



# New Challenges in Corpora Studies – or is it the other way around: Corpora Studies as a new challenge for language professionals?

Stella E. O. Tagnin  
ABRAPUI – ENPULI  
June 2007




# Outline

- ◆ What is **Corpus Linguistics**: a discipline, an approach, a methodology? Tipos de corpora
- ◆ Tools
- ◆ Uses in teaching and in translation

# Is Corpus Linguistics a discipline?



- ◆ Not as we understand Linguistics, Sociolinguistics, Semantics etc.
- ◆ It does not have a specific object
- ◆ It is applicable to other disciplines: syntax as seen by CL, literature studied with CL tools, terminology based on CL etc.



# Is Corpus Linguistics an approach – a methodology?

- ◆ It is not just instrumental, i.e. a set of tools
- ◆ It is a typical way of applying a set of theoretical presuppositions (Berber Sardinha 2000)
- ◆ **Approach:** new theoretical concepts
  - Lexicon: collocations – patterns; “idiom principle”
  - Discovery: Johns (1990, 1991), Sinclair (2004): “trust the text” – no a priori categorization



## LC

- ◆ **Not** just a methodology as it may produce new knowledge
- ◆ Also, it may question established presuppositions



# Basically

**Corpus Linguistics =  
Observation of natural  
language**

**- use -**

as opposed to Chomky's  
introspection

**- grammar -**



# What is the difference

## Firth/ Halliday

- ◆ Performance
- ◆ Linguistic description
- ◆ Empirical
- ◆ Observation - use
- ◆ **Probability**

## Chomsky

- ◆ Competence
- ◆ Linguistic universals
- ◆ Rational
- ◆ Intuition - introspection
- ◆ **Possibility**



# What is a corpus?

1 *formal* a collection of writing, for example all the writings of one person

2 *linguistics* a collection of written and spoken language stored on computer and used for language research and writing dictionaries.

Macmillan English Dictionary (2002)





# Furthermore

- texts compiled according to specific criteria
  -
- collection should be representative of
  -
- language to be studied or
  -
- part of language to be studied

# Corpus linguist vs armchair linguist



- “... I don’t think there can be any corpora, however large, that contain information about all of the areas of English lexicon and grammar that I want to explore...”



# Corpus linguist vs armchair linguist

“... Every corpus [...] however small, has taught me facts that I couldn't imagine finding out about in any other way.”

(Fillmore, C.J. “Corpus linguistics” or “Computer-aided armchair linguistics”. In Svartvik, Jan (ed.) Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991, Berlin/NY: Mouton de Gruyter, 1992, 35-60)



# Types of corpora

- ◆ General language, specialized language, learner
- ◆ Monolingual, bi- ou multilingual
- ◆ Parallel (COMPARA) ou comparable (CorTec)



# Types of corpora

- ◆ online (BNC, Cobuild, WebCorp, Lácio-Web, Comet)
- ◆ DIY – customized: always at hand



# Research Tools

- WordList
- Key Words
- Concordance lines (KWIC)



# Wordlist

- ◆ Identification of topic
- ◆ Most reccurent words in text / corpus

WordList - [new wordlist (F)]

File Settings Comparison Index Window Help

WordList by Frequency

N	Word	Freq.	%	Lemmas
1	THE	7.987	6,68	
2	A	3.086	2,58	
3	IN	3.035	2,54	
4	AND	2.824	2,36	
5	TO	2.690	2,25	
6	OF	2.482	2,07	
7	FOR	1.112	0,93	
8	ON	1.019	0,85	
9	THAT	971	0,81	
10	WITH	934	0,78	
11	WAS	892	0,75	
12	WORLD	873	0,73	
13	IT	804	0,67	
14	HE	778	0,65	
15	CUP	745	0,62	
16	IS	726	0,61	
17	HIS	712	0,60	
18	BUT	682	0,57	
19	SAID	670	0,56	
20	BY	630	0,53	
21	AS	623	0,52	
22	HAVE	597	0,50	
23	AT	567	0,47	
24	TEAM	566	0,47	
25	WE	566	0,47	
26	FROM	537	0,45	

Windows taskbar: Iniciar, S., M., C. C., W, W, 17:03



WordList - [wcing.lst wordlist (A)]

File Settings Comparison Index Window Help

N	Word	Freq.	%	Lemmas
4074	GLOW	1		
4075	GLUE	1		
4076	GLUED	4		
4077	GO	102	0,08	
4078	GOAL	352	0,26	
4079	GOALIE	7		
4080	GOALIES	1		
4081	GOALKEEPER	133	0,10	
4082	GOALKEEPER'S	3		
4083	GOALKEEPERS	4		
4084	ALKEEPERSPOR+	1		
4085	GOALKEEPING	2		
4086	GOALLESS	1		
4087	GOALMOUTH	2		
4088	GOALPOST	4		
4089	GOALS	181	0,14	
4090	GOALSCORERS	1		
4091	GOALSCORING	1		
4092	GOALTENDER	1		
4093	GOALTIROS	2		
4094	GOAT	1		
4095	GOATEE	1		
4096	GOD	6		
4097	GODDESS	1		
4098	GODS	1		

Windows Taskbar: Iniciar, Internet Explorer, Mail, Word, Photos, Music, Windows Explorer, Google Chrome, A-Z, Mail, Calendar, C, M, G, 5, PT, 13:52

WordList - [wcing.lst wordlist (A)]

File Settings Comparison Index Window Help

N	Word	Freq.	%	Lemmas
4074	GLOW	1		
4075	GLUE	1		
4076	GLUED	4		
4077	GO	102	0,08	
4078	GOAL	533	0,40	goals(181)
4079	GOALIE	7		
4080	GOALIES	1		
4081	GOALKEEPER	133	0,10	
4082	GOALKEEPER'S	3		
4083	GOALKEEPERS	4		
4084	ALKEEPERSPOR+	1		
4085	GOALKEEPING	2		
4086	GOALLESS	1		
4087	GOALMOUTH	2		
4088	GOALPOST	4		
4089	GOALS	0		
4090	GOALSCORERS	1		
4091	GOALSCORING	1		
4092	GOALTENDER	1		
4093	GOALTIROS	2		
4094	GOAT	1		
4095	GOATEE	1		
4096	GOD	6		
4097	GODDESS	1		
4098	GODS	1		

Windows Taskbar: Iniciar, Internet Explorer, Mail, Word, Firefox, Chrome, A-Z, Outlook, Word, G, 5, PT, 13:54



# WordList - Statistics

Identification of peculiarities

- ◆ corpus (Overall)
- ◆ each text in the corpus

N	1	2	3	4	5	6	
Text File	<b>OVERALL</b>	SP2706_3.TXT	SP2706_2.TXT	SP2706_1.TXT	SP2606_1.TXT	SP2506_2.TXT	SP2506_1.TXT
Bytes	451,027	1,947	3,097	2,510	5,105	2,570	
Tokens	69,418	328	531	417	869	425	
Types	7,375	199	259	240	409	241	
Type/Token Ratio	10,62	60,67	48,78	57,55	47,07	56,71	
Standardised Type/Token	43,40						
Ave. Word Length	4,70	4,66	4,59	4,80	4,67	4,79	
Sentences	2,001	20	39	19	46	26	
Sent. length	18,94	14,30	12,64	21,42	18,89	15,04	
sd. Sent. Length	17,88	9,43	5,92	16,98	14,55	8,03	
Paragraphs	1,714	2	2	1	2	3	
Para. length	36,18	164,00	265,50	417,00	434,50	141,67	
sd. Para. length	58,95	166,88	316,08		198,70	101,90	
Headings	0	0	0	0	0	0	
Heading length							
sd. Heading length							
1-letter words	8,164	32	74	46	106	42	
2-letter words	11,014	59	77	59	121	67	
3-letter words	7,970	41	65	57	121	67	
4-letter words	7,836	40	60	38	87	48	
5-letter words	8,217	35	71	64	121	45	
6-letter words	6,862	40	65	38	104	41	
7-letter words	7,354	31	34	41	69	38	
8-letter words	4,861	21	31	27	55	26	



# Comparison of 2 WordLists

- ◆ Words which have **positive keyness** (vocabulary of text)
- ◆ Words which have **negative keyness** (vocabulary which does NOT occur in text)

WordList - [key words (keyness)]

File Settings Comparison Index Window Help

N	WORD	FREQ.	PTG.LST %	FREQ.	CA~1.TXT %	KEYNESS	P
785	MUDANÇAS	6		0		24,4	0,000001
786	MÜLLER	6		0		24,4	0,000001
787	SCALONI	6		0		24,4	0,000001
788	TÍTULOS	6		0		24,4	0,000001
789	ÉMERSON	6		0		24,4	0,000001
790	CHAVE	11	0,02	6		24,3	0,000001
791	EXPLICOU	11	0,02	6		24,3	0,000001
792	ITALIANA	11	0,02	6		24,3	0,000001
793	RESULTADO	28	0,04	53	0,01	24,2	0,000001
794	ARBITRAGEM	13	0,02	10		24,1	0,000001
795	ASS	0		86	0,02	24,2	0,000001
796	SES	0		86	0,02	24,2	0,000001
797	FOLLADAS	0		86	0,02	24,2	0,000001
798	CLIP	0		86	0,02	24,2	0,000001
799	IL	1		113	0,02	24,4	0,000001
800	BEEN	0		87	0,02	24,5	0,000001
801	ONLY	0		87	0,02	24,5	0,000001
802	ELECTIONS	0		87	0,02	24,5	0,000001
803	THAN	0		87	0,02	24,5	0,000001
804	GÃ	0		87	0,02	24,5	0,000001
805	CHE	0		88	0,02	24,7	0,000001
806	LANÃ	0		88	0,02	24,7	0,000001
807	VIDA	11	0,02	256	0,06	25,0	0,000001
808	CIOS	0		89	0,02	25,0	0,000001
809	READ	0		89	0,02	25,0	0,000001

Iniciar

# Key- Words

**Study corpus**

**VS**

**Reference corpus**





# Identification of prevailing vocabulary

- ◆ By frequency



KeyWords - [key words (keyness)]

File Settings Window Help

N	WORD	FREQ.	PTG. LST %	FREQ.	CA~1.TXT %	KEYNESS	P
1	NAO	491	0,71	0		1.998,2	0,000000
2	COPA	523	0,75	26		1.926,5	0,000000
3	É	447	0,64	8		1.740,6	0,000000
4	SELEÇÃO	304	0,44	0		1.236,4	0,000000
5	FOLHA	324	0,47	55	0,01	1.019,3	0,000000
6	ALEMANHA	267	0,38	13		984,3	0,000000
7	O	2.351	3,39	7.144	1,55	953,4	0,000000
8	JOGO	250	0,36	49	0,01	763,6	0,000000
9	JÁ	169	0,24	0		687,1	0,000000
10	TÉCNICO	165	0,24	0		670,8	0,000000
11	NA	770	1,11	1.519	0,33	635,9	0,000000
12	GOLS	159	0,23	5		603,1	0,000000
13	E	1.791	2,58	5.899	1,28	598,4	0,000000
14	EQUIPE	171	0,25	22		564,4	0,000000
15	GOL	187	0,27	39		563,3	0,000000
16	PARTIDA	147	0,21	8		536,8	0,000000
17	DA	1.163	1,68	3.336	0,73	526,7	0,000000
18	TAMBÉM	129	0,19	0		524,4	0,000000
19	FINAL	209	0,30	94	0,02	500,9	0,000000
20	ITÁLIA	123	0,18	0		500,0	0,000000
21	DO	1.346	1,94	4.235	0,92	499,8	0,000000
22	PÁGINA	122	0,18	0		495,9	0,000000
23	CROÁCIA	120	0,17	0		487,8	0,000000
24	UOL	141	0,20	13		487,7	0,000000
25	JOGADORES	132	0,19	11		462,1	0,000000

Iniciar

S. w C U 2 PT 14:30



# Concordance lines

**KWIC**

**Key Word In Context**



# Identification of patterns

- lexical patterns - collocations
- syntactic patterns - colligations



# What is a concordance line?

- ◆ **KWIC:** “key word in context” –
- ◆ **Vertical reading:** identification of grammatical and lexical patterns
- ◆ **Horizontal reading:** identification of collocates, clusters and different senses of word

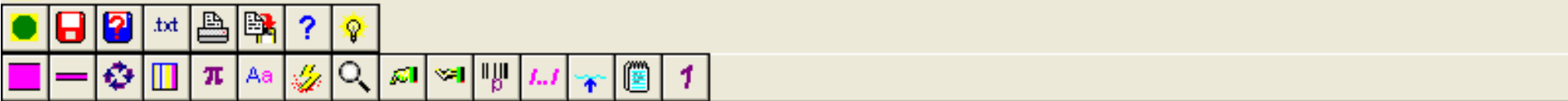


N	Concordance	Set	Tag	Word No.	File	%
1	Austrália nas oitavas por 1 a 0 com um gol marcado nos acréscimos, fruto de u			519	_27~3.txt	71
2	éxico por 2 a 1, neste sábado, com um gol de Maxi Rodríguez na prorrogação,			43	406~2.txt	7
3	ma, Cech passou a ser mais exigido no gol nos minutos iniciais do segundo te			360	_22~2.txt	75
4	inicial, apenas a Austrália jogou após o gol croata. Todos os ataques dos socc			418	_22~4.txt	84
5	de gol. A Argentina conseguiu marcar o gol nos acréscimos, com Messi, mas o			447	_24~1.txt	83
6	bolas cruzadas que a Croácia fez o seu gol, numa bola levantada para a ár			285	806_1.txt	49
7	sueco Allback teve a honra de marcar o gol de número 2000 da história das Cop			324	_23~1.txt	80
8	final. A seleção foi a única a não tomar gol em toda primeira fase do Mundial.			356	306~2.txt	67
9	GE.Net – A defesa cochilou no segundo gol argentino? Zinha – Foi um descuido			342	_25~4.txt	63
10	ce confuso, a Austrália chegou perto do gol. Atento em uma bola espirrada, Chi			231	_26~1.txt	50
11	ício contra a Ucrânia. "Espero que esse gol tenha dado impulso a ele", disse Ma			559	_29~2.txt	93
12	lha.com.br/poll/0616401/ Autor do gol da vitória, Kaká diz que já esperava			1.060	306_1.txt	90
13	goleiro Abbondanzieri, que saiu bem do gol, fechando o ângulo. Mas logo os			277	406~2.txt	44
14	rgetti acabou cabeceando para o próprio gol. O juiz suíço Massimo Busacca, no			138	406~2.txt	21
15	algumas oportunidades para marcar um gol, principalmente no primeiro tempo,			312	306~3.txt	89
16	sul-americanos definiram o placar com gol de Maxi Rodriguez na prorrogação.			369	_27~3.txt	50
17	as as equipes continuaram buscando o gol, mas a forte marcação dos mexican			174	406~2.txt	27
18	foi o atacante Marcus Allback, autor do gol número 2.000 na história das Copas			755	106_1.txt	95
19	tudo para o ataque e quase chegou ao gol de empate aos quatro minutos. Dep			348	506~1.txt	48
20	riativas e atã arriscou bons chutes a gol. Mas foi Toni, apesar de não ter			541	w12_1.txt	59
21	ue pode até fazer mais de um gol no Brasil. FOLHA - É o s			617	806_4.txt	82
22	ria contra a Austrália e autor do primeiro gol no triunfo contra os alemães, pelas			281	707~2.txt	71
23	de sair do banco de reservas, marcou o gol da vitória contra a Austrália. Ele pod			538	_29~2.txt	90
24	Time europeu bate a Tunísia com um gol de pênalti e avança na Copa LAN			18	306~1.txt	6
25	ildunterschrift: laquinta vibra ao marcar gol "Azzurra" deixa crise de lado contr			272	w12_1.txt	33



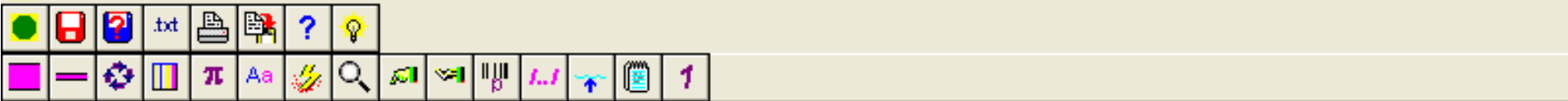
Everton's qualification for the European **Cup** Winners' Cup means Royle is seeking an English which gave the Merseysiders a 3-2 win in this **Cup** Winners' Cup clash in Reykjavik. [p] But if Lomu double; [h] [p] **WORLD CUP** sensation Jonah Lomu has won the New Zealand The America's **Cup** was successfully defended by the U.S. yacht right?" sat down and poured herself a **cup** of coffee. `Very sweet," she said. England 19 years later to regain the **cup**, we returned the favor and made it begin to prepare the rest of your meal, pour 1 **cup** of rice and 2 She sat me down loads of times with a **cup** of tea and that but she said that I could





N	Concordance	Set	Tag
42	jogo Os croatas começaram na frente, com um gol contra de Pablo Ibañez, logo aos 14min do p		
43	te Mundial, a Azzurra só só foi vazada uma única vez: um gol contra do lateral Zaccardo, no empate diante dos EUA		
44	francês Franck Ribery, o argentino Maxi Rodríguez - cujo gol contra o México foi o mais bonito da Copa, segundo p		
45	dois gols e ainda deu o chute que foi desviado por Petit no gol contra. - É ótimo marcar dois gols para a seleção. F		
46	jogo? Zinha - Todos deram o máximo, mas faltou fazer o gol. Criamos chances, mas não sabemos fazer o gol.		
47	13 - Jogador com maior número de cobranças de falta a gol: Cristiano Ronaldo (POR) e Totti (ITA) - 5 - Equipe c		
48	e aos 17min, a seleção voltou a ameaçar concretamente o gol croata, quando, depois de um bom cruzame		
49	ários e bateu de direita. O goleiro Kalac falhou e aceitou o gol croata. Depois de tomar o segundo gol, a Austrália e		
50	im como na etapa inicial, apenas a Austrália jogou após o gol croata. Todos os ataques dos socceros passavam pe		
51	uartas-de-final ficou nos pés de Maxi Rodríguez, que fez o gol da Argentina aos sete minutos da prorrogação. O meia		
52	m, o grandalhão Crouch não precisou entrar para marcar o gol da Inglaterra. Na segunda etapa, aos 14 minutos, Davi		
53	Totti, que, depois de sair do banco de reservas, marcou o gol da vitória contra a Austrália. Ele pode até voltar a jogar		
54	eia, chamado de "estrela em ascensão", que conseguiu o gol da vitória brasileira usando um pouco da "ma		
55	rdo... .. superou o goleiro Pletikosa... .. e comemorou o gol da vitória TUDO SOBRE A PARTIDA Eleito melhor e		
56	página 13/06/2006 - 18h11 Autor do gol da vitória, Kaká diz que já esperava dificuldade na		
57	ol.com.br/folha/esporte/ult92u103430.shtml Autor do gol da vitória, Kaká diz que já esperava dificuldade na estr		
58	, talvez porque beneficie o estilo de jogo deles. Autor do gol da vitória, o zagueiro Juanito disse que não é todo dia		
59	recorde de vitórias seguidas Autor do gol da vitória, Kaká diz que já esperava dificuldade na		
60	r maior confusão. Leia mais Autor do gol da vitória, Kaká diz que já esperava dificuldade na		
61	lha Online - Esporte - Autor do Online - Esporte - Autor do gol da vitória, Kaká diz que já esperava dificuldade na estr		
62	oas chances e de uma delas acabou surgindo o gol da vitória, com Fernando Torres, que já havia		
63	http://polls.folha.com.br/poll/0616401/ Autor do gol da vitória, Kaká diz que já esperava dificuldade na estr		
64	ação para a segunda fase. Leia mais Autor do gol da vitória, Kaká diz que já esperava dificuldade na estr		
65	o de Togo chegou apenas duas vezes com perigo perto do gol de Barthez durante os 45 minutos iniciais, a primeira c		

re-sort the data



N	Concordance	Set	Tag
162	nfatizar o resultado. "Comecei bem, fazendo um gol. Mas o que realmente importa é que a seleç		
163	u que ainda pode melhorar. "Eu comecei bem, fazendo um gol, participando das jogadas da seleção. Acho que fiz a		
164	ia anterior, ainda na Inglaterra, Rooney chegou a fazer um gol de voleio num treinamento. A foto do belo chute do ata		
165	território germânico sem ter tido o gostinho de ter feito um gol sequer. Dentre as defesas, o destaque ficou por con		
166	Virar um jogo assim tem mérito. Nos fizeram um gol e nos pressionaram, mas mesmo assim fomos além.		
167	utelosos. Pelo menos até que Maxí Rodríguez fizesse um gol inacreditável, aos oito minutos do primeiro tempo. Sorí		
168	burgo. Já a Suíça se despede da Copa sem ter levado um gol sequer. Os suíços começaram pressionando e tinha		
169	s seleções parecendo mais preocupadas em não levar um gol. A Espanha voltou a assustar aos 33 minutos, quando		
170	essão nos últimos minutos, mas acabou levando mais um gol nos acréscimos. Zidane recebeu livre na esquerda, de		
171	ar a torcida. Com este retrato do jogo, aconteceu mais um gol. Gilberto recebeu grande lançamento de Ronaldinho e		
172	e Reyes, que teve algumas oportunidades para marcar um gol, principalmente no primeiro tempo, deixou o campo de		
173	r em campo para ganhar ritmo de jogo e, claro, marcar um gol. Quando viu a placa com seu número, fez sinal de que		
174	campista do Milan, o "fantasista" Kaká, que marcou um "gol de anjo" no primeiro tempo. O diá		
175	spanhola às vésperas do início do Mundial. "Um gol do estreante Mariano Pernía, de falta, e uma extraordi		
176	Suécia sofreu, em 12 minutos, contra a Alemanha. Ou um gol, como a Argentina sofreu do México. Com um bom in		
177	A França chegando, sufocando, e Trezeguet perdendo um gol atrás do outro. Ribery, com "ciúmes", repetiu o lance		
178	s 15 do segundo tempo e sete minutos depois perdeu um gol incrível, após falha da zaga tcheca. A respota veio mai		
179	o pelo resultado do jogo da República Checa - que tem um gol a mais de saldo e jogará mais cedo contra Gana. Ma		
180	inutos da etapa inicial. Após o susto de ter tomado um gol logo no começo da partida, a Austrália partiu para o at		
181	te Mundial, a Azzurra só só foi vazada uma única vez: um gol contra do lateral Zaccardo, no empate diante dos EUA		
182	os, sem nem sequer termos conseguido marcar um único gol. Agora é o momento de darmos a volta por cima e mo		
183	er 3 a 0 e eliminar a Suíça do Mundial sem levar um único gol com bola gol com bola rolando.		
184	das Copas do Mundo a ser eliminada sem levar um único gol na competição. A "iniciativa" do jogo ficou por conta		
185	/06/2006 18:40:24) Suíça deixa Copa sem levar um único gol: 3 a 0 nos pênaltis Colônia (Alemanha) - O duelo e		

re-sort the data





# Compilação de um Glossário

## Seleção dos Termos

- ◆ **Keywords**
- ◆ **Clusters** - unidades multipalavras

Concord - [ (F) ]

File View Settings Window Help

N	cluster	Freq.
1	com um gol	13
2	gol da vitória	12
3	autor do gol	11
4	do gol da	8
5	o gol de	7
6	kaká diz que	7
7	um gol de	7
8	o gol da	6
9	perto do gol	6
10	chutes a gol	6
11	o primeiro gol	5
12	um único gol	5
13	ao gol de	5
14	chances de gol	5
15	marcou o gol	4
16	um gol #	4
17	o segundo gol	4
18	o colocado com	4
19	do gol de	4
20	gol de kaká	4
21	colocado com um	4
22	com gol de	4
23	perigo ao gol	3
24	para o gol	3
25	cima do gol	3

Iniciar

T... C... U... 4 N EN 15:28

Concord - [ (A) ]

File View Settings Window Help

N	cluster	Freq.
1	ao gol de	5
2	após o gol	3
3	autor do gol	11
4	chance de gol	3
5	chances de gol	5
6	chutes a gol	6
7	cima do gol	3
8	colocado com um	4
9	com gol de	4
10	com um gol	13
11	de chutes a	3
12	do gol da	8
13	do gol de	4
14	gol da vitória	12
15	gol de empate	3
16	gol de kaká	4
17	gol de maxi	3
18	kaká diz que	7
19	levar um único	3
20	marcar o gol	3
21	marcou o gol	4
22	o gol croata	3
23	o gol da	6
24	o gol de	7
25	o primeiro gol	5

Clusters de 3 palavras

Iniciar

T... C... U... 4 N EN 15:29



# Translational equivalence

- ◆ Functional equivalents

- marcar um gol

- gol contra



N	Concordance	Set	Tag	Word
24	ne, but locked into their defensive mindset, they get a goal and sit back. Then they lose on penalty kicks. Itâ			
25	they went down to 10 men we gave everything to get a goal and it worked," coach Juergen Klinsmann said. "			
26	, the bartender called for last drinks. "What if we get a goal?" one drinker shouted. Then the Aussies got thre			
27	, the bartender called for last drinks. "What if we get a goal?" one drinker shouted. Then the Aussies got thre			
28	d in the 84th and 89th minutes, and John Aloisi got a goal in injury time. Shunsuke Nakamura gave Japan a			
29	at the other end, substitute DaMarcus Beasley had a goal disallowed for offside after Gianluigi Buffon had al			
30	Sancho's dreadlocks to create the space to head in a goal in the 83rd minute. Sancho had a rubber band ar			
31	's beautiful. It's a heck of a challenge, but I think it's a goal we can make." U.S. soccer officials often point to			1.
32	he apparently grew frustrated after narrowly missing a goal on a header in overtime and at being roughed up			
33	guay that scored it. Trinidad and Tobago didn't need a goal to feel like champions, holding Sweden to a 0-0 dr			
34	st-scoring teams in European qualifying, couldn't net a goal through two losing matches against Ecuador and			1.
35	red almost the entire 90 minutes before finally netting a goal to beat Paraguay. And France, demonstrating wh			1.
36	game, they'd better realize this is a World Cup, not a goal of the month contest. © 2006 The Associated P			
37	ages still replayed on TV today, the aftershocks of a goal scored in the final by a 17-year-old named Pelé,			
38	undane shot into a adventure, nearly gifted Portugal a goal in the 78th minute. Cristiano Ronaldo's free kick w			
39	es matter is whether the Americans can finally score a goal or two on their own and make their case for a se			
40	yes wide open. Don't blame us that they didn't score a goal." Instead, blame the 37-year old Hislop, who face			
41	ut we wanted to win the game. We wanted to score a goal at the end of it all. That would have made the whol			
42	driano and Kaka are the others) who has yet to score a goal. His last international goal came a year ago, in t			
43	's only other trip was in 1974, when it failed to score a goal in losing to host West Germany and East Germ			
44	aymaker Juan Roman Riquelme, who failed to score a goal but masterfully distributed from midfield, was na			
45	02 when, as defending champions, they didn't score a goal and went home after the first round. If Switzerland			
46	's only other trip was in 1974, when it failed to score a goal in losing to host West Germany and East Germ			
47	ampionship in 1998 on home soil, then failed to score a goal or win a game in Korea/Japan in 2002. Reminders			



N	Concordance	Set	Tag	Word No.	File	%
1	anck Ribery, o argentino Maxi Rodríguez - cujo gol contra o México foi o mais bonito da Copa,			246	707~2.txt	63
2	a Angola. É muito perigoso e criou a jogada do gol contra Angola jogando atrás", lembrou. "É			605	_30~1.txt	71
3	ainda deu o chute que foi desviado por Petit no gol contra. - É ótimo marcar dois gols para a			72	907~1.txt	46
4	defesa, ainda teve fôlego para finalizar contra o gol de Van Der Sar. "Sinto confiança em ataca			179	706_3.txt	53
5	depois de sair do banco de reservas, marcou o gol da vitória contra a Austrália. Ele pode até v			538	_29~2.txt	90
6	a vitória contra a Austrália e autor do primeiro gol no triunfo contra os alemães, pelas semifin			281	707~2.txt	71
7	u, em 12 minutos, contra a Alemanha. Ou um gol, como a Argentina sofreu do México. Co			343	706_2.txt	65
8	Os croatas começaram na frente, com um gol contra de Pablo Ibañez, logo aos			173	806~3.txt	44
9	a Azzurra só só foi vazada uma única vez: um gol contra do lateral Zaccardo, no empate dian			240	_08~1.txt	38

Concord - [GOAL: 12 entries (sort: 1L,2R)]

File View Settings Window Help

Concordance

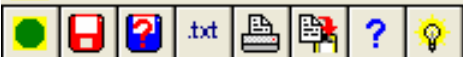
N	Set	Tag	Word No.
1			1.201 606
2			549 707
3			700 006
4			508 206
5			884 907
6			322 507
7			918 007
8			302 007
9			914 807
10			3.050 406
11			782 807
12			297 106

Windows taskbar: Iniciar, Internet Explorer, Outlook, Word, Firefox, Chrome, VLC, A-Z, Mail, Calendar, T..., C..., U..., 4 N, EN, 15:56



N	Concordance	Set	Tag	Word No.	File
1	e scored on," Buffon said. Italy defender Cristian Zaccardo scored an own-goal 27 minutes into Ital			151	907_4.tx
2	manageable for the Italians, but defender Cristian Zaccardo flailed at it with his left foot and spun t			715	806_1.tx
3	was canceled out by a an own goal from Cristian Zaccardo, but the real talking point of an explosiv			105	706_3.tx
4	ith the United States three weeks ago on Cristian Zaccardo's own goal. The Azzurri had been atte			536	007_1.tx
5	heir Group E clash but an own goal from Cristian Zaccardo put the Americans back on level terms i			138	706_2.tx
6	s bumped into the net by his teammate Cristian Zaccardo in the bruising June 17 tie with the Uni			386	907_3.tx
7	scored in the 27th when Italian defender Cristian Zaccardo knocked the ball into his own net. Itali			412	706_1.tx
8	y's cross was sliced into his own net by Cristian Zaccardo. Things got even worse for Italy a minut			414	706_3.tx
9	moment of panic by a reserve defender, Cristian Zaccardo, who put the ball into his own net. Fra			870	807_4.tx
10	te things in the middle. GK BUFFON Gianluigi D ZACCARDO Cristian D CANNAVARO Fabio D			1.691	206_2.tx





N	Concordance	Set	Tag	Word
208	e to end in a draw, with calculators needed to sort out goal differences and determine the two that go throug			
209	at two minutes for U.S. soccer there -- we had an own goal and a red card," said defender Jimmy Conrad, wh			
210	has given up one goal in six games, and it was an own goal. And the Italians have done it with a diverse offen			
211	EAL ESTATE England lead after Paraguay own goal By Martyn Herman Reuters Saturday, June 10			
212	eaching the interval 1-0 in front thanks to an early own goal by Carlos Gamarra. It took just over three minute			
213	ilardino's opener for Italy was canceled out by a an own goal from Cristian Zaccardo, but the real talking point			
214	lead with a header in their Group E clash but an own goal from Cristian Zaccardo put the Americans back on			
215	Y 0 Paraguay captain Carlos Gamarra scores an own goal to give England the lead as the ball glances off hi			
216	tar defender, Andres Escobar, scored an own goal to help the United States to a 2-1 upset			
217	contingent was overpowered from the onset. The own goal only made it worse, then in the sixth minute goalk			
218	gland lead after Paraguay own lead after Paraguay own goal NEWS   OPINIONS   SPORTS   ARTS & LIVING			
219	what we wanted." England had to rely on an early own goal by Paraguay captain Carlos Gamarra to squeak			
220	ad been repaired, after the three red cards and the own goal and the disallowed goal, after the breathtaking s			
221	d States three weeks ago on Cristian Zaccardo's own goal. The Azzurri had been attempting to tie a World C			
222	eam has yielded just one goal in six games -- an own goal against the United States in a first-round game in			
223	ut the United States tied it five minutes later on an own goal. The scoring ended, the drama did not. Italy's D			
224	the elimination of his team in the 1998 Cup by an own goal, was assassinated when he returned home. Mani			
225	opponent from the run of play. One score was an own-goal by an Italian defender, the second was a penalty k			
226	after losing its opener 1-0 to England on an early own-goal by captain Carlos Gamarra. The South American			
227	me. Paraguay had more to deal with than just the own-goal in Frankfurt. Besides a decidedly pro-England cr			
228	ago 0 KAISERSLAUTERN, Germany (AP) _ This own-goal went in Paraguay's favor and knocked Trinidad a			
229	n said. Italy defender Cristian Zaccardo scored an own-goal 27 minutes into Italy's 1-1 draw with the United S			
230	the 2-0 loss to Italy in the semifinals _ set up the own-goal as Petit deflected his free kick behind Ricardo. I			
231	the 2-0 loss to Italy in the semifinals, set up the own-goal as Petit deflected his free kick behind Ricardo. I			

re-sort the data



# Descobrimos novos termos

- ◆ Goal differential
- ◆ Injury time
- ◆ overtime



## In short

- ◆ A corpus will allow you to identify the

**most probable term**

**- not just a possible term!**



# Uses of a parallel corpus

## COMPARA

([www.linguateca.pt/COMPARA](http://www.linguateca.pt/COMPARA))

- ◆ Literary texts
- ◆ Originbals and translations
- ◆ Portuguese: Brazil, Portugal, Mozambique, Angola
- ◆ English: USA, UK, South Africa



# Translation strategies

Synonyms for often used words

The case of "then"

# Then → Então – Aí -- Depois



EBDL1T1(193): There was another long journey through more corridors -- **then** suddenly everything changed.

Seguiu-se outra longa viagem ao longo de mais corredores, **até que**, de repente, tudo mudou.

EBDL1T1(294): But **then** neither of them seemed to go with the shirt I was wearing, so I had to change that.

Mas **depois** achei que nem uma nem outra ficavam bem com a camisa que tinha vestida e tive de pôr uma terceira.

EBDL1T2(294): But **then** neither of them seemed to go with the shirt I was wearing, so I had to change that.

Mas **aí** nenhuma das duas combinava com a camisa que estava usando, por isso tive de trocá-la.

EBDL1T2(1159): I smiled at the thought, **then** felt a bit of a bastard for smiling. Sorri com esse pensamento, **depois** me senti um pouco canalha por estar rindo.

EBDL2(284): `Who was the clock radio for, **then**?´

-- **Então** para quem era o rádio-despertador?

EBDL2(297): `Perhaps one of the boys would like it, **then**.´

-- **Nesse caso**, talvez um dos rapazes o queira.



# Cultural words: *cachaça*

COMPARA - Resultados da pesquisa *cachaça* - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço <http://www.linguateca.pt/cgi-bin/paraalelo.pl>

Google Pesquisar na Web PageRank 80 bloqueado(s) Opções

Procura: "*cachaça*".

<a href="#">PBAA1</a> (181):	Afinal, já lhe não bastava sortir o seu estabelecimento nos armazéns fornecedores; começou a receber alguns gêneros diretamente da Europa: o vinho, por exemplo, que ele dantes comprava aos quintos nas casas de atacado, vinha-lhe agora de Portugal às pipas, e de cada uma fazia três com água e <b>cachaça</b> ; e despachava faturas de barris de manteiga, de caixas de conserva, caixões de fósforos, azeite, queijos, louça e muitas outras mercadorias.	After a while, he began to buy less from wholesalers and ordered some products directly from Europe -- wine, for example. Before, he had purchased it in demijohns, but now he bought barrels straight from Portugal. He turned each barrel into three by adding water and rum. Likewise, he ordered kegs of butter, crates of canned goods, big boxes of matches, oil, cheese, crockery, and much else besides.
<a href="#">PBAA2</a> (24):	O leiloeiro tinha piscos de olhos significativos; de martelo em punho, entusiasmado, o ar trágico, mostrava com o braço erguido um cálice de <b>cachaça</b> , ou, comicamente acororado esbrocava com o furador os paneiros de farinha e de milho.	The auctioneer winked his eyes meaningfully, excited, with the gavel always in one fist, his countenance serious, he would raise his arm to display a jug of crude brandy, or squat comically to poke at the baskets of flour and corn with a sharp-pointed rod.
<a href="#">PBAA2</a> (1389):	Adorava os perfumes ativos, as jóias e as cores vivas, para ele, nada havia, porém, como um passeio ao sítio embarcado, à fresca da madrugada, bebericando o seu trago de <b>cachaça</b> e pitando o seu fumo do Codó.	He adored strong colognes, jewels, and lively colors. But there was nothing, in his view, equal to a short trip out to his rural property. He would set off in the fresh early morning, swigging gulps of sugarcane brandy and smoking his Codó mixture.
<a href="#">PBAD2</a> (447):	Apesar da ordem para que as vendas se mantivessem fechadas, muita gente fizera de véspera a sua provisão de <b>cachaça</b> e patifaria, e cana corria alegre, bebida mesmo na boca da botija, no bafo e no arrote.	Despite an order that the stores should stay closed, many people had stocked up with spirits and baccy the day before, and the drink was running freely, drunk straight from the bottle, accompanied by burps and stinking breath.
<a href="#">PBAD2</a> (770):	Januário tirou a rolha de sabugo com os dentes, deixou a <b>cachaça</b> cair quente goela abaixo.	Januario pulled the corn-cob stopper out with his teeth and poured the spirit hot down his throat.
<a href="#">PBMRI</a> (1622):	Um ano depois da tragédia, eu substituía o uísque pelo finalô, imediatamente substituído pelo gim e mesmo pela <b>cachaça</b> com limão.	A year after the tragedy, I had replaced whiskey with brandy and water, which was immediately supplanted by gin, and then by sugarcane rum with lime.

Esperamos que o COMPARA lhe tenha sido útil!

[UTILIZAR O COMPARA](#) [MAIS INFORMAÇÕES SOBRE O CORPUS](#) [AGRADECIMENTOS](#) [COMO COLABORAR](#)

Iniciar Explora... Microsoft... Microsoft... COMP... Mapa d... Explora... 10:05

# Cachaça in the COMPARA

PBAA  
1(181)  
):

Afinal, já-lhe não bastava sortir o seu estabelecimento nos armazéns fornecedores; começou a receber alguns gêneros diretamente da Europa: o vinho, por exemplo, que ele dantes comprava aos quintos nas casas de atacado, vinha-lhe agora de Portugal às pipas, e de cada uma fazia três com água e **cacheça**; e despachava faturas de barris de manteiga, de caixas de conserva, caixões de fósforos, azeite, queijos, louça e muitas outras mercadorias.

After a while, he began to buy less from wholesalers and ordered some products directly from Europe -- wine, for example. Before, he had purchased it in demijohns, but now he bought barrels straight from Portugal. He turned each barrel into three by adding water and **rum**. Likewise, he ordered kegs of butter, crates of canned goods, big boxes of matches, oil, cheese, crockery, and much else besides.

PBAA  
2(24)  
:

O leiloeiro tinha piscos de olhos significativos; de martelo em punho, entusiasmado, o ar trágico, mostrava com o braço erguido um cálice de **cacheça**, ou, comicamente acorocado esbrocava com o furador os paneiros de farinha e de milho.

The auctioneer winked his eyes meaningfully; excited, with the gavel always in one fist, his countenance serious, he would raise his arm to display a jug of **crude brandy**, or squat comically to poke at the baskets of flour and corn with a sharp-pointed rod.

PBAA2(  
1389):

Adorava os perfumes ativos, as jóias e as cores vivas, para ele, nada havia, porém, como um passeio ao sitio embarcado, à fresca da madrugada, bebericando o seu trago de **cacheça** e pitando o seu fumo do Codó.

He adored strong colognes, jewels, and lively colors. But there was nothing, in his view, equal to a short trip out to his rural property. He would set off in the fresh early morning, swigging gulps of **sugarcane brandy** and smoking his Codó mixture.



# Cachaça in the COMPARA



PBAD2  
(447):

Apesar da ordem para que as vendas se mantivessem fechadas, muita gente fizera de véspera a sua provisão de **cacheça** e patifaria, e cana corria alegre, bebida mesmo na boca da botija, no bafo e no arrote.

Despite an order that the stores should stay closed, many people had stocked up with **spirits** and baccy the day before, and the drink was running freely, drunk straight from the bottle, accompanied by burps and stinking breath.

PBAD2  
(770):

Januário tirou a rolha de sabugo com os dentes, deixou a **cacheça** cair quente goela abaixo.

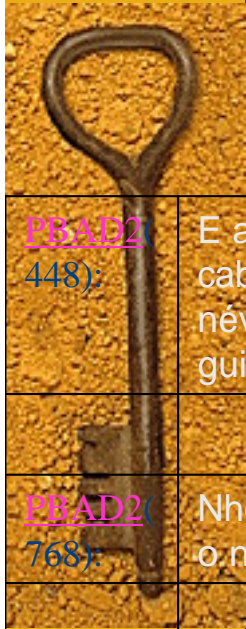
Januario pulled the corn-cob stopper out with his teeth and poured the **spirit** hot down his throat.

PBMR1  
(1622):

Um ano depois da tragédia, eu substituía o uísque pelo finalô, imediatamente substituído pelo gim e mesmo pela **cacheça** com limão.

A year after the tragedy, I had replaced whiskey with brandy and water, which was immediately supplanted by gin, and then by **sugarcane rum** with lime.

# Pinga in the COMPARA



<p>PBAD2(448):</p>	<p>E alguns mais altos, a <b>pinga</b> subida na cabeça, já riam e antegozavam, na névoa estúpida da bebida, o grã-guinhol, a fantástica ópera de títeres.</p>	<p>And those who were already high, the <b>drink</b> having gone to their heads, were already laughing in the stupefying cloud of drunkenness and looking forward to the Grand Guignol, the fantastic puppet theatre.</p>
<p>PBAD2(768):</p>	<p>Nhonhô quer um gole de <b>pinga</b>? disse o negro sentindo frio.</p>	<p>Do you want a drop of <b>rum</b>, massa? asked the Negro, feeling the cold.</p>
<p>PBAD2(769):</p>	<p>Como Januário custasse a responder, ele acrescentou é <b>pinga</b> da cabeça, Nhonhô, esquenta os peitos.</p>	<p>As Januario was slow to answer, he added it's a <b>drop</b> of the best, massa, warms the cockles of the heart.</p>
<p>PBAD2(1026):</p>	<p>A <b>pinga</b> desceu áspera arranhando a goela, foi queimar o estômago vazio.</p>	<p>The <b>spirit</b> was sharp going down, rasping his throat, burning his empty stomach.</p>
<p>PBPM2(407):</p>	<p>Quando abri a porta, encontrei um monte de pacotes na soleira: cigarros, carne moída, cerveja, <b>pinga</b> e flores.</p>	<p>When I opened the door I found a pile of packages on the threshold: cigarettes, ground beef, <b>rum</b> and flowers.</p>



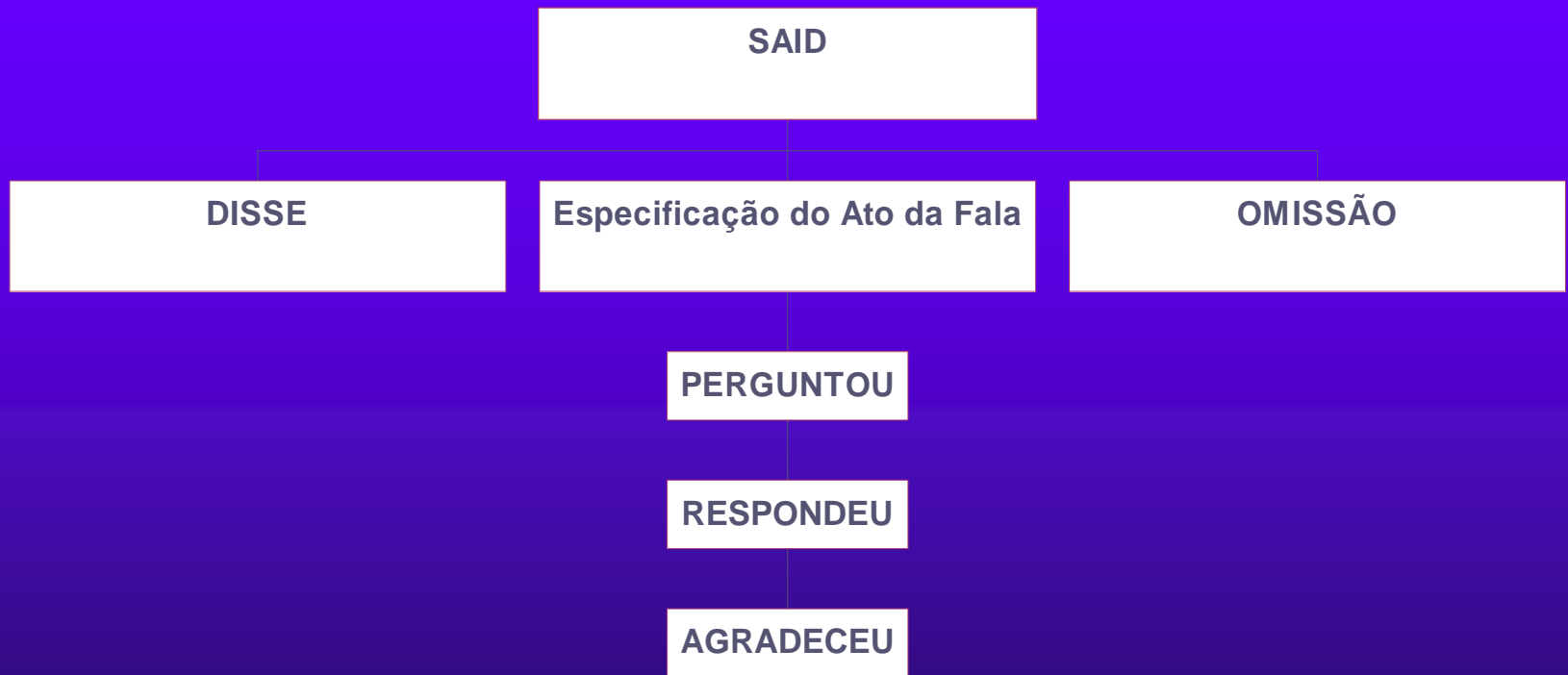
# Contrastive Analysis

Peculiarities of each  
language

# The case of “said”

<a href="#">EBDL1T2</a> (240):	«Well, frankly old bean, I'm baffled,» he <b>said</b> .	-- Bem, francamente, meu velho camarada, estou surpreso -- <b>disse</b> ele.
<a href="#">EBDL1T2</a> (247):	«Particularly for me,» I <b>said</b> .	-- Principalmente para mim.
<a href="#">EBDL1T2</a> (167):	«Got a smoke on you?» <b>said</b> the nurse.	-- Tem um cigarro aí? -- <b>perguntou</b> a enfermeira.
<a href="#">EBDL1T2</a> (98):	«Ward 3J,» I <b>said</b> .	«Ala 3J», respondi.
<a href="#">EBDL1T2</a> (319):	She <b>said</b> , «When I don't feel the need to see Karl any more.»	<b>Respondeu</b> , «Quando eu não sentir mais necessidade de ver Karl».
<a href="#">EBDL1T2</a> (357):	«Pathological?» she <b>said</b> .	-- Patológica? -- <b>perguntou</b> ela.
<a href="#">EBDL1T2</a> (914):	«But I am writing,» I <b>said</b> .	-- Mas estou escrevendo -- <b>retriquei</b> .
<a href="#">EBDL1T2</a> (1053):	«This doesn't affect Rhoda, believe me,» he <b>said</b> earnestly.	-- Isso não tem nada a ver com Rhoda, acredite em mim.
<a href="#">EBDL1T2</a> (1660):	I <b>said</b> I could.	<b>Respondi</b> que sim.
<a href="#">EBDL1T2</a> (1728):	«I wont stop, thanks,» she <b>said</b> .	Obrigada, mas não vou ficar -- <b>agradeceu</b> .

# Said vs. Disse



# Results



said IO – 310 = **0,4%** → disse PT – 203 =  
0,25%

disse PO – 936 = 0,23% → said IT – 772 =  
0,18%

told IT - 59 =

0,013%  
**0,193%**

Conclusion:

IT – 1106 said: 772 ← disse – 69,8%

334 ← others – 30,19%

1106/421.725 = **0,26%**



# DIY Corpora

- ◆ Construction depends on purpose of research
- ◆ Corpus tools: Wordsmith Tools (Scott 1996)





# Advantages of a DIY Corpus

- ◆ Up-to-date
- ◆ Reliable
- ◆ Wealth of information
  - lexical
  - syntactic
  - contextual
  - discursive/textual
- ◆ Become familiar with the topic
- ◆ Easy to consult





# Contrasting languages

A quadrilingual corpus:

- ◆ Brazilian Portuguese (PB) vs European Portuguese (PE)
- ◆ American English (IA) vs British English (IB)

# European English – positive

N	WORD	FREQ.	TOTPOP.LST %	FREQ.	TOTPOB.LST %	KEYNESS	P
1	SE	385	2,18	90	0,36	314,8	0,000000
2	LUME	82	0,46	0		144,1	0,000000
3	Q	77	0,44	0		135,3	0,000000
4	B	77	0,44	0		135,3	0,000000
5	PREPARAÇÃO	77	0,44	3	0,01	113,0	0,000000
6	COZER	59	0,33	0		103,7	0,000000
7	DL	57	0,32	0		100,2	0,000000
8	GR	89	0,50	13	0,05	92,6	0,000000
9	NATAS	49	0,28	0		86,1	0,000000
10	SUMO	39	0,22	0		68,5	0,000000
11	GRS	43	0,24	1		67,1	0,000000
12	CONFECÇÃO	37	0,21	0		65,0	0,000000
13	DURANTE	61	0,35	9	0,04	63,1	0,000000
14	S	29	0,16	0		50,9	0,000000
15	DEITE	27	0,15	0		47,4	0,000000
16	JUNTA	23	0,13	0		40,4	0,000000
17	TACHO	21	0,12	0		36,9	0,000000
18	MARGARINA	57	0,32	19	0,08	35,1	0,000000
19	ARREFECER	19	0,11	0		33,4	0,000000
20	PURÉ	19	0,11	0		33,4	0,000000
21	AS	246	1,39	202	0,81	32,5	0,000000
22	LHE	17	0,10	0		29,8	0,000000
23	DEITA	17	0,10	0		29,8	0,000000
24	BOCADOS	17	0,10	0		29,8	0,000000
25	COM	445	2,52	434	1,75	29,8	0,000000
26	COZA	16	0,09	0		28,1	0,000000
27	DEIXA	25	0,14	3	0,01	28,1	0,000000
28	C	40	0,23	12	0,05	27,0	0,000000
29	TAPE	15	0,08	0		26,3	0,000000
30	ALOURAR	14	0,08	0		24,6	0,000001

# Brazilian Portuguese - negative



N	WORD	FREQ.	TOTPOP.LST %	FREQ.	TOTPOB.LST %	KEYNESS	P
31	XÍCARAS	2	0,01	34	0,14	24,6	0,000001
32	GELADEIRA	0		24	0,10	25,8	0,000000
33	REFOGUE	2	0,01	36	0,15	26,6	0,000000
34	RESERVE	9	0,05	60	0,24	26,9	0,000000
35	OLIVA	0		30	0,12	32,3	0,000000
36	SALSINHA	0		30	0,12	32,3	0,000000
37	COZINHE	6	0,03	58	0,23	33,1	0,000000
38	PORÇÕES	1		38	0,15	33,3	0,000000
39	GOSTO	12	0,07	78	0,31	34,3	0,000000
40	PARA	89	0,50	252	1,02	35,8	0,000000
41	RENDIMENTO	0		35	0,14	37,6	0,000000
42	BAIXO	1		43	0,17	38,5	0,000000
43	ACRESCENTE	9	0,05	81	0,33	44,4	0,000000
44	DO	70	0,40	233	0,94	46,1	0,000000
45	PREPARO	9	0,05	87	0,35	49,7	0,000000
46	POR	77	0,44	262	1,06	53,9	0,000000
47	PANELA	15	0,08	115	0,46	57,1	0,000000
48	SUCO	0		59	0,24	63,5	0,000000
49	CHÁ	21	0,12	156	0,63	75,9	0,000000
50	REINO	0		73	0,29	78,6	0,000000
51	COLOQUE	23	0,13	183	0,74	93,3	0,000000
52	FOGO	6	0,03	135	0,54	106,3	0,000000
53	XÍCARA	0		119	0,48	128,2	0,000000



# Lexical Contrasts

## PB vs PE

Fogo	135	6
Lume	0	82
Fogo baixo	32	0
Fogo brando	8	0
Lume (mto) brando	0	17
Fogo alto	11	0
Lume forte	0	4
Fogo médio	6	0
Xic/ Xícara(s)	155	2(?)
C/ch/chávena	0	44



# Intercultural Equivalents

- ◆ Some words are typically Brazilian: *coco, bananas, farofa, mandioca, abacaxi;*

- *salsão,*
- *suco,*
- *xícara,*
- *salsinha,*
- *geladeira,*
- *coentro,*
- *filé,*
- *pimentão,*
- *pimenta-do-reino,*
- *dourar,*
- *[claras em] neve*
- *aipe* (also in BP)
- *sumo,*
- *chávena,*
- *salsa,*
- *frigorífico,*
- *coentros,*
- *filete,*
- *pimento,*
- *pimenta preta,*
- *alourar,*
- *[claras em] castelo.*



# Lexical Contrasts

## IA vs IB

◆ Broth	41		3
◆ Stock	19		55
◆ Skillet	38		4
◆ Frying pan	0		21
◆ Can	11		6
◆ Tin	0		7
◆ Tin (baking)	4		36
◆ Cornstarch	13		0
◆ Cornflour	0		21
◆ Gas (mark)		1	47
◆ Courgettes	0		23
◆ Zucchini	8		1

# *DUBLINERS* SOB A LUPA DA LINGÜÍSTICA DE CORPUS:

Uma contribuição para a análise e a  
avaliação da tradução literária

Lourdes Bernardes Gonçalves

Universidade Federal do Ceará (Depto Letras Estrangeiras)

**Orientadora:** Dra. Stella Esther Ortweiler Tagnin

2006





# Corpus-driven research in Literature: "A Little Cloud" and its translations

Tradutores: Hamilton Trevisan  
José Roberto O'Shea

Nº	Joyce	oc.	Trevisan	oc.	O'Shea	oc.
1	little	57	little	33	little	40
2	life	16	amigo	14	amigo	23
3	child	13	vida	14	vida	17
4	friend	13	criança	12	criança	16
5	head	10	uísque	10	olhos	13
6	time	10	noite	9	gente	10
7	eyes	8	cabeça	8	coisa	9
8	face	8	olhos	8	rosto	9
9	glass	8	vez	8	copo	8
10	man	7	coisa	7	vez	8

**Tabela 6.7:** Os 10 substantivos mais ocorrentes (texto original e traduções)





# Comparison - Keywords

"A Little Cloud" (**alitle**)

"Dubliners" (minus "A Little Cloud") (**dub-alc**):

N	WORD	FREQ.	ALITCL.TXT	%FREQ.	DUB-ALC.LST	%KEYNESS
1	GALLAHER	48	0,95	0		251,5
2	CHANDLER	41	0,81	0		214,8
3	IGNATIUS	38	0,75	0		199,0
4	LITTLE	57	1,13	153	0,24	76,2
5	LONDON	12	0,24	4		45,4

(Gonçalves 2006)

# Comparison - Keywords

O' Shea: upn-os x d-upn-os:

N	WORD	FREQ.	UPN-OS.TXT %	FREQ.	D-UPN-OS.LST	%KEYNESS
1	GALLAHER	58	1,18	0		300,6
2	CHANDLER	44	0,90	0		227,9
3	LITTLE	40	0,82	0		207,2
4	IGNATIUS	38	0,78	0		196,8
5	AMIGO	23	0,47	23	0,04	58,9

Trevisan: upn-ht x d-upn-ht:

N	WORD	FREQ.	UPN-HT.TXT %	FREQ.	D-UPN-HT.LST	%KEYNESS
1	GALLAHER	43	1,06	0		228,4
2	CHANDLER	36	0,88	0		191,1
3	LITTLE	33	0,81	0		175,2
4	LONDRES	11	0,27	3		44,2
5	ANNIE	8	0,20	0		42,4

(Gonçalves 2006)

# Comparison – Keywords

O’Shea (upn-os) vs Trevisan (upn-ht):

N	WORD	FREQ.	UPN-OS.TXT %	FREQ.	UPN-HT.TXT %	KEYNESS
1	IGNATIUS	38	0,78	5	0,12	23,1
2	PRA	13	0,27	0		15,7
3	TÁ	7	0,14	0		8,5
4	TE	6	0,12	0		7,3
5	VOLTA	6	0,12	0		7,3

(Gonçalves 2006)



# Text Aligner

JJ	When we knew him first he used to be rather interesting, talking of faints and worms; but I soon grew tired of him and his endless stories about the distillery.
HT	No princípio, quando o conhecemos, costumava ser interessante com suas conversas sobre vermes e desmaios, mas logo cansara-me dele e de suas intermináveis histórias a respeito da destilaria.
OS	Quando o conhecemos era um sujeito cativante, que falava de bagaço e de serpentinas; mas logo cansei-me dele e de suas histórias intermináveis a respeito do alambique.

(Gonçalves 2006)



# Advantages of CL in Literary Studies

- Analysis of literary text

- Analysis and evaluation of translation(s)

(Gonçalves 2006)



# Wrapping up

CL lets you

- ◆ handle a large quantity of data/text at the same time;
- ◆ extract technical terms from a corpus;
- ◆ compile terminological glossaries;
- ◆ find translational equivalents both in a comparable or parallel corpus;
- ◆ analyze translation strategies in a parallel corpus;
- ◆ evaluate translations in a parallel corpus
- ◆ and whatever else the researcher can come up with...



# Some references

- ◆ Berber Sardinha, T. **Lingüística de Corpus: Histórico e Problemática**, São Paulo: DELTA vol.16 no.2, 2000.  
([http://www.scielo.br/scielo.php?pid=S0102-44502000000200005&script=sci\\_arttext&tling=pt](http://www.scielo.br/scielo.php?pid=S0102-44502000000200005&script=sci_arttext&tling=pt))
- ◆ Halliday, M.A.K. Corpus studies and probabilistic grammar. In KarimAijmer & Bengt Altenberg (Eds.) **English Corpus Linguistics - Studies in honor of Jan Svartvik**, London: Longman, 1991, 30-43.
- ◆ Gonçalves, Lourdes Bernardes. **DUBLINERS SOB A LUPA DA LINGÜÍSTICA DE CORPUS: Uma contribuição para a análise e a avaliação da tradução literária**, 2006.
- ◆ Johns, Tim. 'From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data- driven Learning', CALL Austria 10, pp. 14-34. 1990
- ◆ Johns, Tim. Should you be persuaded - two samples of data-driven learning materials. In Johns and King (eds.) **Classroom Concordancing** (Birmingham University: English Language Research Journal 4). 1-13. 1991.
- ◆ Sinclair, J. M. **Trust the Text: Language, Corpus, and Discourse**, London: Routledge, 2004.





Thank you

Stella

seotagni@usp.br



This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.