Um **Agente** manipula um **Preendedor** para abrir ou fechar um **Objeto_tipo_contêiner** (por exemplo: casaco, jarra). Algumas vezes uma **Região_cercada** ou um **Portal_do_contêiner** podem ser expressos. Uma vez que o **Manipulador** é sintaticamente omissível, vários verbos neste frame incorporam o **Preendedor**.

EFs Nucleares

Ela **DESTAMPOU** a garrafa

D. Odete **DESABOTOAVA** a braguilha.

EFs Não-nucleares

Ale, **TIRE** a tampa do porta malas para mim, **rápido**

Legenda dos Elementos de Frame

<i>EF</i>	<i>Nuclearidade</i>
Agente	Central
Preendedor	Central
Objeto_tipo_contêiner	Central
Portal_do_contêiner	Central
Manipulador	Periférico
Tempo	Periférico
Modo	Periférico
Lugar	Periférico
Finalidade	Periférico
Grau	Periférico
Possuidor	Periférico
Beneficiário	Extra-temático
Iteração	Extra-temático
Razão	Extra-temático

Frame Fechamento

1- ULs Monolexêmicas (EF PRENDEDOR incorporado)

2- ULs Polilexêmicas (EF PRENDEDOR em interação com a Palavra-alvo verbal)

1- ULs Monolexêmicas

Unidade Lexical *Destampar*

- **2º Padrão:**
- **Elementos de Frame:** Agente – Objeto_tipo_contêiner – Prendedor (incorporado à palavra-alvo)
- **Realizações sintáticas:** Agente IND; Objeto_tipo_contêiner Obj/SN; Prendedor (incorporado morfológicamente)

Que você acha se **DESTAMPAMOS** **uma caixa de champanha?** -Se for só uma caixa. - Esse convite foi coisa sua. - Sim, para me divertir. -Por que não se diverte comigo? E **IND**

Camada		DESTAMPAMOS	uma caixa de champanha?
EF	Agt=IND	Prend	Obj_cont
FG			Obj
TS			SN
Verbo			

Padrões de Valência- UL *Destampar*

Número anotado	Padrões		
6 TOTAIS	Agente	Objeto_tipo_contêiner	Prendedor
(2)	SN Ext	SN Obj	Incorporado
(2)	IND	SN Obj	Incorporado
(1)	INI	SN Obj	Incorporado
(1)	INI	IND	Incorporado

2- ULs Polilexêmicas

Unidade Lexical *Abrir_((tampa))*

- 9º Padrão:
- Elementos de Frame: Agente – Objeto_tipo_contêiner – Prendedor
- Realizações sintáticas: Agente Ext/SN; Objeto_tipo_contêiner Dep/SP; Prendedor Obj/SN

Nunca se disse que o Julio acendia fósforos para esconder o cheiro -que deixava no banheiro da casa de Tenoch -ou que **Tenoch** **ABRIA** **a tampa** **do vaso** **com o pé** **na casa de Julio**. -Esses eram detalhes que um não queria saber sobre o outro.

Camada	Tenoch	ABRIA	a tampa	do vaso	com o pé	na casa de Julio.
EF	Agt		Prend	Obj_cont	Manip	Lugar
FG	Ext		Obj	Dep	Dep	Dep
TS	SN		SN	SP	SP	SP
Verbo						

Elementos de Frame e suas realizações sintáticas- UL *Abrir_((tampa))*

Elemento de Frame	Número Anotado	Realizações
Agente	(31)	SN.Ext=(5) IND=(16) INI=(8) INC=(2)
Objeto_tipo_contêiner	(31)	IND=(18) INI=(1) SP.Dep=(12)
Preendedor	(31)	SN.Obj=(28) IND=(3)
Beneficiário	(1)	SNpro.Dep=(1)
Manipulador	(2)	SP.Dep=(2)
Tempo	(2)	SAdv.Dep=(2)
Lugar	(1)	SP.Dep=(1)
Finalidade	(1)	SVinfP=(1)

O Frame Movimento_corporal

Definição

Este frame contém palavras para movimentos ou ações que um **Agente** desempenha usando alguma **Parte_do_Corpo**.

EFs Nucleares

Ezequiel **ABRIU** a boca.

ela não **ABRIU** a mão!

EFs Não-nucleares

tio Pedro **ABRIU** os olhos, **pálido, muito magro**

ele não **ABRE** a mão **nem para cumprimentar**

Unidade Lexical *Abrir_((boca))*

- 1º Padrão:
- Elementos de Frame: Agente – Parte_do_Corpo
- Realização Sintática: Agente INC; Parte_do_Corpo Obj/SN

Qual o problema, baby? -Você está alto. Está me envergonhando. Não quero bolo. - Vamos lá. **ABRA** a boca. Estão vendo? **INC**

Camada		ABRA	a boca
EF	Agt=INC		P_corpo
FG			
TS			
Verbo			

- 2º Padrão:
- Elementos de Frame: Agente – Parte_do_Corpo
- Realização Sintática: Agente Ext/SN; Parte_do_Corpo Obj/SN

depois fez o mesmo às pernas, e acabou por amarrar as duas mandíbulas, de modo que **a onça** não pudesse **ABRIR** a boca.

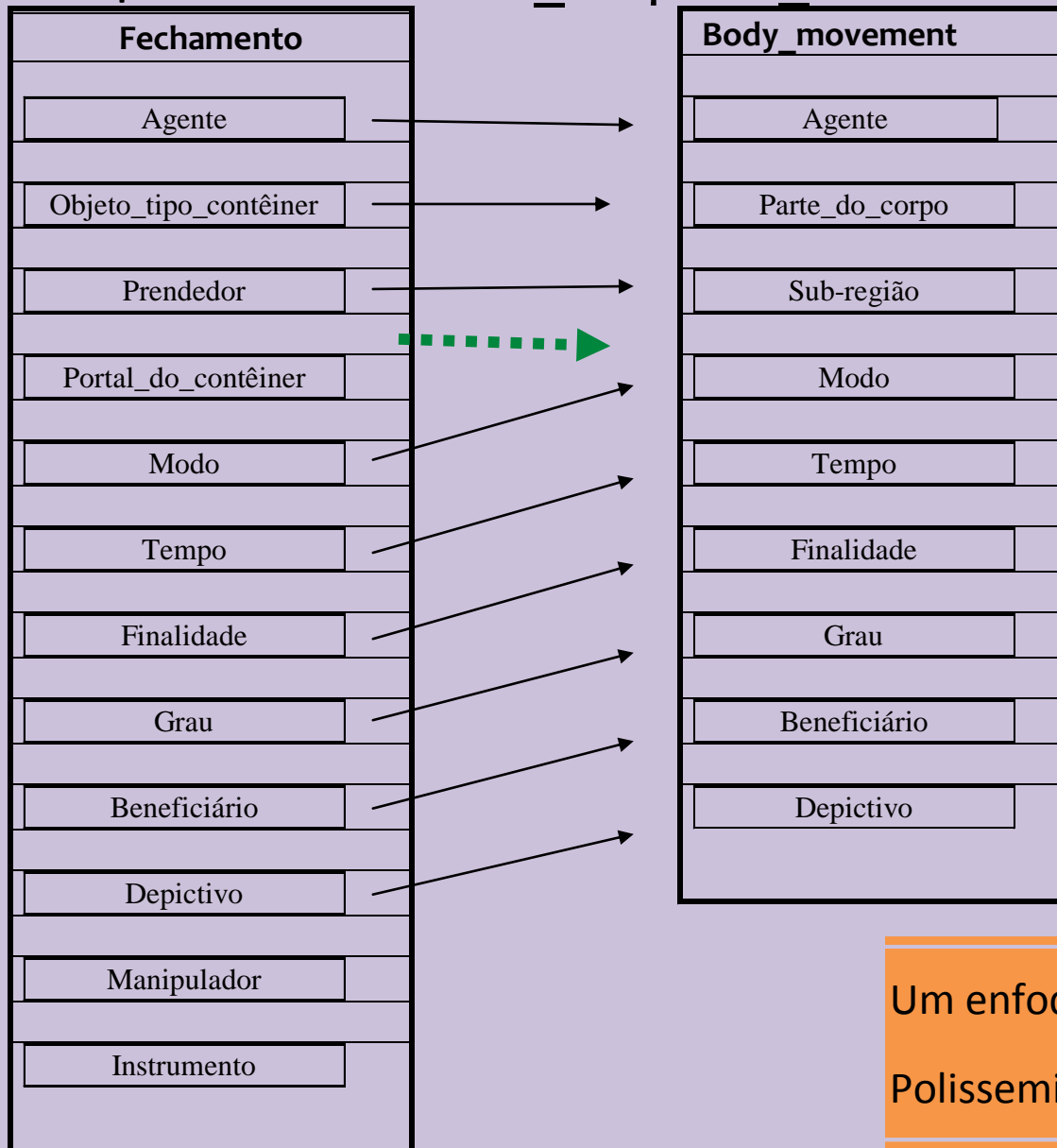
Camada	a onça	não pudesse	ABRIR	a boca
EF	Agt			P_corpo
FG	Ext			Obj
TS	SN			SN
Verbo		Modal		

Elementos de Frame e suas realizações sintáticas- UL *Abrir_((boca))*

Elemento de Frame	Número Anotado	Realizações
Agente	(104)	SN.Ext=(21) INC=(66) IND=(16) INI=(1)
Parte_do_corpo	(104)	SN.Obj=(88) IND =(16)
Tempo	(7)	Or.Sub. Adv. Temporal. Dep= (4) SAdv.Dep= (2) SP.Dep=(1)
Finalidade	(4)	SvinfP.Dep=(3) SP.Dep=(1)
Lugar	(1)	SAdv.Dep=(1)
Modo	(1)	SAdv.Dep= (1)
Grau	(09)	SAdv.Dep= (7) SA.Dep=(1) SP.Dep=(1)
Causa_interna	(2)	SP.Dep=(1) SVger.Dep=(1)
Depictivo	(2)	SA.Dep=(2)
Beneficiário	(1)	SP.Dep=(1)

Comentários

Relação de *Uso* entre Fechamento e Movimento_corporal, que se especializa para Movimento_corporal abertura



Um enfoque para a
Polissemia de *ABRIR*

Considerações metodológicas

Avanços no desenvolvimento do Projeto FrameNet Brasil

- *Corpora* com variantes do Português do Brasil
- Aumento de *corpora* no Projeto FrameNet BR que contemplam língua falada ou escrita oralizada (maior variedade de padrões)
- Consolidação dos padrões sintáticos considerando-se apenas os **Elementos Centrais** (que efetivamente definem o frame) de forma manual

Considerações

- No então estágio do Projeto, os procedimentos de anotação e de formulação das tabelas consolidadoras são praticados manualmente, o que se configura uma tarefa exaustiva.
- No entanto, tais procedimentos contribuem para o desenvolvimento da ferramenta computacional *FrameNet Brasil DeskTop*

Referências

- FILLMORE, C. J. Frame semantics. In: **Linguistics in the morning calm**. Seoul, Hanshing Publishig Co.: 1982.
- MAJID, A. et al. “The Semantic categories of cutting and breaking events: a crosslinguistic perspective”. In: *Cognitive Linguistics*. Ed. Ewa Dabrowska. 2007.
- SALOMÃO, M.M.M. “FrameNet Brasil: um trabalho em progresso” In: **Calidoscópico**, Vol. 7.2, 2009a.
- RUPPENHOFER, J. et al. *FrameNet II: Extended Theory and Practice*. 2006.