

Curso avançado de estudos contrastivos usando o COMPARA como ferramenta

Diana Santos

Linguatca
www.linguatca.pt

Três partes

1. A idade da inocência
2. A hora da verdade
3. A tempo?

Pressuposto deste curso:

As pessoas interessadas em estudos contrastivos, e em corpos paralelos, estão mais habilitadas do que um vulgar engenheiro a compreender uma sintaxe de procura não trivial, visto que são especialistas em línguas e linguagens 😊

Índice

- Estudos contrastivos?
- Corpos paralelos
 - A história do COMPARA
 - Outros corpos que incluem o português
- O COMPARA por miúdos
 - Os textos
 - A marcação
 - A anotação
 - A interacção
- Questões técnicas
 - O sistema de procura
 - A anotação automática ... e a revisão
- Problemas no COMPARA
- Problemas/questões difíceis em geral
- E várias perguntas, exercícios, exemplos

Estudos contrastivos

- O que são? Estudos que comparam mais do que uma língua para iluminar a forma como as línguas “funcionam”
- Para que servem?
 - Para a ciência, para a comunicação, para a paz, para a tecnologia
- O meu credo:
 - A melhor forma de fazer estudos contrastivos é usar a tradução feita sem essa preocupação, ou seja, a tradução real
- Diferença fundamental entre estudos contrastivos e de tradução!
 - A tradução é um meio, não um fim
 - Os tradutores são informantes inconscientes

Apresentação oficial do COMPARA

- o maior corpus paralelo editado do mundo
 - 3 milhões de palavras, 100.000 “frases alinhadas”
- textos literários (originais e traduções) em **português e inglês**, 35 autores, 45 tradutores
- projecto no âmbito da **Linguateca**, iniciado em 2000
- liderado por **Ana Frankenberg-Garcia e Diana Santos**
 - Rosário Silva, Susana Inácio, Pedro Sousa Martins
- anotação morfossintáctica
 - parte portuguesa - PALAVRAS, parte inglesa – CLAWS
- <http://www.linguateca.pt/COMPARA>



Breve história dos corpos na Linguateca

TALC 2008

- 1. tornar acessível (na rede) o que já existia
- 2. criar e/ou melhorar (analisando) esses corpos
- 3. adicionar a dimensão da tradução → COMPARA
- 4. adicionar a dimensão da revisão humana → Floresta Sintá(c)tica
- 5. adicionar a possibilidade de criar corpos próprios
- 6. adicionar a dimensão de corpos comparáveis

Projecto AC/DC

Corpógrafo

Relação com outras iniciativas do AC/DC

TALC 2008

- Textos fixos, anotação básica do PALAVRAS

Anotação hierárquica
Revisão humana

AC/DC

Alinhamento
Revisão humana

Floresta

COMPARA

- Textos da responsabilidade e escolha dos participantes

Corpógrafo

Breve história do COMPARA

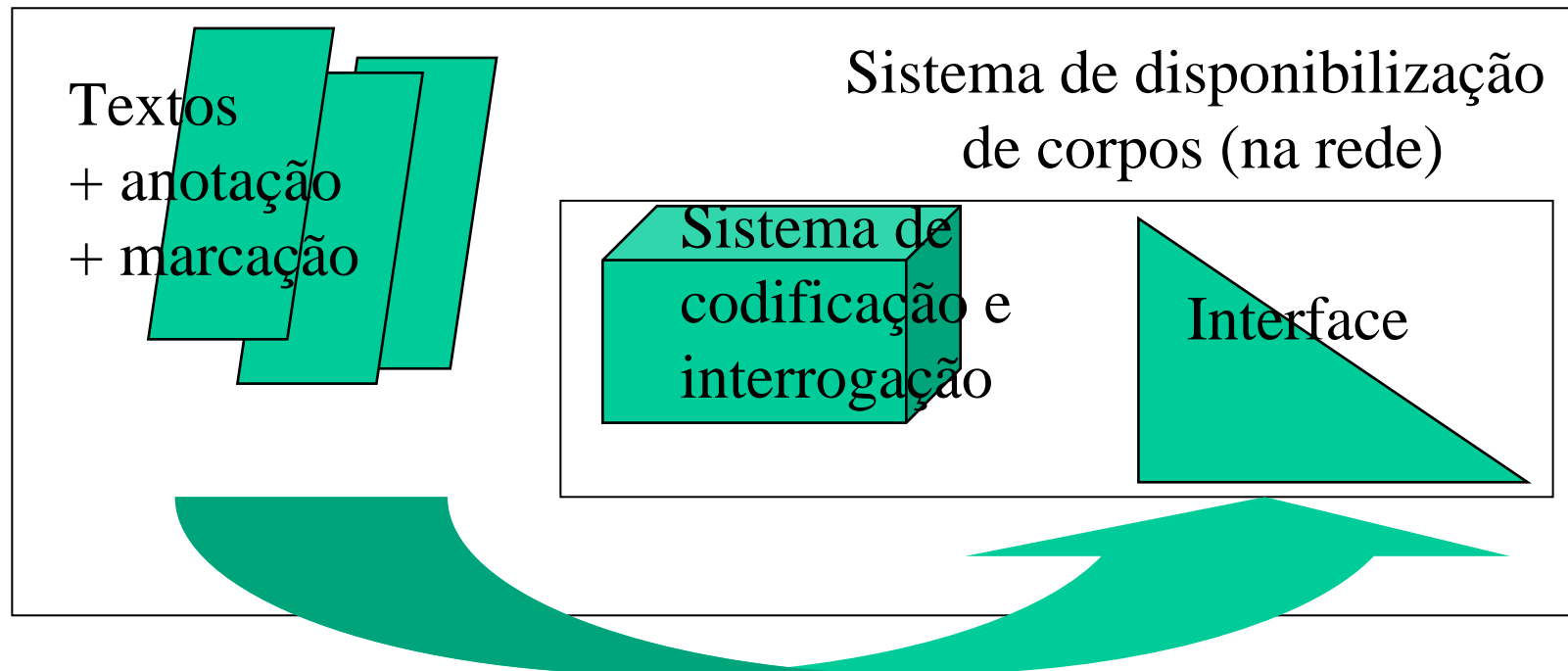
- 19 de Fevereiro de 1999: Encontro electrónico entre a Ana e eu
- 2 de Novembro de 1999: Convite meu para fazermos em conjunto um corpo paralelo
 - Eu estava a fazer o AC/DC
 - A Ana estava em Oxford para começar a criar um corpo para o ensino
- 3 a 7 de Abril 2000: visita da Ana a Oslo
- 3 de Maio de 2000: separação conceptual DISPARA e COMPARA
- 11 de Maio de 2000: primeira versão (não anunciada) na Rede
- 18 de Setembro de 2000: primeira apresentação exterior do COMPARA (Diana, Univ. Oslo)

Breve história do COMPARA (cont.)

- 3 de Novembro de 2000: primeira apresentação internacional do COMPARA (Ana, no CULT 2000 em Bertinoro, Itália)
- 9 de Janeiro de 2001: anúncio do COMPARA nas listas
- Janeiro de 2001: primeiros estágios, não remunerados, no COMPARA
- Novembro de 2002: URL passa a www.linguateca.pt/COMPARA/
- Novembro de 2004: começo da revisão da anotação do PALAVRAS
- Janeiro de 2005: primeira versão do COMPARA com toda a parte portuguesa anotada (ainda não revista) pelo PALAVRAS
- Fevereiro de 2008: primeira versão do COMPARA com toda a parte inglesa anotada (ainda não revista) pelo CLAWS
- Julho de 2008: começo da revisão da anotação da parte inglesa

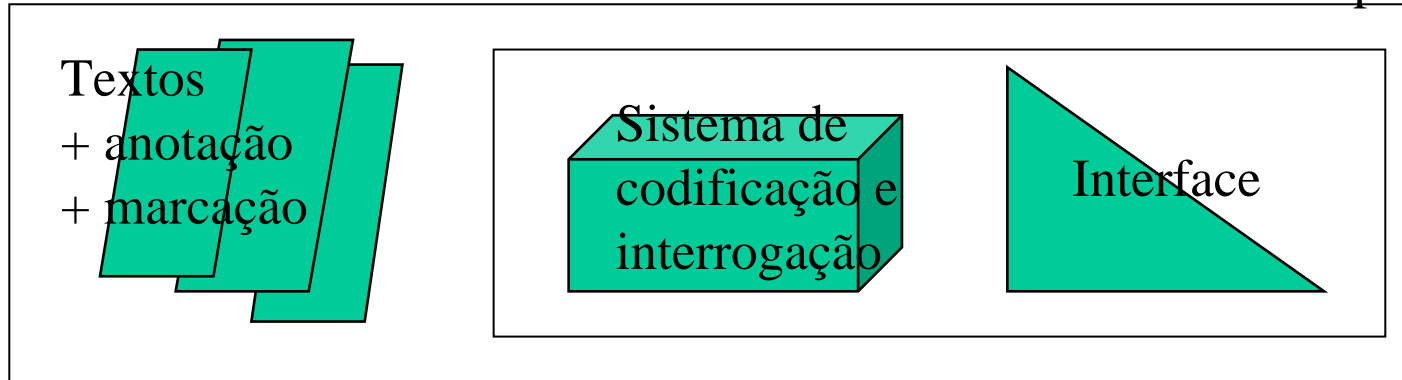
Multidimensionalidade da noção de corpo

Corpo / projecto

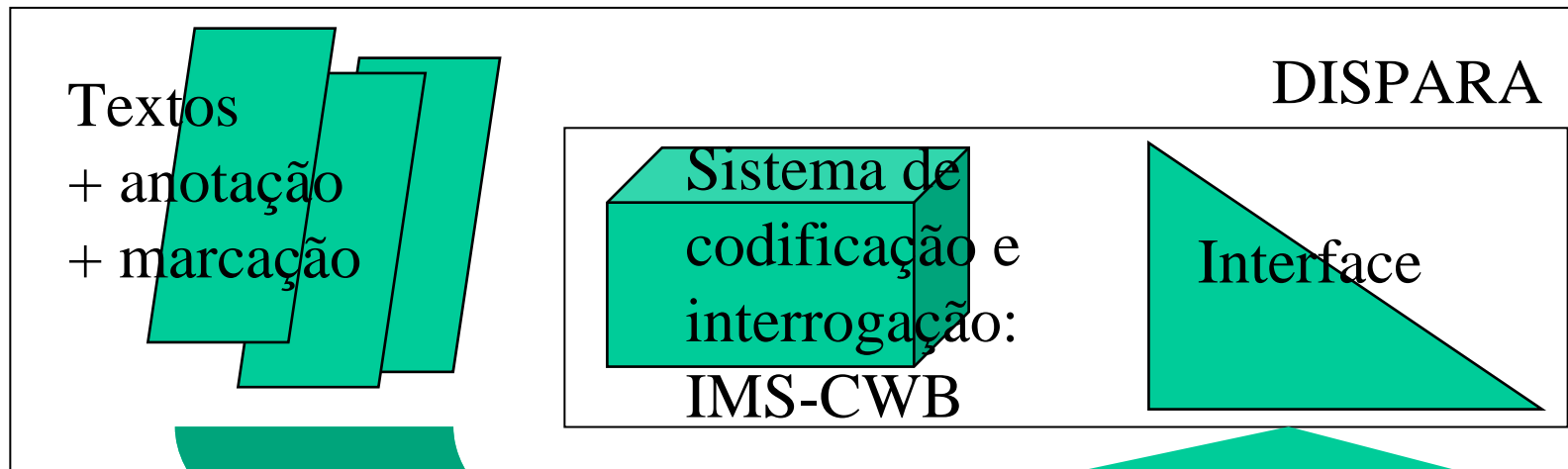


Aplicando o modelo ao COMPARA

Corpo



COMPARA

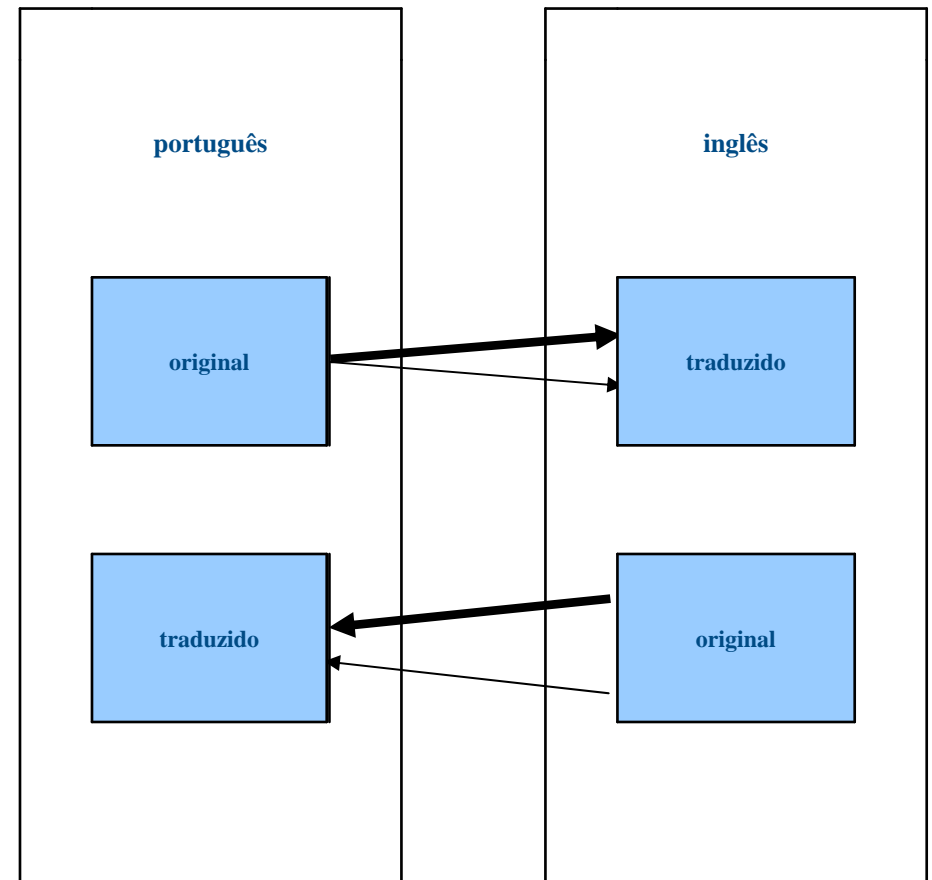


Corpos paralelos que incluem o português

- ENPC <http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/engorig.html>
- EuroParl <http://www.statmt.org/europarl/>
- JRCAquis <http://langtech.jrc.it/JRC-Acquis.html>
- OPUS <http://urd.let.rug.nl/tiedeman/OPUS/>
- NAT-QI <http://linguateca.di.uminho.pt/nat>
- CorTrad <http://nilc.icmc.usp.br/dispara/CorTrad/>
- ...

A estrutura do COMPARA

- Inspirada pelo ENPC
- Integrável no AC/DC (parte portuguesa) e no BNC e outros (parte inglesa)
- Mais do que uma tradução por original (2 casos)



Usos do COMPARA

- Para estudar a tradução
- Para estudar a diferença entre as línguas

- Para ensinar tradução
- Para ensinar línguas

- Para criar dicionários bilingues
- Para criar materiais de ensino

- Para simples instrumento de consulta

Um exemplo de cada

- Como é que se traduz *supposed to*
- Os modais *poder* e *can/could* são usados da mesma maneira?
- Observar os casos de tradução de *andar a –ar /actually*
- Quais as traduções de *bom*?
- Através do estudo das concordâncias paralelas de palavras com frequência maior do que 15...
- Obtendo bons exemplos para testes de substituição ou preenchimento
- Como se escreve *publicamente* em inglês: *publicly* ou *publically*?

Exemplos práticos de exploração

- Que verbos são mais usados com o advérbio *home*?
- Que tradução é mais comum para *despedir* ?
- O que é mais azul em português? E em inglês?
- E cor-de-rosa?

- Qual a construção mais comum com o verbo *dizer*?
- Que verbos são mais usados em orações relativas?
- Com que verbo *não* é mais frequente?

- Quando é que *sempre* é usado?

Questões complexas nos bastidores

- Nem todos os textos têm a mesma informação
 - os textos já digitalizados têm uma qualidade muito variável
 - e escolhas diferentes no que marcaram (notas de tradução, ênfase, etc.)
- Textos com traduções diferentes contam duas vezes
 - Por vezes é legítimo
 - A maioria das vezes não o é
- Editoras diferentes têm normas de publicação diferentes
 - Discurso directo
 - Títulos de capítulos
- Tempos diferentes... línguas diferentes
 - Ortografia
 - Estrangeirismos

A marcação

(algo com que que faz sentido marcar pedaços de texto)

- Diferenças de alinhamento

- Distinção entre reordenamento e alinhamento

- Diferenças na pontuação

- O que é uma frase? Qual a unidade para contar frases?

- Classificação entre: palavra ou expressão estrangeira, título, nome de marca ou simples ênfase?

- Duas opções: classificação exclusiva, ou independente?

Interlúdio sobre categorização

- A que categoria semântica pertence *Salamandra Roxa* e *Globo Azul* ?
PPLJ1(1071): *Helena passou a outro envelope, **Salamandra Roxa** , spoilt*
- PBPC2(123): *Petrus me disse que eu sabia despertar *Ágape*, através do **Globo Azul**.*
- A que categoria gramatical pertence *A Grã-Bretanha Precisa das Suas Universidades*. ?
EBDL2(724): *O seu carro está estacionado lá fora, um Renault vermelho com seis anos, com um autocolante amarelo no vidro traseiro: «**A Grã-Bretanha Precisa das Suas Universidades**.»*
- A que grupo de cor pertence *cor de osso* ?

O que significa *verde*? Como interpretá-lo?

- *PBOL1(248): Desejaria ser, em parte, como essa adolescente, e sustentar com doçura, ano após ano, também emoldurada, meu ramo sempre **verde**, sua corola imortal.* [sema="cor_naomaduro"]
- *PBCB2(202): Um cacho de bananas **verdes** no chão da cozinha lembra-me que passei o dia a chá e bolacha.* [sema="cor_naomaduro"]
- *PBJS1(1126): Não entendo por que é que, pelo menos, não alugaram um cupê para andar em volta do parque -- disse o capitão, referindo-se às elegantes carruagens de aluguel decoradas com espelhos, seda **adamascada** e contornos de prata, verdadeiras camas ambulantes, que eram anunciadas diariamente nos jornais.* [sema="cor_0"]
- *EBDL1T2(867): A área tem muito **verde**, é arborizada e quieta.* [sema="cor_0"]

A anotação

(algo com que faz sentido marcar todo o texto)

- Se há coisa que não é consensual é a gramática de uma língua quando se chega aos próprios textos...
- O que é um adjectivo?
- O que é um substantivo?
- O que é uma palavra?
- É preciso fazer escolhas e decisões e documentá-las

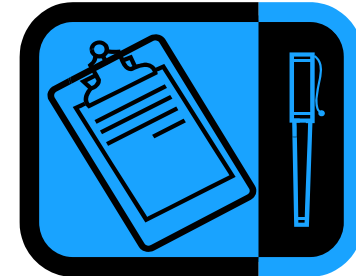
- Quando se usa um sistema automático com as suas próprias decisões, é preciso conhecê-lo...

O PALAVRAS

- Desenvolvido por Eckhard Bick desde 1994
- O primeiro analisador sintáctico para o português disponível na rede e usável por terceiros
- Um sistema dependencial baseado em restrições (constraint grammar dependency parsing)
- Uma teoria subjacente sobre o português
- Atribui sempre um par **forma:função**
- Usado pela Liguatca desde 1999-2000
 - Nova atomização
 - Muitas melhorias sugeridas através da colaboração
 - Como todos os sistemas automáticos, não é perfeito...

Anotando os corpora

Cada corpus é anotado sintacticamente.



PALAVRAS

```

$START
Cada      [cada] <quant> DET M S @>N
corpus    [corpus] N M S @SUBJ>
é         [ser] <fmc> V PR 3S IND VFIN @FAUX
anotado   [anotar] V PCP M S @IMV @#ICL-AUX<
sintacticamente ALT sintaticamente [sintático] ADV @<ADVL
$.
    
```

Formato AC/DC

Cada	cada	DET_quant	0	S	M	>N	0
corpus	corpus	N	0	S	M	SUBJ>	0
é	ser	V_fmc	PR_IND	3S	0	FAUX	0
anotado	anotar	V	PCP	S	M	IMV_#ICL-AUX<	0
sintacticamente	sintático	ADV	0	0	0	0	0

Um exemplo pouco animador

CF185-7 Ele sequestrou e violentou três meninos com a intenção de lhes transmitir o vírus da Aids de que se sabia portador.

com	[com] PRP @<ADVL
a	[o] <artd> DET F S @>N
intenção	[intenção] N F S @P<
de	[de] PRP @N<
lhes	[eles] PERS M/F 3P DAT @DAT>
transmitir	[transmitir] V INF @IMV @#ICL-P<
o	[o] <artd> DET M S @>N
vírus	[vírus] N M S @<ACC
de	[de] <sam-> PRP @N<
a	[o] <artd> <-sam> DET F S @>N
Aids	[aids] PROP F S @P<
de	[de] PRP @ADVL>
que	[que] KS @SUB @#FS-<ACC @#FS-<ACC
se	[se] PERS M/F 3S/P ACC @SUBJ>
sabia	[saber] V IMPF 3S IND VFIN @FMV
portador	[portador] N M S @<ACC

O mesmo exemplo no formato da Floresta

CF185-7 Ele sequestrou e violentou três meninos com a intenção de lhes transmitir o vírus da Aids de que se sabia portador.

```
=====MV:v-inf('transmitir')      transmitir
=====ACC:np
=====>N:art('o' <artd> M S)      o
=====H:n('vírus' M S)   vírus
=====N<:pp
=====H:prp('de' <sam->)        de
=====P<:np
=====>N:art('o' <-sam> <artd> F S) a
=====H:n('aids' F S)   Aids
=====N<:fcl
=====OC:adjp-
=====N<:pp
=====H:prp('de')   de
=====P<:pron-indp('que' <rel> M S)   que
=====ACC:np
=====H:pron-pers('se' M 3S ACC) se
=====P:vp
=====MV:v-fin('saber' IMPF 3S IND)   sabia
=====OC:adjp
=====H:adj('portador' M S)   portador
```

Santos et al. (2007)

Formato AC/DC e COMPARA

word	lema	pos	T	P	G	func	...
com	com	PRP	0	0	0	<ADVL	0
a	o	DET_artd	0	S	F	>N	0
intenção	intenção	N	0	S	F	P<	0
de	de	PRP	0	0	0	N<	0
lhes	eles	PERS	DAT	3P	M/F	DAT>	0
transmitir	transmitir	V	INF	3	0	IMV_#ICL-P<	0
o	o	DET_artd	0	S	M	>N	0
vírus	vírus	N	0	S	M	<ACC	0
da	de+o	PRP+DET_artd	0	S	F	N<+>N	0
Aids	aids	PROP	0	S	F	P<	0
de	de	PRP	0	0	0	ADVL>	0
que	que	KS	0	0	0	SUB_#FS-<ACC #FS-<ACC	0
se	se	PERS	ACC	3S/P	M/F	SUBJ>	0

T temcagr: tempo, caso, grau **P** pessnum: pessoa, número **G** gen: género

Quantas questões esta frase ilustra?

Santos et al. (2007)

- orações relativas
 - quantos tipos?
- coordenação
 - partilha de argumentos: *três meninos* é objecto de *sequestrou*; *Ele* é sujeito de...
- ligação vaga (2 vezes)
 - *portador do vírus da Aids* ou *portador de Aids*
 - *sequestrou e violentou* ou só *violentou* (?) com a intenção de...
- constituintes descontínuos (ou dependências de longa distância)
- elipse? *saber-se portador*
- *Aids* é um nome próprio?

O quebra-cabeças é...

Santos et al. (2007)

- integrar todas as peças de forma consistente numa mesma árvore
- ao contrário da maior parte dos projectos e/ou estudos de corpora que só observam um fenómeno de cada vez
- cada frase é geralmente um exemplo de tantos fenómenos quantos sintagmas ou palavras (generalização apressada 😊)
- e ainda há o problema do léxico – o que são palavras ou locuções ou morfemas
 - ex-comandante da LUAR

Mas a classificação não é óbvia!

Santos et al. (2007)

- Em última análise, a forma como a distinção está codificada em qualquer floresta é arbitrária! (mas pode ser resolvida com um sistema de procura adequado)
- *icl*, *acl*, *fcl* existem, mas *rcl* ou *subcl* não existem
- O tempo está marcado no verbo, ou só/também na oração?
- O género está marcado no SN, ou só no seu núcleo, ou em cada palavra passível de ter género?
 - *um índio pele vermelha* : que género deve ser marcado em *vermelha?* e em *pele?* e em *pele vermelha?*
- 3/4 razões para ter um adjectivo como núcleo
 - elipse, propriedade, indeterminação: *jovens alemães*

E as necessidades do utilizador não são precisas!

Santos et al. (2007)

- sintagma nominal *é quando* sintagma nominal
- SN complexos, mas sem oração no meio
- verbos que aparecem após uma citação
- frases em que a ordem sujeito-verbo-objecto é quebrada
 - frases que têm os 3, e verbo sem auxiliares?
 - VSO, SOV, OSV, OVS, VOS
- encontre um SN com a maior quantidade possível de dependentes
 - pai [de família [de emigrantes [dos subúrbios [de Moscovo [de 1900]]]]]
 - cão [de caça] [de loiça] [da Bélgica] [do meu pai] [do tempo da Grande Guerra]
- orações em que o particípio não exerce uma função verbal

O CLAWS

- Desenvolvido por Garside (et al) para o BNC
- Usado pela Linguateca apenas há um ano
- Muito menos ambicioso do que o PALAVRAS, sem função
- Teoria sintagmática (phrase structure): *pos tagger*

This is quite amazing.

CLAWS

```
<w id="2.1" pos="DD1">This</w> <w id="2.2" pos="VBZ">is</w>  
<w id="2.3" pos="RG">quite</w> <w id="2.4" pos="JJ">amazing</w>  
<w id="2.5" pos=".">.</w>
```

Formato AC/DC

This	this	DD1	0	0	0	0	0
is	is	VBZ	0	0	0	0	0
quite	quite	RG	0	0	0	0	0
amazing	amazing	JJ	0	0	0	0	0
.	.	PU	0	0	0	0	0

A codificação da reordenação

Robin said sharply, 'Caro was Judy's mother. And my wife. Not Joe's.'

72. <s> Robin said sharply, 'Caro was Judy's mother.

73. <s> And my wife

74. <s> Not Joe's.'

*-- A Caro era mãe de Judy. E minha mulher. Não era nada ao Joe -- **disse Robin com brusquidão.***

72. <s1/2+1> -- A Caro era mãe de Judy.
<reord 3> -- disse Robin com brusquidão.
</reord>

73. <s> E minha mulher.

74. <s2> Não era nada ao Joe <place 3>

A codificação da reordenação

Robin said sharply, 'Caro was Judy's mother. And my wife. Not Joe's.'

72. `<s>` Robin said sharply, 'Caro was Judy's mother.

73. `<s>` And my wife

74. `<s>` Not Joe's.'

*-- A Caro era mãe de Judy. E minha mulher. Não era nada ao Joe -- **disse Robin com brusquidão.***

72. `<s1/2+1>` -- A Caro era mãe de Judy.

`<reord 3>` -- disse Robin com brusquidão.
`</reord>`

73. `<s>` E minha mulher.

74. `<s2>` Não era nada ao Joe `<place 3>`

Sintaxe da procura no IMS-CWB

- [atributo="valor"]
- [atributo!="valor"] } ou uma sequência destes testes de atributos ligada por &
- Uma sequência, que pode ser modificada por {min, max}, ou por * ou +, significando qualquer número incluindo zero ou não

[lema="comer"] [pos="DET.*"]* [pos="N.*" & func="<ACC"]

- Os valores são descritos por expressões regulares
 - . qualquer character
 - [a-d] qualquer dos caracteres a, b, c, ou d
 - [, : ? .] qualquer dos caracteres ,, :, ?, or .
 - [^axl9] todos os caracteres excepto a, x, l ou 9 } caracteres
 - + um ou mais
 - * zero ou mais
 - {2,7} pelo menos 2 e no máximo sete
- } modificadores

Mais informações sobre a sintaxe do IMS-CWB

- Listas de valores: separadas por |
 - (passagem|estrada|via|rua)
 - [lema="forma|maneira|modo"]
- Caracteres especiais: precedidos de \
 - [lema="casar\+.*"]
- Indicação de sobre que unidade deve ser calculada a distribuição: @
 - [lema="dar"] []* @[func="<ACC"]
 - “caixa” @[pos="ADJ.*"]
- Atalhos: [word="pena"] pode ser escrito “pena” (e, se sozinho, pena no COMPARA e no AC/DC)
- Restrição a um atributo estrutural ou a um número de unidades:
 - within ua, within 5

Exploração inicial: Algumas questões talvez surpreendentes

- Adjectivos que modificam a palavra *pedra*
- Lema de substantivos femininos “regulares” ou de advérbios
- Função de gerúndio
- Categoria gramatical de palavras em *-ado*
- Categoria gramatical de *amigo*

- Categoria gramatical das palavras em *ing*
- Categoria gramatical de *home*

- Qual o género morfológico do que é comparado em *mais X do que?*

A interacção (primeira volta)

- Escolha do que se quer procurar
- Escolha do que se quer receber
 - Concordância
 - Distribuição
 - Por palavra
 - Por lema
 - Por categoria gramatical
 - Por autor
 - etc

O estudo de *embora* (inspirado por *home*)

- Quais as categorias gramaticais relevantes?
- Quais os verbos associados?
- Problemas na anotação...

- E *fora*?

Curso avançado de estudos contrastivos usando o COMPARA como ferramenta: segunda parte

Diana Santos

Linguatca
www.linguatca.pt

Problemas no presente COMPARA (Nov 2008)

- Distribuição por data não funciona
- Nomes diferentes de **autor** / **author** para a mesma informação
- Não há maneira fácil de evitar a repetição de originais
 - [fonte!=".*T1"]
- Não calcula bem o tamanho do subcorpo em todos os casos
- Atributos deslocados (acontece...)
- Não apresenta “bem” distribuições com casos vagos
- Listas de frequências dos nomes próprios ainda não corrigidas
 - Número dos nomes próprios deve ser dividido pelo tamanho
 - Sr.=Ryder 284 -> 142
 - Os=Vizinhos=do=Lado 88 -> 22

Não apresenta “bem” distribuições com casos vagos

- Procurar casos vagos
 - Branco / negro
 - POS de *amigo*
 - Casos de “ADJ_N”...

Aqui a questão principal, além da consciencialização das opções de anotação, é o facto de que o número, por exemplo, de casos de *amigo* como N e de *amigo* como ADJ pode ou não incluir os casos vagos

- E, se incluir, a soma dos casos é mais do que o número de ocorrências de *amigo* no corpo

Exemplo complexo de inspecção do COMPARA

- O caso de *sit* e suas relações com formas de expressão em português
- Este exemplo é para não pensarem que os estudos contrastivos com o COMPARA se reduzem a uma ou duas consultas (formuladas por uma expressão complexa de pesquisa)

A relação entre *sit* – *sentar*

45 sentidos diferentes (relevantes para a tradução) de *sit* no COMPARA

- sit up, sit through, sit with, sat on, sat forward, sitting up, sit out
- sit in judgement, sit on the bench
- sit to
- there sit, still sit
- sit things out, who should sit where, sits on the skyline, sat deep in discussion, sat down to a meal, sitting round, sitting about all over...
- sat Xing, sat opposite, behind, next, alone, attentively, bent...

70 traduções diferentes no COMPARA

O que é que é traduzido por *sit*?

123 contextos ou expressões diferentes são traduzidos por *sit*

- ficar *ando
- mesa
- ficar
- ter à frente, ter diante
- de onde estava

muitos deles sem qualquer relação com uma posição sentada

- fiquei só de cueca, gostava de olhar, fiquei com eles à mostra, fodendo num motel, velava o corpo, ali o deixaram ficar, e é a calma personificada, as pessoas não trazem muita pressa, benigna e sossegadamente determinemos, deixa-te estar aí, folheávamos, a vossos pés...

Outras questões sem auxílio automático

- Quantas *correntes* há no COMPARA? Distribuição de sentidos ...
- Oposição privativa: palavras que significam algo e o seu oposto
 - gosto: bom gosto, mau gosto
 - sorte: boa sorte, má sorte
 - ficar: mudar ou permanecer
- Vagos quanto a de propósito ou não
 - envergonhado
 - enganado
- Vagos quanto a uma característica ou a um sentimento
 - apaixonado
 - engraçado

Forma, maneira, modo e caminho

No DiaCLAV (um corpo do AC/DC)

forma	4741	4684	de	735	a	724	para	21
modo	1067	1067	de	69	entre	4	para	3
caminho	728	726	de	232	para	68	a	30
maneira	573	557	de	129	a	31	para	6

- `[word="forma" & pos="N.*"] @[pos="PRP.*"]`
- `[word="modo" & pos="N.*"] @[pos="PRP.*"]` etc.

Duas maneiras diferentes de obter uma tabela grande

`[word="forma|modo|caminho|maneira" & pos="N.*"] [pos="de"]` etc.

Exemplos de procuras “perigosas”

■ Distribuição em português original e português traduzido

- Verbos no gerúndio
- a / em

■ Distribuição por variante

- Verbos no gerúndio
- a / em

■ Distribuição por variante de português

- rapariga
- moça

Há mais texto traduzido em português de Portugal no COMPARA...

Variante e/ou data? E se estamos a misturar problemas de convergência ou divergência na nossa análise?

A interacção (segunda volta)

- Escolha do que se quer procurar
- Escolha do que se quer receber
 - Concordância
 - Distribuição
 - Por palavra
 - Por lema
 - Por categoria gramatical
 - Por autor
 - Etc
 - **Distribuição combinada** das expressões de busca em português e inglês

Restrições de alinhamento

- Procurar nos dois lados do COMPARA ao mesmo tempo...
- Tecnicamente, procura-se num lado, e os resultados são restritos aos casos cujo alinhamento satisfaça a restrição de alinhamento
- Restrições de alinhamento negativas: precedidas de !
 - mais alinhado com !more
- Comportamento do COMPARA quando se pede distribuição combinada:
 - Além de calcular as uas que a satisfazem, conta quantos casos da expressão individual existem nos dois lados do corpo, e apresenta os 3 valores (ou só 2 se a restrição for negativa)

Procuras complexas envolvendo o contraste

- Dois adjetivos modificando um substantivo – traduzidos por substantivo seguido de uma coordenação de adjetivos
- Dois adjetivos modificando um substantivo – traduzidos por adjetivo substantivo adjetivo
- Outras estratégias?

- Diminutivos em português ([temcagr=“DIM”])– como é que são traduzidos em inglês quando não envolvem a palavra *little*?
- *mais ADJ do que* – quando é traduzido por um adjetivo em *-er* ?
- Adjectivos vs. PP (*wooden* – *de madeira*)
[pos=“J.*” & word=“.*en”] ---- “de” [pos=“N.*”]

(Um lado dos) estudos contrastivos

Mas nem sempre é possível encontrar directamente expressões paralelas

- *-ing* deverbal nouns (*the moving, the establishing, the grouping, the computing...*)
- *viajado, passado* como adjectivos
- infinitivo pessoal vs. impessoal: como se traduz?

A interacção (terceira volta)

- Escolha do que se quer procurar
- Escolha do que se quer receber
 - Concordância
 - Distribuição
 - Por palavra
 - Por lema
 - Por categoria gramatical
 - Por autor
 - Etc
 - Distribuição combinada das expressões de busca em português e inglês
- **Seleccção de partes do corpo**

De notar que esta última reduz sempre significativamente o material

Mensagem final

- O mais importante é a informação que se pode obter, e desenvolver a metodologia de como obtê-la
- É importante perceber as potencialidades (e as limitações) de cada corpo, e idealmente usar vários
- **Mas:** não fiquem ofuscados!
- É importante compreender que um corpo ajuda-nos a fazer linguística
- Mas o mais importante é
 - Explorar
 - Descobrir
 - Validar

Mais informação

- Sítio do COMPARA: <http://www.linguateca.pt/COMPARA>
- Página das publicações e apresentações associadas ao COMPARA
- Sítio do AC/DC: <http://www.linguateca.pt/ACDC/>

- Página das publicações da Linguateca (produção escrita)

- Um texto/apostila com exercícios seleccionados relacionados com o presente curso

Obrigada pela atenção 😊