

Susana Afonso

UMA FRAMENET PARA O PORTUGUÊS

Sumário

2

- Projecto Framenet
- Definição de “frame” (quadro conceptual **QC**)
- Estrutura de uma framenet
- Anotação semântica com base em QCs: SALSA
- Framenet para o português e uso dos recursos existentes para a sua construção
- Vantagens no uso deste tipo de recursos

Framenet

3

- Projecto lexicográfico:
 - Baseado na Teoria dos Quadros Conceptuais ('Frame semantics')
 - Extracção de corpora das combinações sintáctico-semânticas ligadas a uma determinada palavras (verbos, principalmente):
 - Anotação por QCs e busca em corpora de exemplos relevantes que ilustrem os elementos dos QCs
 - Estabelece redes de quadros conceptuais

Framenet

4

- Línguas: inglês, japonês, alemão, espanhol e português
- Framenet para o português:
 - ▣ Framenet Brasil liderado por Maria Margarida Salomão
 - ▣ FrameCorp liderado por Rove Chishman

Quadro conceptual ('frame')

5

- Todos os conceitos num quadro conceptual evocam e são evocados pelo quadro na sua totalidade.
- Para entender um conceito, é necessário ter em conta o QC e todos os conceitos nele presentes.

Exemplo de um quadro conceptual

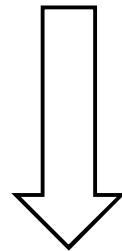
6

Cura em vinte e quatro horas?

CURAR

7

- Sentido 1: curar uma doença
- Sentido 2: processo de preservação



2 quadros conceptuais

Como é um quadro conceptual?

8

- Valência verbal
- Papéis semânticos específicos para cada um dos enquadramentos:
AGENTE: comprador, vendedor, falante,...
- Informação sintáctica (argumentos e adjuntos) e sintagmática (*sn*, *sv*, *sp*, etc.)
- Relações entre quadros conceptuais

CURAR

9

Definição:....

Verbos pertencentes a este QC:...

Elementos do quadro conceptual:

Nucleares/obrigatórios:

- ▣ Affliction [Affl] <def>, <ex>
- ▣ Body_part [B_p] <def>, <ex>
- ▣ Healer [Hlr]
- ▣ Medication [Med] ...
- ▣ Patient [Pat] ...
- ▣ Treatment [Trtm] ...

CURA

10

Elementos do quadro conceptual:

Periféricos/não obrigatórios

- ▣ Degree [Degr] <def>, <ex>
- ▣ Duration [Dur] ...
- ▣ Manner [Manr]...
- ▣ Motivation [Motive]...
- ▣ Place [Place]...

CURAR: sintaxe dos elementos

| Number Annotated | Patterns | | | |
|------------------|---------------------|----------------------------|----------------|--|
| 12 TOTAL | <i>Affliction</i> | <i>Healer</i> | <i>Patient</i> | |
| (1) | 2nd -- | NP External argument | NP Obj | |
| (1) | DNI -- | CNI -- | NP Ext | |
| (1) | DNI -- | NP Ext | NP Obj | |
| (1) | INI -- | NP Ext | NP Obj | |
| (1) | NP Ext | CNI -- | INI -- | |
| (1) | PP[of] Dependent | CNI -- | NP Ext | |
| (1) | PP[of] Dep | CNI -- | NP Obj | |
| (4) | PP[of] Dep | NP Ext | NP Obj | |
| (1) | PPing[of] Dep | NP Ext | NP Obj | |

Exemplo

| Number Annotated | Patterns | | | |
|------------------|-------------------|---------------|----------------|--|
| | <i>Affliction</i> | <i>Healer</i> | <i>Patient</i> | |
| (1) | NP Ext | CNI -- | INI -- | |

[<[Affliction](#)> **Many conditions associated with old age**] can be treated and *alleviated*^{Tgt}, if not cured .

[<[Healer](#)> CNI]

[<[Patient](#)> INI]

CURAR: sintaxe dos elementos (cont.)

| Number Annotated | Patterns | | | |
|-------------------------|-------------------|---------------|----------------|--------------------|
| 5 TOTAL | <i>Affliction</i> | <i>Healer</i> | <i>Patient</i> | <i>Treatment</i> |
| (1) | DNI -- | NP Ext | PP[re] Dep | PPing[by] Dep |
| (1) | NP Obj | NP Ext | 2nd -- | PP[with] Dep |
| (1) | NP Obj | NP Ext | 2nd -- | PPing[by] Dep |
| (1) | NP Obj | NP Ext | INI -- | PP[with] Dep |
| (1) | NP Obj | NP Ext | INI -- | PP[without] Dep |

Exemplo

| Number Annotated | Patterns | | | |
|------------------|-------------------|---------------|----------------|--------------------|
| | <i>Affliction</i> | <i>Healer</i> | <i>Patient</i> | <i>Treatment</i> |
| (1) | NP Obj | NP Ext | INI -- | PP[without] Dep |

There 's a man down by the docks who claims [<Healer> **he**]
can *cure*^{Tgt} [<Affliction> **illness**] [<Treatment> **without medicine**].

[<Patient> INI]

CURAR: sintaxe dos elementos (cont.)

| Number Annotated | Patterns | | | |
|------------------|-------------------|----------------|------------------|--|
| 14 TOTAL | <i>Affliction</i> | <i>Patient</i> | <i>Treatment</i> | |
| (1) | 2nd -- | NP Ext | PP[by] Dep | |
| (1) | DNI -- | NP Ext | PPing[by] Dep | |
| (1) | NP Ext | DNI -- | PP[by] Dep | |
| (2) | NP Ext | INI -- | PP[by] Dep | |
| (1) | NP Ext | INI -- | PPing[by] Dep | |
| (1) | NP Obj | 2nd -- | NP Ext | |
| (5) | NP Obj | INI -- | NP Ext | |
| (2) | NP Obj | PP[in] Dep | NP Ext | |

Exemplo

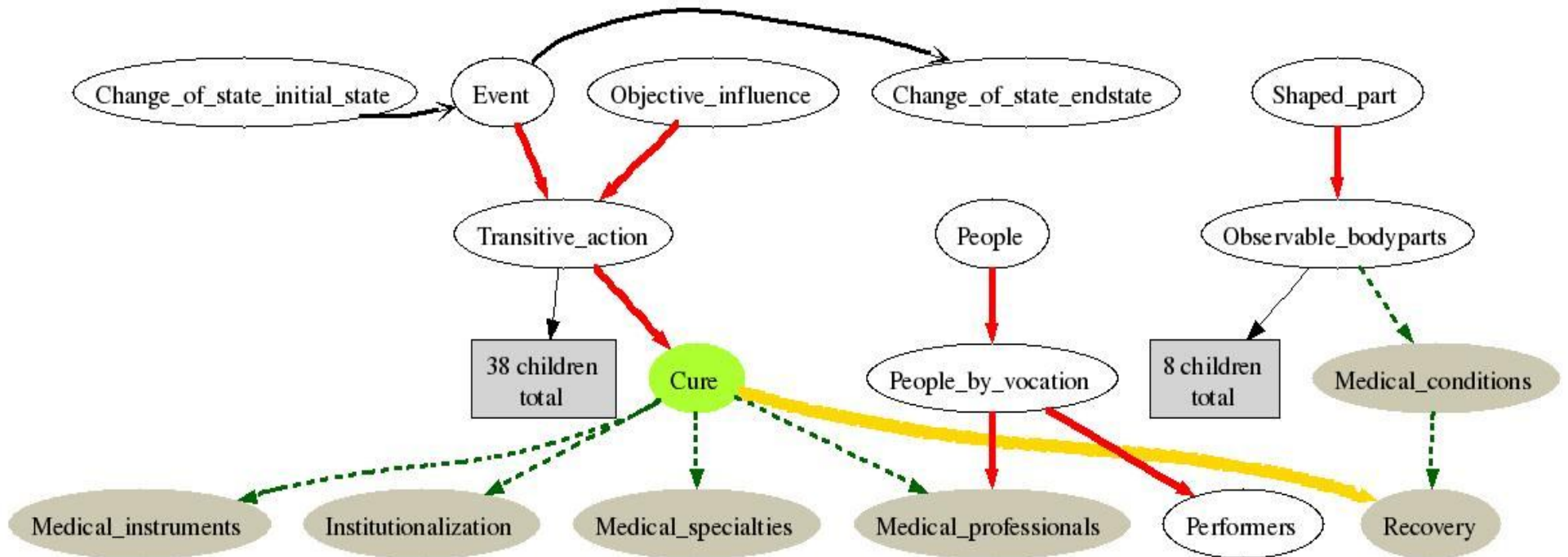
| Number Annotated | Patterns | | | |
|------------------|-------------------|----------------|------------------|--|
| | <i>Affliction</i> | <i>Patient</i> | <i>Treatment</i> | |
| (5) | NP Obj | INI -- | NP Ext | |

The difficulty here is that [Treatment simple lack of stress] may also *alleviate*^{Tgt} [Affliction the symptoms] – so an improvement while on holiday may be a slightly ambiguous result .

[Patient INI]

CURAR: relações entre QCs

17

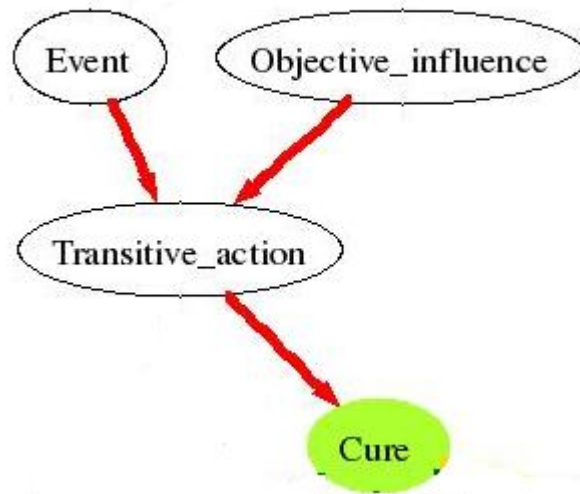


Relações entre QCs

18

□ Herança:

- ▣ relação do tipo "is-a";
- ▣ a totalidade de informação semântica de um QC propaga-se para um QC dependente



Relações entre QCs

19

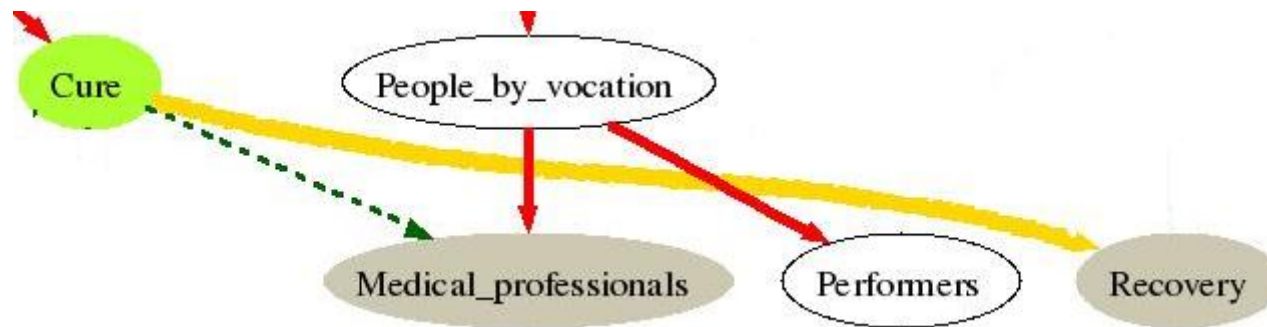
- Precedência:
 - Vários sub-QCs são fases de um ciclo. Cada fase precede a seguinte.



Relações entre QCs

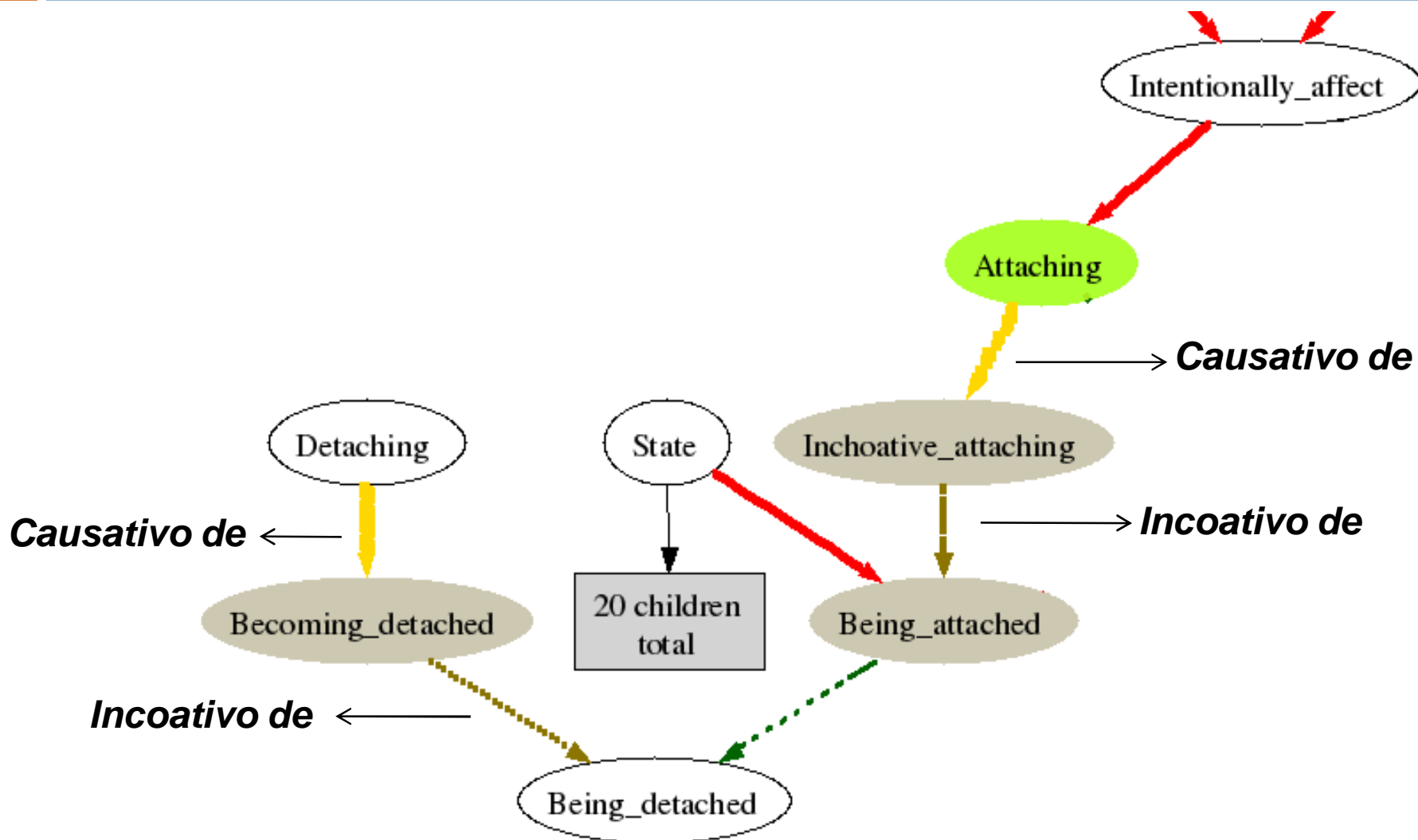
20

- Causativo de; incoativo de:
 - ▣ Relações que envolvem conceptualizações diferente sobre um mesmo evento que constitui um QC genérico.



Relações entre QCs: causativo/incoativo de

21

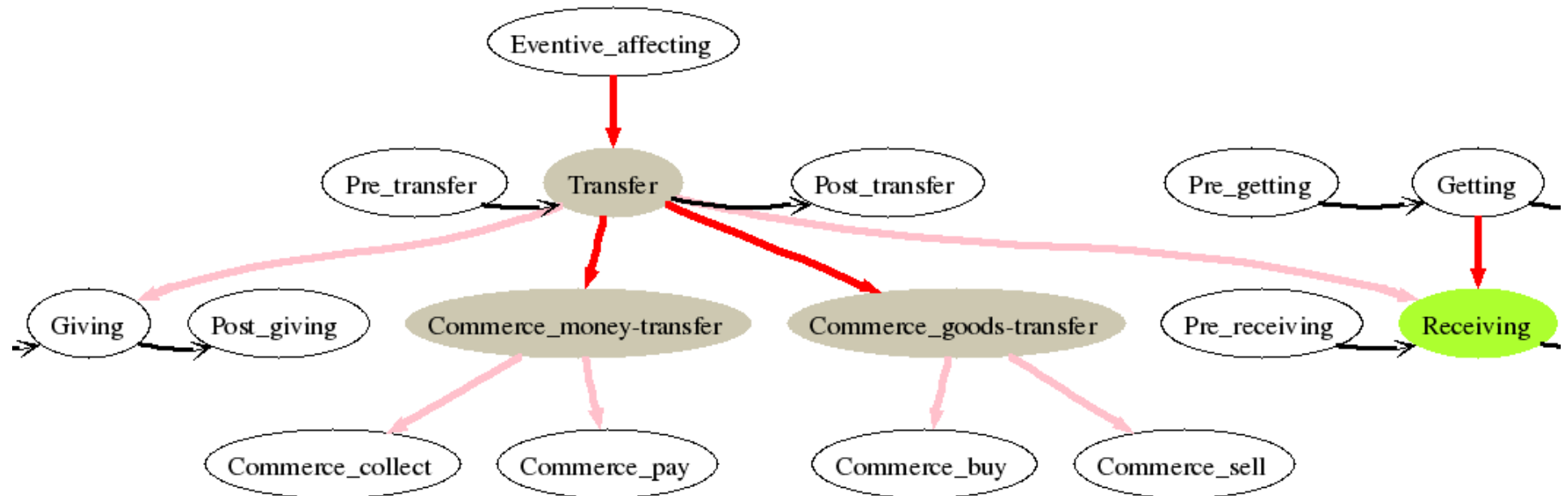


Relações entre QCs

22

□ Perspectiva:

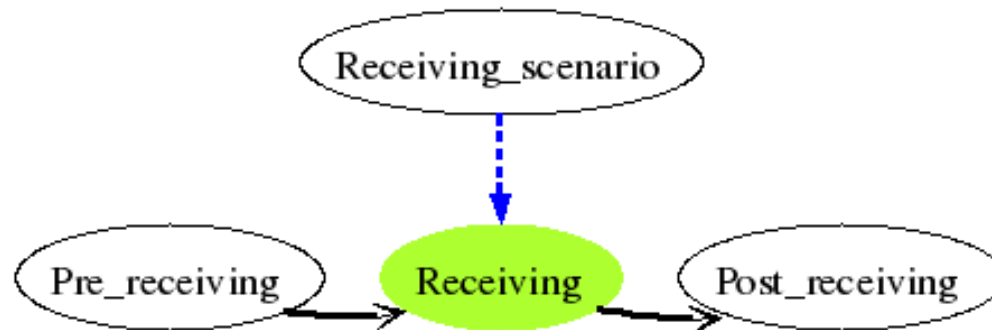
- Dois ou mais QCs constituem diferentes perspectivas sobre um QC mais abstracto (QC transacção comercial: QC que inclui a perspectiva do *comprador* e um QC separado que inclui a perspectiva do *vendedor*)



Relações entre QCs

23

- "Sub-frame":
 - ▣ Relação entre um QC complexo envolvendo transições de estados que constituem também QCs;



Framenet e corpora

24

- Propósito: BASE DE DADOS LEXICAL
 - ▣ As framenets usam corpora para extrair exemplos, ou seja,
 - Os exemplos são escolhidos para ilustrar todas as possibilidades de contextos sintático-semânticos das unidades lexicais de um QC;
 - Os exemplos são estruturas regulares quanto à relação sintaxe-semântica, deixando de fora estruturas mais idiomáticas e menos composicionais;

Framenet e corpora

25

- Propósito: ANOTAÇÃO SEMÂNTICA DE TEXTOS
 - ▣ De recente implementação pela Framenet Berkeley (inglês) para compreensão textual
 - ▣ Anotação com QCs, frase a frase (corpora como ponto de partida)

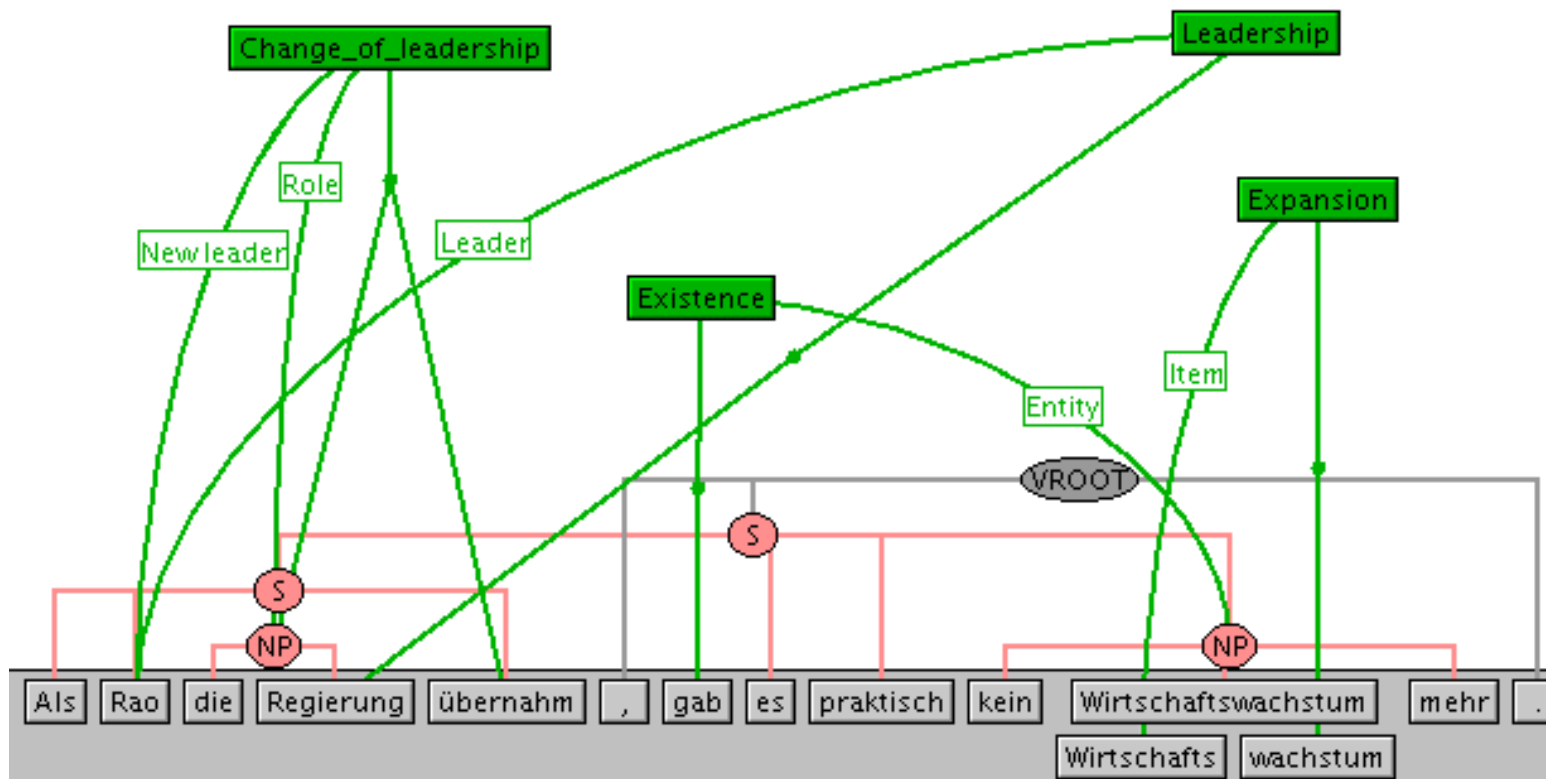
- Projecto **SALSA** (*Saarbrücken Lexical Semantics Annotation and Analysis*) (Burchard et al. to appear)
 - anotação do corpus TIGER (alemão)

Projecto **SALSA**

26

- Anotação de todos os lemas de um corpus com QCs já desenvolvidos para o inglês (os QCs são supostamente universais)
- Anotação com QCs feita sobre a anotação morfossintáctica existente
- Representação é também em formato de árvore

Exemplo



Desde que Rao subiu ao poder, não há praticamente nenhum crescimento económico

SALSA



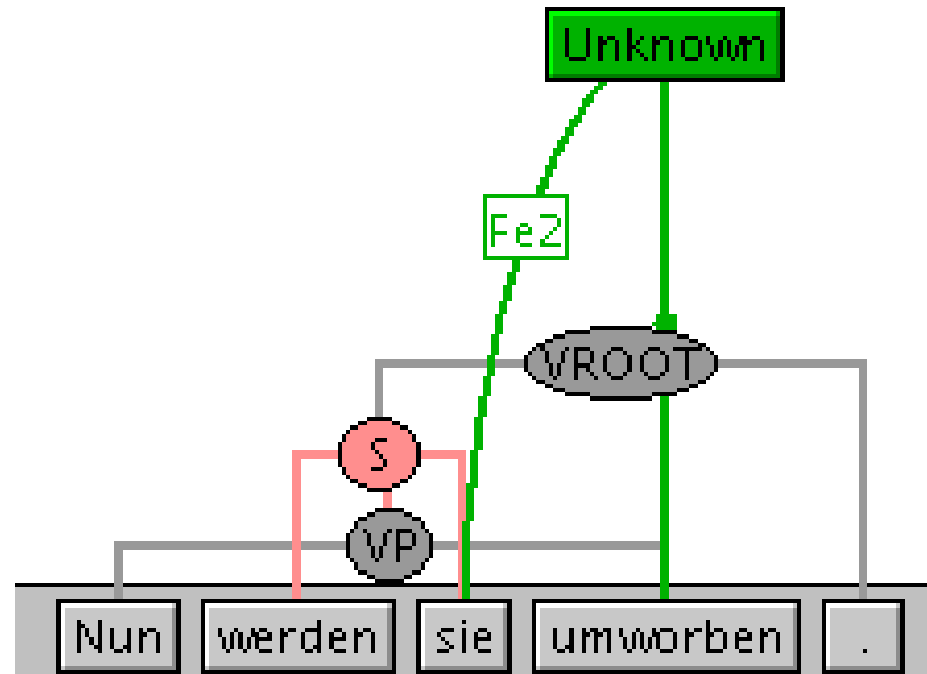
FrameNet

28

- Anotação de corpora com QCs do inglês
- Teste da hipótese da universalidade dos QCs
- Novos Qcs - 'Protoframes'
- Anotação de fenómenos idiossincráticos, de vagueza e ambiguidade
- Disponibilização de frequências

- Desenvolvimento de QCs e ilustração de todas as possibilidades com exemplos retirados de corpora
- Desenvolvimento das protoframes
- Escolha de fenómenos sintáctico-semânticos regulares (ilustração de QCs)
- Sem acesso a frequências

Exemplo



Agora ser.PROGR eles aliciados

'Agora eles estão a ser aliciados'

FrameNet para o português: *FrameNet Brasil e FrameCorp*

30

- FrameNet Brasil
 - Universidade Federal Juiz de Fora, e liderada por Maria Margarida Salomão
 - Implementação do projecto mãe (Berkeley, inglês) para o português
 - Inclui as variedades PB e PE
 - Construção será feita com base nos resultados descritivos do FrameCorp

FrameCorp

31

- FrameCorp (Chishman 2008)
 - UNISINOS, liderado por Rove Chishman
 - Objectivo: investigação lexicográfica numa perspectiva semântica e computacional tendo como modelo a Teoria dos QCs
 - Corpus investigado e anotado: SUMM IT. Português do Brasil, 50 textos (17.123 tokens) do Folha de S. Paulo, secção de ciência, anotado pelo PALAVRAS (Bick 2000), informação anafórica e retórica introduzida manualmente (Coelho et al. 2006)
 - Projecto semelhante ao projecto SALSA

FrameCorp

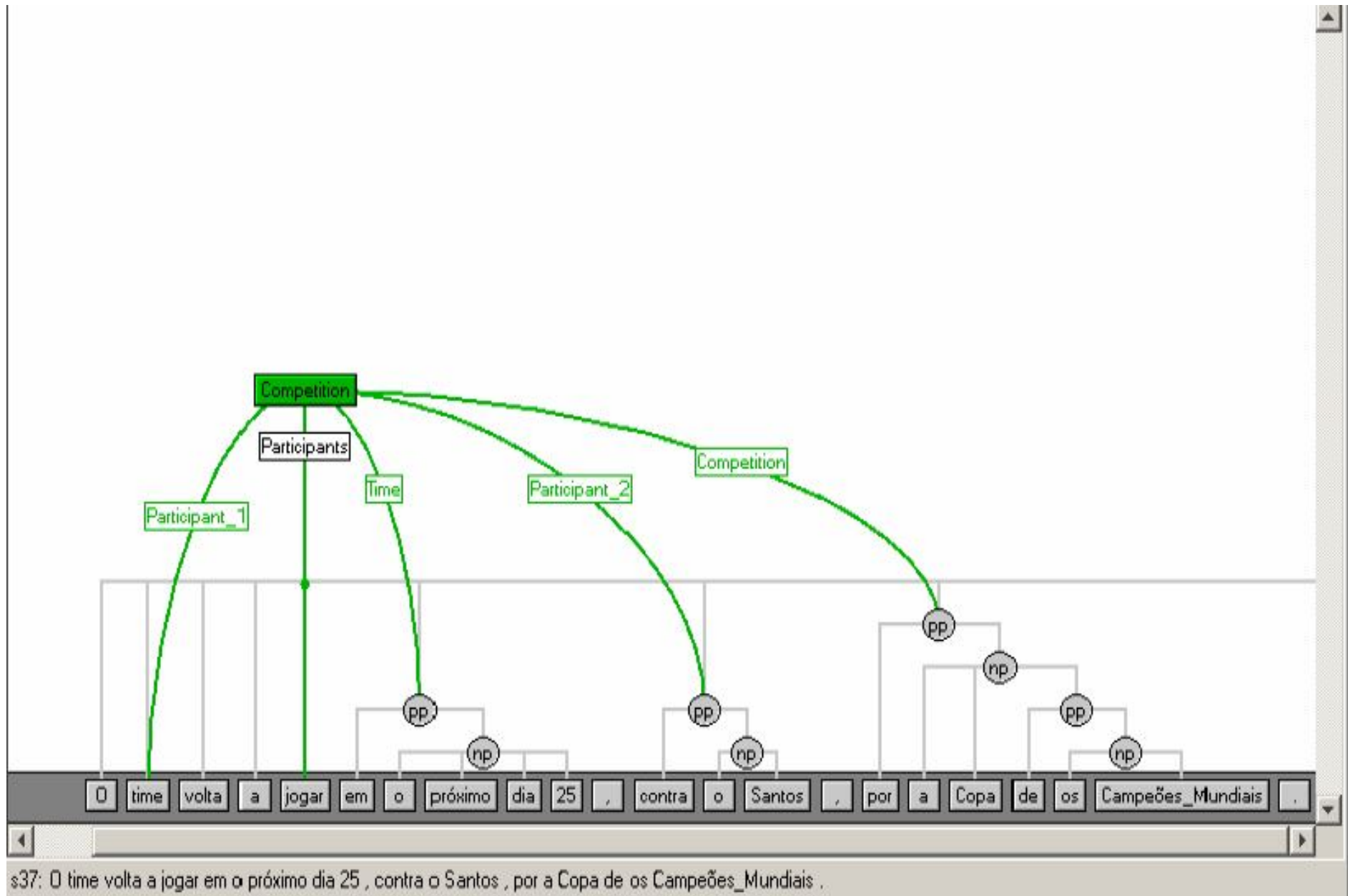
32

- Usa para anotar o corpus a ferramenta SALTO, desenvolvida pelo projecto SALSA
- Os QCs usados na anotação são os do inglês
- Alguns estudos preliminares (Chishman 2008a, b)
 - ▣ 70 frases com o verbo *jogar*
 - ▣ Anotação com o QC “statement”
- Nos dois estudos, a anotação foi bem sucedida, o que parece indicar que SALTO é apropriado para anotar textos em português.

Jogar (Chishman 2008a)

33

| <i>Frame</i> | Ocorrências |
|-----------------------------|-------------|
| Competition | 60 |
| Copmpetition/being employed | 04 |
| Cause motion | 03 |
| Self motion | 03 |
| Total | 70 |



Floresta e Framenet

35

- A Floresta pode contribuir para a construção da FrameNet para o português e ao mesmo tempo beneficiar desse recurso:
 - corpora de vários gêneros, anotados morfossintacticamente;
 - informação semântica: já se testou anotação semântica (Bick 2007) com valores de recall de 86,6% e de precision de 90,5%
 - Não foi ainda anotado nenhum corpus de PE com o SALTO.

Floresta e Framenet

36

- A Framenet permite aumentar a robustez de um treebank, através da identificação de inconsistências da anotação manual.
- Mais geralmente, uma Framenet pode informar opções de tratamento lexicográfico

Vantagens do uso de uma framenet e/ou de um corpus anotado com QCs

37

- Tradução (existência de QCs na língua-fonte e língua-alvo)
- Desenvolvimento de sistemas de pergunta – resposta
- Investigação linguística:
 - ▣ Lexicografia (cp. WordNet)
 - ▣ **Construções:**
 - qualquer estrutura, atômica ou complexa, que constitua um par de forma e significado
 - Itens lexicais, especialmente verbos cuja relação semântica com a construção no seu todo é mais próxima

Impessoalização

38

- Um corpus anotado com QCs implica que os participantes sejam também anotados.
- Há três categorias de elementos omissos que são usadas na anotação dos QCs:
 - **CNI** 'constructional null instantiation': conceptualmente presente mas genérico e omissivo

Ex. CAUSE MOTION

[<Theme> **Muchos pescados muertos**] fueron *arrojados*^{Tgt} [<Goal> **a las playas**] y los pescadores no salieron al mar.

[<Agent> **CNI**]

Impessoalização

39

- **INI** 'indefinite null instantiation': objectos/pacientes, fazendo com que um verbo transitivo ocorra intransitivamente (*comer, beber, ...*)

Ex: INGESTION

[<Ingestor> **He**] *drank*^{Tgt} [<Source> **from the little blue-and-white cup**] . [<Ingestibles> **INI**]

- **DNI** 'definite null instantiation': específico e recuperável através do contexto

Ex: COMMUNICATION

[<Communicator> **Otras familias reales**] *comunicaron*^{Tgt} [<Message> **su renuncia a delegar en uno de sus miembros la representación en las exequias de la princesa**] .

[<Addressee> **DNI**]

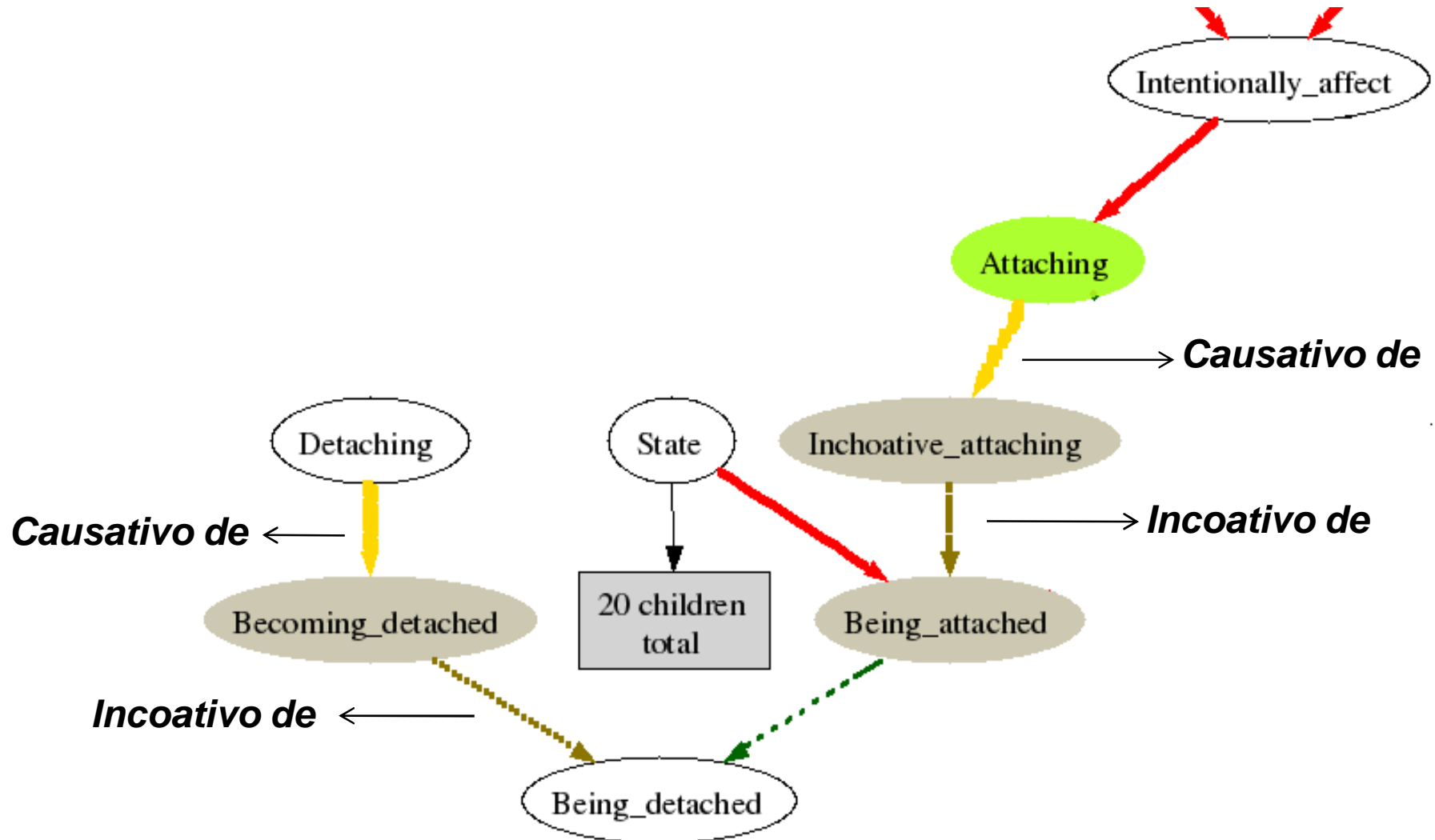
Impessoalização

40

- Em que QCs CNI e INI ocorrem com mais frequência?
- Que tipo de construções estão associadas a estes QCs?
- Qual a relação entre as construções e os QCs?
- Que tipo de itens lexicais (verbos) encontramos nestas construções?
- Qual a relação entre os QCs?

Attaching vs. inchoative attaching

41



Attaching vs. inchoative attaching

| Frame Element | Number Annotated | Realizations(s) |
|---------------|------------------|---|
| Agent | (15) | CNI.-- (5) NP.Ext (10) |
| Connector | (15) | INI.-- (12) PP[with].Dep (2) PP[by].Dep (1) |
| Goal | (15) | PP[into].Dep (1) PP[to].Dep (11) DNI.-- (3) |
| Item | (15) | NP.Obj (14) DNI.-- (1) |

Valence Patterns:

These frame elements occur in the following syntactic patterns:

| Number Annotated | Patterns | | | |
|------------------|-----------|-----------------|-----------------|-----------|
| 15 TOTAL | Agent | Connector | Goal | Item |
| (1) | CNI -- | INI -- | PP[into] Dep | NP Obj |
| (3) | CNI -- | INI -- | PP[to] Dep | NP Obj |
| (1) | CNI -- | PP[with] Dep | DNI -- | NP Obj |
| (1) | NP Ext | INI -- | DNI -- | NP Obj |
| (1) | NP Ext | INI -- | PP[to] Dep | DNI -- |
| (6) | NP Ext | INI -- | PP[to] Dep | NP Obj |
| (1) | NP Ext | PP[by] Dep | PP[to] Dep | NP Obj |

| Frame Element | Number Annotated | Realizations(s) |
|---------------|------------------|--|
| Goal | (8) | DNI.-- (1) PP[to].Dep (6) VPto.Dep (1) |
| Handle | (1) | PP[at].Dep (1) |
| Item | (9) | NP.Ext (9) PP[to].Dep (1) |
| Manner | (1) | AVP.Dep (1) |

Valence Patterns:

These frame elements occur in the following syntactic patterns:

| Number Annotated | Patterns | | | |
|------------------|---------------|-----------|------------|---------------|
| 4 TOTAL | Goal | Item | | |
| (4) | PP[to] Dep | NP Ext | | |
| 3 TOTAL | Goal | Item | Item | |
| (2) | PP[to] Dep | NP Ext | NP Ext | |
| (1) | VPto Dep | NP Ext | NP Ext | |
| 1 TOTAL | Goal | Item | Manner | |
| (1) | DNI -- | NP Ext | AVP Dep | |
| 1 TOTAL | Handle | Item | Item | Item |
| (1) | PP[at] Dep | NP Ext | NP Ext | PP[to] Dep |

Attaching vs. inchoative attaching: verbos

Attaching

Lexical Unit Information

| Lexical Unit | LU Status | Lexical Entry Report | Annotation Report |
|--------------------------|---------------------------|----------------------|-------------------|
| adhere.v | Needs_SCs | LE | Anno |
| affix.v | Finished_Initial | LE | Anno |
| agglutinate.v | Insufficient_Attestations | LE | |
| anchor.v | Finished_Initial | LE | Anno |
| append.v | In_Use | LE | |
| attach.v | Finished_Initial | LE | Anno |
| attachment_act.n | Finished_Initial | LE | Anno |
| attachment_item.n | Finished_Initial | LE | Anno |
| bind.v | Finished_Initial | LE | Anno |
| bond.v | Finished_Initial | LE | Anno |
| cement.v | Finished_Initial | LE | Anno |
| chain.v | Finished_Initial | LE | Anno |
| cinch.v | Insufficient_Attestations | LE | |
| concatenate.v | Insufficient_Attestations | LE | |
| connect.v | In_Use | LE | |
| detach.v | In_Use | LE | |
| fasten.v | Finished_Initial | LE | Anno |
| fetter.v | Insufficient_Attestations | LE | |

Inchoative attaching

Lexical Unit Information

| Lexical Unit | LU Status | Lexical Entry Report | Annotation Report |
|----------------------|------------------|----------------------|-------------------|
| agglutinate.v | Needs_SCs | LE | Anno |
| attach.v | Finished_Initial | LE | Anno |
| bind.v | Needs_SCs | LE | Anno |
| fasten.v | Needs_SCs | LE | Anno |
| moor.v | Needs_SCs | LE | Anno |
| stick.v | Finished_Initial | LE | Anno |
| take hold.v | Created | LE | |