



## SEMÂNTICA – PARTE 3

*SCC5869 Tópicos em Processamento de Língua Natural*

Thiago A. S. Pardo

### ANÁLISE SEMÂNTICA

- Até agora, significado da sentença

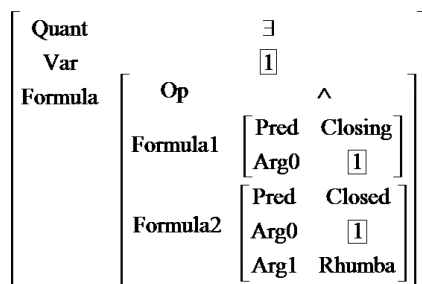
- Rhumba closed.*

Quant		$\exists$	
Var		[	1
Formula	Op		$\wedge$
	Formula1	[	Pred Closing
		Arg0	[
			1
	Formula2	[	Pred Closed
		Arg0	[
			1
		Arg1	Rhumba

## ANÁLISE SEMÂNTICA

### ○ Até agora, significado da sentença

- *Rhumba closed.*



### ○ Mas... o que é Rhumba? Qual sua caracterização semântica?

- É preciso mais do que simplesmente a palavra: **semântica lexical**

3

## SEMÂNTICA LEXICAL

### ○ Exemplos relativamente simples

- *Eu banco sua viagem.*
- *Em qual banco você trabalha?*
- *O banco tinha três pernas.*
- Quais os sentidos? Verbo “financiar”, instituição financeira ou assento?

4

## SEMÂNTICA LEXICAL

### ○ Exemplos relativamente simples

- *Eu banco sua viagem.*
- *Em qual banco você trabalha?*
- *O banco tinha três pernas.*
  - Quais os sentidos? Verbo “financiar”, instituição financeira ou assento?

### ○ Exemplos bem mais complexos

- *O banco está quebrado.*
- *O homem encontrou o banco.*
  - E agora? Mais informação (contextual) é necessária

5

## SEMÂNTICA LEXICAL

### ○ Tarefas envolvidas

- Desambiguação lexical de sentido
  - Significado/sentido das palavras
- Reconhecimento de entidades nomeadas
  - Tipo semântico de nomes próprios, normalmente
    - *Entidades nomeadas vs. mencionadas*
    - Isso é semântica lexical?
- Anotação de papéis semânticos
  - Atribuição de papéis aos elementos sentenciais
- **Outras?**

6

## SEMÂNTICA LEXICAL

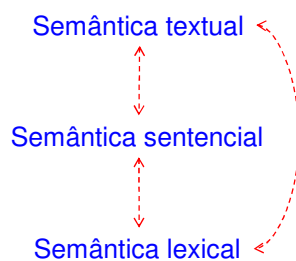
### ○ Tarefas envolvidas

- Desambiguação lexical de sentido
  - Significado/sentido das palavras
- Reconhecimento de entidades nomeadas
  - Tipo semântico de nomes próprios, normalmente
    - *Entidades nomeadas vs. mencionadas*
    - Isso é semântica lexical?
- Anotação de papéis semânticos
  - Atribuição de papéis aos elementos sentenciais
- Relacionamentos entre palavras, construção de ontologias, indexação ontológica, subcategorização, restrições de seleção, metáforas, etc.

7

## SEMÂNTICA

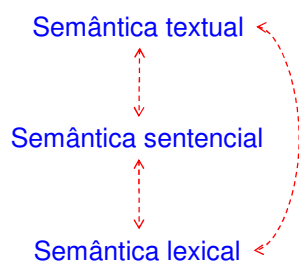
- Níveis de análise semântica mais usuais em PLN e interação entre eles
- Há outros níveis?
- Semântica lexical é sempre necessária para a análise sentencial?
- Semântica textual???



8

## SEMÂNTICA

- Níveis de análise semântica mais usuais em PLN e interação entre eles
  - Há outros níveis? **SIM**, por exemplo, morfologia
  - Semântica lexical é sempre necessária para a análise sentencial? **NÃO**, depende da aplicação e do que se quer
  - Semântica textual???  
**RST**, por exemplo



9

## SEMÂNTICA LEXICAL

- **Principais recursos** utilizados para as tarefas relacionadas à semântica lexical
  - Relações entre sentidos
    - WordNet de Princeton (Fellbaum, 1998) e outras wordnets, por exemplo
  - Eventos e seus participantes
    - FrameNet (Baker et al., 1998), por exemplo
  - Decomposição em primitivas de significados
    - Schank (1975), por exemplo

10

## RELAÇÕES ENTRE SENTIDOS

- **Sentido**: representação discreta de um aspecto do significado da palavra
- Diferentes tipos de relações entre sentidos
  - **Sinonímia**
    - *Ameaça e perigo*
  - **Antonímia**
    - *Feio e bonito*
  - **Meronímia** (e **holonímia**): parte-todo (todo-parte)
    - *Roda* é merônimo de *carro*, mas *carro* é holônimo de *roda*
  - **Hiperônimo** (e **hipônimo**): mais genérico (mais específico)
    - *Veículo* é hiperônimo de *carro*, mas *carro* é hipônimo de *veículo*

11

## RELAÇÕES ENTRE SENTIDOS

- **Sentido**: representação discreta de um aspecto do significado da palavra
- Diferentes tipos de relações entre sentidos
  - **Homonímia**: mesma forma, sentidos não relacionados
    - *Manga* (da camisa) e *manga* (fruta)
  - **Polissemia**: mesma forma, sentidos relacionados
    - *Banco* (de areia) e *banco* (instituição financeira)
    - Onde se depositam “coisas”
  - **Metonímia**: uso de um termo no lugar de outro
    - *Leitura de Machado de Assis*

12

## SINONÍMIA

- **Atenção:** relação entre palavras, não sentidos
  - Indicação: as palavras podem ser usadas de forma intercambiável
  - **Não há sinônimos perfeitos**
    - Por exemplo, **água** e **H<sub>2</sub>O**
      - **H<sub>2</sub>O** é um termo “mais científico”, não apropriado para um guia turístico, por exemplo
        - Essa diferença de gênero faz parte do significado dos termos

13

## ANTONÍMIA

- **Atenção:** vários tipos de oposição conceitual, podendo tornar a identificação difícil
  - Exemplos
    - *Vivo vs. morto, homem vs. mulher* (oposição complementar)
    - *Frio vs. quente* (gradação)
    - *Grande vs. pequeno* (escala)
    - *Esquerda vs. direita* (direção)
    - *Filho vs. pai* (reversível)

14

## ANTONÍMIA

- **Atenção:** vários tipos de oposição conceitual, podendo tornar a identificação difícil
  - Alguns antônimos são **marcados morfologicamente**
    - *Gostoso* vs. *desgostoso*

15

## HIPERONÍMIA/HIPONÍMIA

- **Atenção:** relações usadas em vários contextos
  - Taxonomias: somente hiperonímias/hiponímias
    - *Cão de caça* e *filhote* são hipônimos de *cão*
    - Nem sempre é simples
      - *Poodle* e *pastor alemão* também são tipos de cães, mas expressam relações diferentes das anteriores
  - Ontologias: outras relações também

16



## HIPERONÍMIA/HIPONÍMIA

- **Atenção:** relações usadas em vários contextos
  - Taxonomias: somente hiperonímias/hiponímias
    - *Cão de caça e filhote* são hipônimos de *cão*
    - Nem sempre é simples
      - *Poodle e pastor alemão* também são tipos de cães, mas expressam relações diferentes das anteriores
  - Ontologias: outras relações também

17

**Atenção: não há consenso**

## CAMPO SEMÂNTICO

- Identifica um conjunto de palavras de um mesmo domínio, dentre as quais há várias relações
  - Por exemplo: *reserva, vôo, viagem, preço, tarifa, refeição, avião*
    - Campo semântico de **viagem de avião**

18

## WORDNET DE PRINCETON

- **Maior e mais famoso** repositório de relações
- Conceitos representados por listas de palavras sinônimas
  - Synsets (*synonym sets*)
    - Exemplo: {casa, lar, moradia}
    - Palavras organizadas por frequência, muitas vezes
- Relações entre synsets, formando uma grande rede
- Entradas acompanhadas por glosas e exemplos de uso
  - Glosas: definições no estilo de dicionários
- Sem consenso, mas muitos a classificam como uma “ontologia lingüística”

19

## WORDNET DE PRINCETON

### ○ WordNet 3.0

WordNet Search - 3.0

wordnetweb.princeton.edu/perl/webwn?s=bass&sub=Search+WordNet&o2=&o0=1&o7=&o5=&o1=1&o6=&o4=

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options: (Select option to change)

Key: "S." = Show Synset (semantic) relations, "W." = Show Word (lexical) relations

**Noun**

- [S. \(n\) bass](#) (the lowest part of the musical range)
- [S. \(n\) bass, bass part](#) (the lowest part in polyphonic music)
- [S. \(n\) bass, basso](#) (an adult male singer with the lowest voice)
- [S. \(n\) sea bass, bass](#) (the lean flesh of a saltwater fish of the family Serranidae)
- [S. \(n\) freshwater bass, bass](#) (any of various North American freshwater fish with lean flesh (especially of the genus *Micropterus*))
- [S. \(n\) bass, bass voice, basso](#) (the lowest adult male singing voice)
- [S. \(n\) bass](#) (the member with the lowest range of a family of musical instruments)
- [S. \(n\) bass](#) (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

**Adjective**

- [S. \(adj\) bass, deep](#) (having or denoting a low vocal or instrumental range) "*a deep voice*"; "*a bass voice is lower than a baritone voice*"; "*a bass clarinet*"

[WordNet home page](#)

## WORDNET DE PRINCETON

- Exemplo: *entity* (**unique beginner**) – para substantivos

WordNet Search - 3.0  
wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&r=1&s=bass&l=3&h=0010000000#c

- [S: \(n\) bass, basso](#) (an adult male singer with the lowest voice)
  - [direct hypernym / inherited hypernym / sister term](#)
    - [S: \(n\) singer, vocalist, vocalizer, vocaliser](#) (a person who sings)
      - [S: \(n\) musician, instrumentalist, player](#) (someone who plays a musical instrument (as a profession))
        - [S: \(n\) performer, performing artist](#) (an entertainer who performs a dramatic or musical work for an audience)
          - [S: \(n\) entertainer](#) (a person who tries to please or amuse)
            - [S: \(n\) person, individual, someone, somebody, mortal, soul](#) (a human being) *"there was too much for one person to do"*
              - [S: \(n\) organism, being](#) (a living thing that has (or can develop) the ability to act or function independently)
                - [S: \(n\) living thing, animate thing](#) (a living (or once living) entity)
                  - [S: \(n\) whole, unit](#) (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"*; *"the team is a unit"*
                    - [S: \(n\) object, physical object](#) (a tangible and visible entity, an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                      - [S: \(n\) physical entity](#) (an entity that has physical existence)
                        - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
                          - [S: \(n\) causal agent, cause, causal agency](#) (any entity that produces an effect or is responsible for events or results)
                            - [S: \(n\) physical entity](#) (an entity that has physical existence)
                              - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

## WORDNETS

- Redes semânticas
- Iniciativas para várias línguas
- Possibilidade de indexação entre as wordnets
  - Por exemplo, EuroWordNet
- E para o português?

## EXERCÍCIO

- Em grupos de 3, respondam/ façam:
  - Como construir manualmente uma wordnet para uma língua sem recursos semânticos significativos?

23

## EXERCÍCIO

- Em grupos de 3, respondam/ façam:
  - Como construir automaticamente (mesmo que parcialmente) uma wordnet para uma língua sem recursos semânticos significativos?

24

## RELAÇÕES ENTRE SENTIDOS

- Algumas relações entre sentidos são sistemáticas
  - Exemplo
    - *Eu moro perto da USP <local>.*
    - *A USP <organização> elegeu um novo reitor.*
      - <local> ⇔ <organização>
    - Padrão recorrente: *escolas, hospitais, organizações, etc.*

25

## EVENTOS E SEUS PARTICIPANTES

- **Semântica de eventos**
  - Predição de participantes/argumentos de um predicador
- 2 tipos principais de restrições sobre argumentos
  - Papéis semânticos
  - Restrições de seleção

26

## PAPÉIS SEMÂNTICOS

### ○ Exemplos

- *Sasha broke the window.*
  - $\exists e \text{ Breaking}(e) \wedge \text{Breaker}(e, \text{Sasha}) \wedge \text{BrokenThing}(e, \text{Window})$
- *Pat opened the door.*
  - $\exists e \text{ Opening}(e) \wedge \text{Opener}(e, \text{Pat}) \wedge \text{OpenedThing}(e, \text{Door})$
  - Tipos de sujeitos para cada verbo: Breaker e Opener
  - Papéis profundos são específicos para cada evento
    - Eventos do tipo *break* têm Breakers e BrokenThings
    - Eventos do tipo *open* têm Openers e OpenedThings

27

## PAPÉIS TEMÁTICOS

- Os papéis específicos não são muito úteis em PLN
- Mas **há coisas em comum** entre eles
  - Breakers e Openers são atores volitivos, animados
  - BrokenThings e OpenedThings são inanimados e afetados pela ação
- **Papéis temáticos** são um modelo particular de papéis semânticos e tentam capturar tais propriedades
  - Breakers e Openers são AGENTES de seus verbos
  - BrokenThings e OpenedThings são TEMAS de seus verbos

28

## PAPÉIS TEMÁTICOS

- Papéis temáticos propostos pelo gramático indiano Panini alguns séculos AC
- **Reformulação recente** por Fillmore (1968), principalmente
- Não há consenso sobre os papéis temáticos, mas alguns são bem aceitos

29

## PAPÉIS TEMÁTICOS

- Exemplos
  - AGENTE
  - EXPERIENCIADOR
  - TEMA
  - INSTRUMENTO
  - BENEFICIÁRIO
  - OBJETIVO
  - ORIGEM
  - RESULTADO
  - CONTEÚDO
  - Etc.

30

## REALIZAÇÕES VARIADAS DE PAPÉIS TEMÁTICOS

- Papéis temáticos e realizações variadas/alternadas
  - *John*<sub>AGENTE</sub> *broke the window*<sub>TEMA</sub>
  - *John*<sub>AGENTE</sub> *broke the window*<sub>TEMA</sub> *with a rock*<sub>INSTRUMENTO</sub>
  - *The rock*<sub>INSTRUMENTO</sub> *broke the window*<sub>TEMA</sub>
  - *The window*<sub>TEMA</sub> *broke*
  - *The window*<sub>TEMA</sub> *was broken by John*<sub>AGENTE</sub>
- Conjunto de papéis temáticos de um verbo: grade temática (**θ-grid**) ou estrutura de caso do verbo

31

## REALIZAÇÕES VARIADAS DE PAPÉIS TEMÁTICOS

- Algumas possibilidades para verbo *break*
  - AGENTE:sujeito, TEMA:objeto
  - AGENTE:sujeito, TEMA:objeto, INSTRUMENTO:SP
  - INSTRUMENTO:sujeito, TEMA:objeto
  - TEMA:sujeito
- Realizações múltiplas da estrutura de argumentos do verbo: *verb alternations* ou *diathesis alternations*
- Classes de verbos possuem alternâncias similares, por exemplo, verbos do tipo “ter” (*allocate, offer, owe*) e “enviar” (*forward, hand, mail*)
  - Levin (1993) e as classes de verbos
  - Incorporadas em repositórios como a *VerbNet* (Kipper et al., 2000)

32



## PAPÉIS TEMÁTICOS

- Por que isso é considerado semântica lexical?

33

## PAPÉIS TEMÁTICOS

- Por que isso é considerado semântica lexical?
  - É o verbo, o item léxico, que evoca uma estrutura de casos
- Muito útil para reconhecer variações de realização lingüística
- Repositórios amplamente usados: PropBank (Kingsbury e Palmer, 2002) e FrameNet (Baker et al., 1998)

34

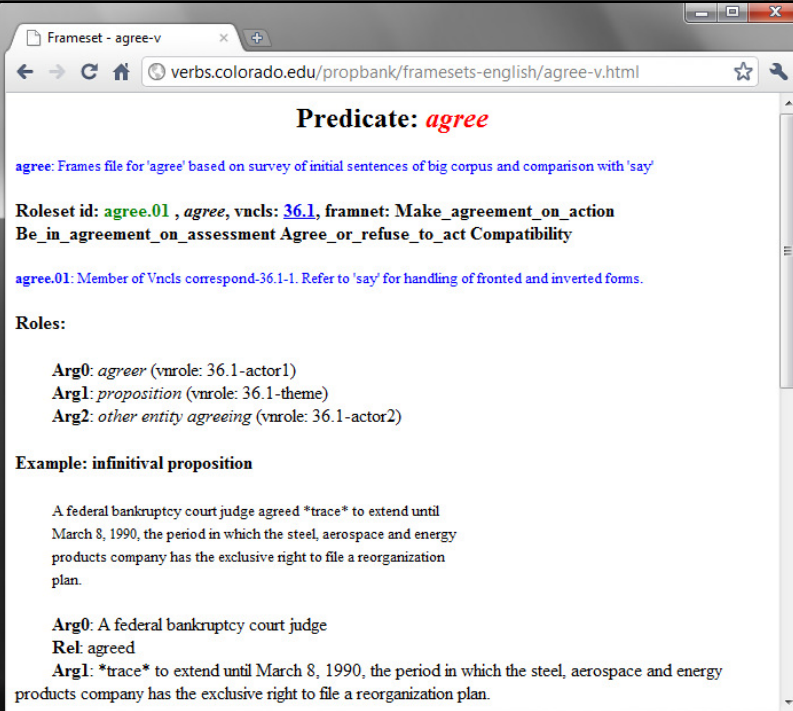
## PROPBANK

### o *Proposition Bank*

- Repositório de sentenças com papéis semânticos anotados
  - o Para o inglês, cópula *Penn Treebank* anotado
- Papéis genéricos do tipo Arg0, Arg1, Arg2, etc., específicos para cada verbo
  - o Há alguns padrões mais recorrentes: por exemplo, na maior parte dos casos, Arg0 é o AGENTE
- Uso de papéis genéricos
  - o Dificuldade de se ter papéis universais
  - o Vistas à automatização desse nível de análise

35

EXEMPLO



Frameset - agree-v

verbs.colorado.edu/propbank/framesets-english/agree-v.html

**Predicate: *agree***

agree: Frames file for 'agree' based on survey of initial sentences of big corpus and comparison with 'say'

Roleset id: **agree.01**, *agree*, vncls: [36.1](#), framnet: **Make\_agreement\_on\_action**  
**Be\_in\_agreement\_on\_assessment Agree\_or\_refuse\_to\_act Compatibility**

agree.01: Member of Vncls correspond-36.1-1. Refer to 'say' for handling of fronted and inverted forms.

**Roles:**

**Arg0:** *agreeer* (vnrole: 36.1-actor1)  
**Arg1:** *proposition* (vnrole: 36.1-theme)  
**Arg2:** *other entity agreeing* (vnrole: 36.1-actor2)

**Example: infinitival proposition**

A federal bankruptcy court judge agreed \*trace\* to extend until March 8, 1990, the period in which the steel, aerospace and energy products company has the exclusive right to file a reorganization plan.

**Arg0:** A federal bankruptcy court judge  
**Rel:** agreed  
**Arg1:** \*trace\* to extend until March 8, 1990, the period in which the steel, aerospace and energy products company has the exclusive right to file a reorganization plan.

## FRAMENET

- **FrameNet**: representação via frames e elementos de frames
  - Frames são para várias classes de palavras, não só verbos
  - Permitem reconhecer, por exemplo, a correspondência entre as estruturas de argumentos abaixo
    - [The price of bananas]<sub>ARG1</sub> increased [5%]<sub>ARG2</sub>
    - [The price of bananas]<sub>ARG1</sub> rose [5%]<sub>ARG2</sub>
    - There has been a [5%]<sub>ARG2</sub> rise [in the price of bananas]<sub>ARG1</sub>

37

## FRAMENET

- **FrameNet**
  - Frames têm elementos essenciais e não essenciais
  - Exemplo: frame **change\_position\_on\_a\_scale**
    - Válido para diversas palavras
      - Verbos: *advance, climb, decline, decrease*, etc.
      - Substantivos: *rise, shift*, etc.
      - Advérbios: *increasingly*
    - Há herança entre frames e também relações entre eles

38

The screenshot shows a web browser window with the URL `framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=118&frame=Change_position`. The page title is "Frame Report (recent data)" and the main heading is "Change\_position\_on\_a\_scale".

**Definition:**  
 This frame consists of words that indicate the change of an **Item**'s position on a scale (the **Attribute**) from a starting point (**Initial value**) to an end point (**Final value**). The direction (**Diff**) of the movement can be indicated as well as the magnitude of the change (**Difference**). The rate of change of the value (**Rate**) is optionally indicated. Another scale (**Scale**), which the values are correlated with, is indicated if it is not the default correlate (namely, absolute time).

The distinction between **Attributes** and **Items** is not always an easy one. The clear cases involve the expression of the **Attribute** in an *i*-PP.  
 Hawke's Bay winery **DOUBLED** in size last year.  
 The amount you can deduct, then, depends on whether or not the stock has **INCREASED** in value during the period you have owned it.

Other clear cases of **Attributes** involve NP's like *size, quality, number, value* that denote abstract attributes rather than events or classes of things.

The analysis is more complicated when an event-denoting noun phrase occurs as the subject and the sentence has no *i*-PP. In the simplest cases, where the **Attribute** is the number of occurrences of the event, the **Attribute** is usually left implicit, as in exx. 1 and 2, the **Attribute** below.

(1) Accidents **INCREASED** 20% to 345.  
 (1--second FE layer) Accidents **INCREASED** 20% to 345.  
 (2) Attacks on civilians **DECREASED** over the last 4 months DNI.

In cases like (1), we will tag the **Attribute** on the second layer below the **Final value** label, as that FE makes it clear that the **Attribute** is simply cardinality. In cases like (2), we tag it as DNI because in context the hearer

The screenshot shows the "FEs" section of the FrameNet website. It lists various frame elements (FEs) and their core meanings, each with an example sentence.

**FEs:**

**Core:**

- Attribute [att]**: The **Attribute** is a scalar property that the **Item** possesses.  
 ex) XOSE in price by 2%.
- Difference [Diff]**: The distance by which an **Item** changes its position on the scale.  
 ex) XOSE in price by 2%.
- Final state [finis]**: A description that presents the **Item**'s state after the change in the **Attribute**'s value as an independent predication.  
 It was never bad (1 or 2 seizures a year), but this past decade, it has **INCREASED** to having them 1 day a month and on that 1 day I have 6 or 7 seizures.
- Final value [val2]**: The position on the scale where the **Item** ends up.  
 Microsoft shares **FELL** to 7.5/8.
- Initial state [inis]**: A description that presents the **Item**'s state before the change in the **Attribute**'s value as an independent predication.  
 Diesels have **INCREASED** from having a 20% market share in 1995 to just over 30% in 2004.
- Initial value [val1]**: The initial position on the scale from which the **Item** moves away.  
 Microsoft shares **FELL** from 12.3/8 to 7.5/8.
- Item [item]**: The entity that has a position on the scale.

FrameNet

framenet.icsi.berkeley.edu/index.php?option=com\_wrapper&Itemid=118&frame=Change\_position

**Non-Core:**

**Circumstances [Gr]** The **circumstances** under which the change of position takes place.  
The eggs will only **INCREASE** **under pressure**.

**Correlate [cor]** The **correlate** is a directional path (or ordered collection of values) that the (dependent) **Attribute** is measured against. Note that normally, the **correlate** is simply (forward-directional) time.  
The amount of power **INCREASES** with the frequency of the laser.  
As you travel along the **Quan-duo road**, the poverty level **SOARS** dramatically.

**Degree [Deg]** **Degree** to which event occurs  
**Semantic Type**  
Degree

**Duration [Dur]** The length of time over which the change takes place.  
**Semantic Type**  
Duration  
The amount of troops in the war will no doubt **DOUBLE** over the next several years.

**Final\_correlate [cor2]** The value, location, or state-of-affairs that corresponds to the **Final value** of the **Attribute**.  
From Monterrey **to southern Oregon**, tree height **INCREASES** steadily in this species to a stunning 50 meters.

**Group [Grp]** The **Group** in which an **Item** changes the value of a **Attribute** in a specified way.  
Colon cancer incidence **FELL** by 50 % **among men over 30**.

FrameNet

framenet.icsi.berkeley.edu/index.php?option=com\_wrapper&Itemid=118&frame=Change\_position

**Semantic Type**  
Locative\_relation

**Result [Result]** **Result**. of an event

**Speed [Speed]** The rate of change of the Value.  
**Semantic Type**  
Speed  
Prices are **RISING** by 2% a year.

**Time [T]** The **Time** is the time-frame in which the change of position occurs.  
**Semantic Type**  
Time

Inherits From:  
Is Inherited By: **Proliferating\_in\_number**  
Subframe of:  
Has Subframes:  
Precedes:  
Is Preceded by:  
Uses:  
Is Used By: **Change\_of\_quantity\_of\_possession**, **Remainder**  
Perspective on:  
Is perspectivized in:  
Is Causative of:  
See Also:

**Lexical Units**

*advance.v, balloon.v, dim.v, decline.n, decline.v, decrease.n, decrease.v, diminish.v, dip.v, double.v, drop.v, dwindle.v, edge.v, escalation.n, explode.v, explosion.n, fall.n, fall.v, fluctuate.v, fluctuation.n, gain.n, gain.v, grow.v, growing.a, growth.n, hike.n, increase.n, increase.v, increasingly-adv, jump.v, move.v, mushroom.v, plummet.v, reach.v, rise.n, rise.v, rocket.v, shift.n, shift.v, skyrocket.v, slide.v, soar.v, swell.v, swing.v, triple.v, tumble.n, tumble.v*

Created by josef on Mon Mar 05 17:02:03 PST 2001

Frame Index

framenet.icsi.berkeley.edu/fnReports/data/frameIndex.xml?banner=/fnReports/banner.html

## Frame Index

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#)  
[J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#)  
[T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

[Abandonment](#)  
[Abounding\\_with](#)  
[Absorb\\_heat](#)  
[Abundance](#)  
[Abusing](#)  
[Access\\_scenario](#)  
[Accompaniment](#)  
[Accomplishment](#)  
[Accountrements](#)  
[Accuracy](#)  
[Achieving\\_first](#)  
[Active\\_substance](#)  
[Activity](#)  
[Activity\\_abandoned\\_st](#)  
[Activity\\_done\\_state](#)  
[Activity\\_finish](#)  
[Activity\\_ongoing](#)

### Lexical Units:

*advance.v, balloon.v, climb.v, decline.n, decline.v, decrease.n, decrease.v, diminish.v, dip.v, double.v, drop.v, dwindle.v, edge.v, edged\_down.v, escalation.n, explode.v, explosion.n, fall.n, fall.v, fluctuate.v, fluctuation.n, gain.n, gain.v, grow.v, growing.a, growth.n, hike.n, increase.n, increase.v, increasingly.adv, jump.v, move.v, mushroom.v, plummet.v, reach.v, rise.n, rise.v, rocket.v, shift.n, shift.v, skyrocket.v, slide.v, soar.v, swell.v, swing.v, triple.v, tumble.n, tumble.v*

Created by 605 on 03/05/2001 05:02:03 PST Mon

| Lexical Unit | LU Status        | Lexical Entry Report          | Annotation Report          | Annotator ID | Created Date                |
|--------------|------------------|-------------------------------|----------------------------|--------------|-----------------------------|
| advance.v    | Finished_Initial | <a href="#">Lexical entry</a> | <a href="#">Annotation</a> | 605          | 03/09/2001 04:46:11 PST Fri |
| balloon.v    | Created          | <a href="#">Lexical entry</a> |                            | 361          | 08/02/2007 12:22:16 PDT Thu |
| climb.v      | Finished_Initial | <a href="#">Lexical entry</a> | <a href="#">Annotation</a> | 605          | 11/15/2005 01:46:20 PST Tue |

Change\_position\_on\_a\_s...

framenet.icsi.berkeley.edu/fnReports/data/lu/lu1893.xml?mode=lexentry&banner=/fnReports/banr

## Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

| Frame Element                 | Number Annotated | Realization(s)   |
|-------------------------------|------------------|--|
| <a href="#">Attribute</a>     | (39)             | 2nd.-- (3)<br>CNI.-- (7)<br>DNI.-- (13)<br>NP.Ext (15)<br>PP[with].Dep (1)                             |
| <a href="#">Circumstances</a> | (16)             | 2nd.-- (1)<br>PP[in].Dep (9)<br>PP[as].Dep (2)<br>PP[after].Dep (1)<br>VPing.Dep (1)<br>PP[on].Dep (2) |
| <a href="#">Correlate</a>     | (5)              | 2nd.-- (1)<br>PP[against].Dep (1)<br>PP[as].Dep (2)<br>PP[in].Dep (1)                                  |
| <a href="#">Degree</a>        | (4)              | AVP.Dep (4)  |
|                               |                  | NP.Dep (7)   |

[Clear Sentences](#) [Turn Colors Off](#)

Change\_position\_on\_a\_s... x

framenet.icsi.berkeley.edu/fnReports/data/lu/lu1893.xml?mode=annotation&banner=/fnReports/b

- T-Wat-Wpace.rate-(1)
  1. The announcement stated that `` [Item economic activity] has continued to *ADVANCED Target* [Speed at a substantial pace] . [Attribute DNI][Difference INI]
  2. A flurry of data released Thursday said [Item spending] *ADVANCED Target* [Speed at its slowest pace since April] and renewed speculation that the Federal Reserve is poised for another interest-rate cut . [Attribute DNI][Difference INI]
  3. A month ago , the government said [Item gross domestic product , measuring all the goods and services produced in the nation] , *ADVANCED Target* [Speed at a 1.1 percent rate] [Duration in the April -June quarter] . [Attribute DNI][Difference INI]
  4. [Item It] *ADVANCED Target* [Speed at a 5.4 percent rate] [Initial correlate from October] [Final correlate through December] , compared with 4.3 percent from July through September . [Attribute DNI][Difference INI]
  5. [Item Manufacturing productivity] *ADVANCED Target* [Speed at a 2.6 percent rate] [Initial correlate from April] [Final correlate through June] , [Speed slower than the 3.5 percent gain three months earlier] . [Attribute DNI][Difference INI]
  6. [Item The economy] grew 4.1 percent in 1994 , its best showing in a decade , and closed out the year by *ADVANCING Target* [Speed at a booming 5.1 percent annual rate] [Time in the fourth quarter] . [Attribute CNI][Difference INI]
  7. The government reported earlier [Item the gross domestic product] *ADVANCED Target* [Speed at a 2.8 percent annual rate] [Initial correlate from January] [Final correlate through March] . [Speed much slower than the 5.1 percent rate posted during the final three months of 1994] . [Attribute DNI][Difference INI]
- T-Wmoderately-(1)
  1. TOKYO ( AP ) --- [Item The Tokyo Stock Exchange 's main index] *ADVANCED Target* [Degree moderately] [Time Thursday] , [Result rising for the third session in a row to close at its highest level this year] . [Attribute CNI][Difference INI]
  2. HONG KONG ( AP ) --- Asian stock markets closed mostly higher Thursday , with [Attribute share prices in Tokyo] *ADVANCING Target* [Degree moderately] [Result to close at their highest level this year] . [Difference INI] [Item share]

## FRAMENET

### o Exemplo de anotação

- [Oil]<sub>ITEM</sub> *rose* in [price]<sub>ATTRIBUTE</sub> by [2%]<sub>DIFFERENCE</sub>
- A [5%]<sub>DIFFERENCE</sub> [dividend]<sub>ITEM</sub> *increase*...
- [Microsoft shares]<sub>ITEM</sub> *fell* [to 7 5/8]<sub>FINAL\_VALUE</sub>

## REPOSITÓRIOS

- Várias iniciativas para integrar repositórios
  - Por exemplo, projeto SemLink: PropBank, VerbNet, FrameNet e WordNet (Loper et al., 2007)
- Vários repositórios disponíveis, principalmente para o inglês
- E para o português?

47

## RESTRICÇÕES DE SELEÇÃO

- Assim como papéis semânticos, definem informações sobre argumentos de predicadores
- Restrições sobre os tipos de argumentos que um predicador pode ter
  - Não somente verbos, mas substantivos e adjetivos também
- Exemplo
  - *Eu ingeri uma fruta.*
    - Ingerir pede um AGENTE do tipo animado e um TEMA do tipo ingerível
      - Ingerir(animado,ingerível)

48



## RESTRICÇÕES DE SELEÇÃO

- Assim como papéis semânticos, definem informações sobre argumentos de predicadores
- Restrições sobre os tipos de argumentos que um predicador pode ter
  - Não somente verbos, mas substantivos e adjetivos também
- Exemplo
  - Outras classes além dos verbos?

49

## RESTRICÇÕES DE SELEÇÃO

- Assim como papéis semânticos, definem informações sobre argumentos de predicadores
- Restrições sobre os tipos de argumentos que um predicador pode ter
  - Não somente verbos, mas substantivos e adjetivos também
- Exemplo
  - Outras classes além dos verbos?
    - *Ela está pálida.* → adjetivo *pálida* pede algo que tenha cor
    - *To bake a cake.* → *cake* pede *bake* (com traço “que assa”), e não *make*

50

## RESTRIÇÕES DE SELEÇÃO

- Alguns predicadores são menos exigentes, outros são mais
- Conjunto de tipos semânticos dos argumentos é aberto
- Cada predicador tem várias restrições, dependendo do sentido do predicador
  - Grande elenco de restrições é necessário
    - Poucos repositórios disponíveis
  - Alternativa: associar os tipos dos argumentos aos **synsets da WordNet**

51

## RESTRIÇÕES DE SELEÇÃO

- Restrições e WordNet
  - Exemplo
    - *Ingerir({*living thing, animate thing*}, {*food, nutrient*})*
      - Quaisquer argumentos que pertençam aos ramos ontológicos desses synsets são aceitos
- Logicamente, WordNet não é completa e outros repositórios podem ser usados

52

## DECOMPOSIÇÃO EM PRIMITIVAS DE SIGNIFICADO

- Uma maneira de definir uma palavra é via seus atributos/componentes de significado
    - Papéis temáticos são um aspecto disso
  - Várias abordagens
    - Atributos/traços semânticos (Wierzbicka, 1992, 1996)
      - *Hen*                    +female, +chicken, +adult
      - *Rooster*              -female, +chicken, +adult
      - *Chick*                 +chicken, -adult
- Variações de galinha (em relação à idade e ao sexo)

53

## DECOMPOSIÇÃO EM PRIMITIVAS DE SIGNIFICADO

- Uma maneira de definir uma palavra é via seus atributos/componentes de significado
  - Papéis temáticos são um aspecto disso
- Várias abordagens
  - Atos primitivos (a la Schank)
    - João matou sua esposa.
      - MATAR(x,y) ⇒ CAUSAR(x, TORNAR-SE(NÃO(VIVO(y))))
    - João abriu a porta.
      - ABRIR(x,y) ⇒ CAUSAR(x, TORNAR-SE(ABERTA(y)))

54

## OUTROS REPOSITÓRIOS

- Diversas fontes de conhecimento relevantes
  - Bases de **senso comum**
    - Projetos OMCS: MIT, UFSCar
  - Conhecimento de mundo/**enciclopédico**
    - Wikipédia
  - Gazeteers/**almanaques** (listagens de nomes classificados)
    - REPENTINO – REpositório para reconhecimento de ENTidades com NOme (Sarmiento et al., 2006)
  - **Thesaurus** e **wordnets** afetivas, terminológicas, de domínio, etc.
  - **LDOCE** – *Longman Dictionary of Contemporary English*
  - Listagens de paráfrases, marcadores discursivos, etc.

55



## MÉTODOS COMPUTACIONAIS: ALGUNS EXEMPLOS

**Desambiguação Lexical de Sentido**  
(*Word Sense Disambiguation*)

## DESAMBIGUAÇÃO LEXICAL DE SENTIDO (DLS)

- **Definição genérica**
  - Dada uma palavra e seu contexto
  - Dada uma listagem de possíveis sentidos da palavra
  - Tarefa de **atribuição de um sentido à palavra**
  
- Análise semântica completa requer a associação de sentido a cada palavra
  
- Pode ser útil em diversas tarefas de PLN
  - Tradução automática, sumarização, perguntas e respostas, etc.

57

## DESAMBIGUAÇÃO LEXICAL DE SENTIDO (DLS)

- Adaptável para cada aplicação
  - Tradução automática
    - Palavra em português → possíveis traduções em inglês
  
  - Síntese de fala
    - Palavra → homógrafos com diferentes pronúncias
  
  - Indexação de artigos
    - Título → organização do campo de conhecimento, thesaurus de domínio
  
- Genericamente, usa-se muito a WordNet e o LDOCE como inventário de sentidos

58

## DESAMBIGUAÇÃO LEXICAL DE SENTIDO (DLS)

### ○ Exemplo

- Em inglês, *bass*
  - Instrumento musical, cantor ou peixe?

Sentidos na WordNet de Princeton

#### Noun

- [S](#): (n) **bass** (the lowest part of the musical range)
- [S](#): (n) **bass**, [bass part](#) (the lowest part in polyphonic music)
- [S](#): (n) **bass**, [basso](#) (an adult male singer with the lowest voice)
- [S](#): (n) [sea bass](#), **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- [S](#): (n) [freshwater bass](#), **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- [S](#): (n) **bass**, [bass voice](#), [basso](#) (the lowest adult male singing voice)
- [S](#): (n) **bass** (the member with the lowest range of a family of musical instruments)
- [S](#): (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

## TIPOS DE DLS

### ○ 2 principais variantes

- **Algumas palavras** em particular e seus possíveis sentidos
  - Tarefa mais controlada
    - Vários “pequenos” cópus disponíveis, eventos internacionais (por exemplo, avaliações do SENSEVAL)
  - Aprendizado de máquina supervisionado pode ajudar
    - Uso de atributos do contexto das palavras: por exemplo, palavras e etiquetas morfossintáticas anteriores e posteriores
  - Exemplo: aprendizado de como escolher o sentido adequado da palavra **bass** em seus vários contextos de uso

## TIPOS DE DLS

### ○ 2 principais variantes

- **Algumas palavras** em particular e seus possíveis sentidos

- Exemplos de “regras” aprendidas automaticamente

palavra *fish* no contexto → peixe  
*striped bass* → peixe  
 palavra *guitar* no contexto → instrumento  
*bass player* → instrumento

...

61

## TIPOS DE DLS

### ○ 2 principais variantes

- **Todas as palavras de conteúdo de um texto** e um repositório enorme de sentidos

- Dados esparsos

- Poucos córpus disponíveis

- Por exemplo, SemCor (parte do Brown Corpus), manualmente anotado com sentidos da WordNet

- Aprendizado supervisionado é problemático

- Uso de dicionários e *bootstrapping*

62

## TIPOS DE DLS

### ○ 2 principais variantes

- **Todas as palavras de conteúdo de um texto** e um repositório enorme de sentidos

#### ○ Uso de dicionários

##### ○ Método de Lesk e suas variantes

- Método simplificado de Lesk (Kilgarriff e Rosenzweig, 2000)
  - Dada uma palavra e seu contexto
  - E suas entradas em um dicionário
  - Escolhe-se o sentido cuja entrada tenha mais palavras em comuns com a palavra e seu contexto

63

## TIPOS DE DLS

### ○ 2 principais variantes

- **Todas as palavras de conteúdo de um texto** e um repositório enorme de sentidos

#### ○ Uso de dicionários

##### ○ Método de Lesk e suas variantes

- **Exemplo** do método simplificado de Lesk

- *The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.*

- Entradas na WordNet

- *bank: a financial institution that accepts **deposits** and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the **mortgage** on my home"*
- *bank: sloping land - especially the slope beside a body of water; "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*

64



## TIPOS DE DLS

### o 2 principais variantes

- Todas as palavras de conteúdo de um texto e um repositório enorme de sentidos

- o Restrições/preferências de seleção

- o Argumentos esperados pelos verbos podem ajudar

- *"In our house, everybody has a career and none of them includes washing dishes", he says.*

[+lavável]

- *In her tiny kitchen at home, Ms. Chen works efficiently, stir-frying several simple dishes, including braised pig's ears and chicken livers with green peppers.*

[+comestível]

65

## MÉTODOS COMPUTACIONAIS: ALGUNS EXEMPLOS

Anotação de Papéis Semânticos  
(*Semantic Role Labeling*)

## ANOTAÇÃO DE PAPÉIS SEMÂNTICOS (APS)

- Um tipo “mais raso” de análise semântica
- **Definição genérica**
  - Tarefa de se identificar os argumentos e seus papéis semânticos para cada predicado em uma sentença
  - Exemplos
    - Estilo **FrameNet**: [You]<sub>COGNIZER</sub> can't [blame]<sub>TARGET</sub> [the program]<sub>EVALUEE</sub> [for being unable to identify it]<sub>REASON</sub>
    - Estilo **PropBank**: [The San Francisco Examiner]<sub>ARG0</sub> [issued]<sub>TARGET</sub> [a special edition]<sub>ARG1</sub> [yesterday]<sub>ARGM-TMP</sub>
- Comum para extração de informação e perguntas e respostas
- **Útil** para diversas ferramentas e aplicações de PLN
  - Exemplos?

67

## MÉTODO DE ANÁLISE

- **Método usual**: aprendizado de máquina supervisionado
  - **Passo 1 – treinamento** (com base na FrameNet ou no PropBank, por exemplo)
    - Sentenças são analisadas sintaticamente
    - Para cada predicado da sentença
      - Para cada possível argumento na árvore
        - Extrai-se um conjunto de atributos
        - Associa-se o tipo de argumento (classe)
        - Gera-se um classificador
  - **Passo 2 – teste/uso**
    - Dada uma nova sentença
      - Analisa-a sintaticamente
      - Buscam-se os predicados
      - Para cada possível argumento do predicado, faz-se classificação do tipo do argumento

68

## MÉTODO DE ANÁLISE

- Método usual: aprendizado de máquina supervisionado
  - **Atributos usuais de um possível argumento**
    - Predicado
    - Tipo de sintagma
    - *Head* do sintagma
    - Etiqueta morfossintática
    - Caminho na árvore sintática
    - Voz da sentença (ativa vs. passiva)
    - Posição do argumento em relação ao predicado (antes ou depois)
    - Subcategorização/regra do predicado
    - Tipo de entidade nomeada
    - Etc.
  - **Cuidado especial**
    - Muitos argumentos não devem ser classificados
    - Não se devem ter argumentos com sobreposição

69

## MÉTODO DE ANÁLISE

- Método usual: aprendizado de máquina supervisionado
  - **Possibilidade de processamento em várias etapas**
    - Poda da árvore sintática da sentença
    - Identificação de argumentos válidos
    - Classificação dos argumentos “filtrados”
  - Em vez da árvore sintática completa, pode-se ter *chunks*, dependências, etc.

70