



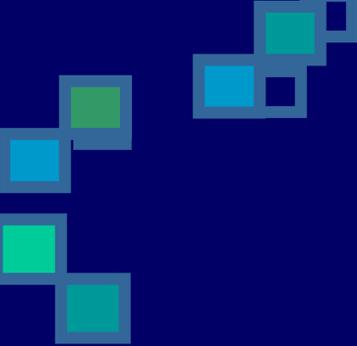
Uma metodologia para o estudo
comparativo de elementos que
garantem a naturalidade
(ou a falta dela)
em textos originais e traduzidos



Elisa Duarte Teixeira

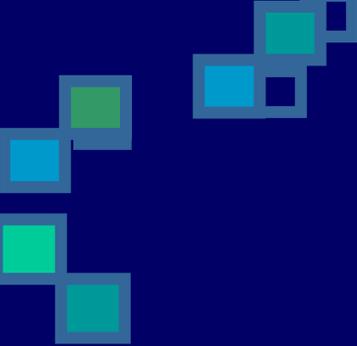
doutorado – USP

SILEL, novembro de 2006



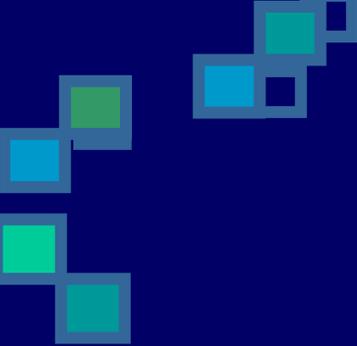
Projeto de doutorado

- Áreas de concentração:
 - Tradução
 - Lingüística de Corpus
 - Terminologia
 - Data de início:
 - julho de 2004
 - Orientação:
 - Profa. Dra. Stella E. O. Tagnin
 - Co-orientação:
 - Prof. Dr. Tony Berber Sardinha
 - Tema:
 - Culinária
- 



Esta apresentação

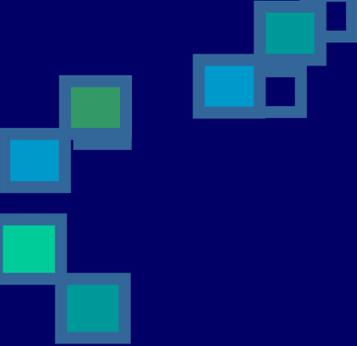
- Naturalidade na tradução
 - Tradução e Lingüística de Corpus
 - Como identificar evidências da naturalidade usando a Lingüística de Corpus
 - Metodologia de pesquisa
 - Caracterização dos corpora
 - Exploração dos corpora com o WordSmith Tools
 - Análise dos resultados
 - Conclusões preliminares
- 



Naturalidade na tradução

- A boa qualidade de uma tradução depende, em grande parte, da naturalidade – especialmente nos textos técnicos
- naturalidade = proximidade / semelhança com outros textos, produzidos por falantes nativos
- Dificuldade de se estudar a naturalidade:
 - ainda não foi feita uma descrição exaustiva dos elementos caracterizadores da naturalidade no texto de partida → não é possível reproduzi-los no texto de chegada



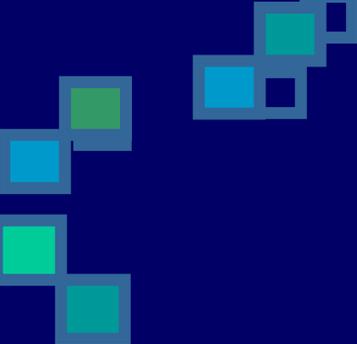


Tradução e L.C.

- Vários autores têm se dedicado a pesquisar a Tradução usando a Lingüística de Corpus
 - McEnery & Wilson (1993) e Sommers (1993) discutem o potencial de corpora paralelos alinhados para diversas áreas do PLN, como a tradução automática
 - Baker (1995), Kenny (1997) e Laviosa (1998) – usando um corpus monolíngüe comparável (TEC), descreveram os universais da tradução
 - Malmkjaer (1998) e Peters & Picht (1998) apontam para a utilidade de corpora paralelos alinhados como fonte de equivalentes na prática da tradução
- 

Tradução e L.C.

- Corpora monolíngües comparáveis (como o TEC) e paralelos alinhados mostraram-se insuficientes para o estudo e a prática da tradução
 - São difíceis de conseguir / compilar
 - Muitas vezes oferecem equivalentes não confiáveis – “unidiomatic constructions” (Teubert, 1996; Bowker, 2000)
- Pesquisadores como Bowker e Pearson (e vários pesquisadores depois delas) demonstraram a utilidade de corpora comparáveis e *specialized target language corpora* na tradução



Naturalidade, Tradução e L.C.

- Os corpora de textos autênticos possibilitam a identificação de padrões lingüísticos próprios da área de especialidade em questão = elementos que garantem a naturalidade no texto traduzido, MAS:
 - O que exatamente diferencia um texto traduzido de um originalmente produzido na língua de chegada?
 - Como identificar e quantificar esses elementos nos corpora de forma automática e/ou semi-automática?
 - Como conscientizar o aprendiz da existência desses elementos?
- 



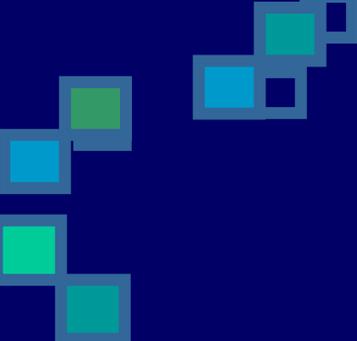
Um estudo piloto



na área da Culinária

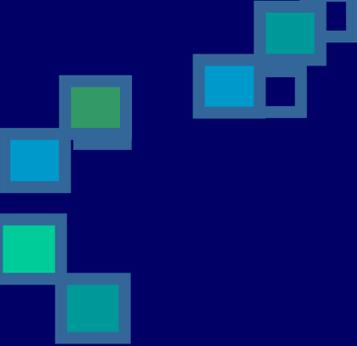
Metodologia: os corpora

- Para o presente estudo, foram usados 3 corpora
 - A versão para o inglês do livro *Cozinhando para amigos*, de Heloisa Bacellar (São Paulo: DBA, 2005):
 - livro completo de capa a capa (127.033 tokens / 6.199 types);
 - somente as receitas e introduções dos menus (100.015 tokens / 4.367 types)
 - somente as receitas (78.292 tokens / 2.708 types)
 - Um corpus de receitas originalmente escritas em inglês (britânico) retiradas da internet – BBC e Great British Cooking (1.551.837 tokens e 11.138 types)
 - Brown Corpus – corpus de referência (língua geral)



Metodologia: ferramentas

- Programa usado: WordSmith Tools (Scott, 1996)
 - Passos:
 - Wordlist das 3 versões do corpus de estudo (livro vertido para o inglês)
 - Wordlist do corpus de comparação (receitas originalmente escritas em inglês)
 - Palavras-chave de cada uma dessas wordlists usando a o Brown Corpus como corpus de referência
 - Comparação entre as wordlists do corpus de estudo e de comparação (Compare 2 wordlists)
 - Concordância e clusters de palavras “suspeitas”
- 



Alguns resultados

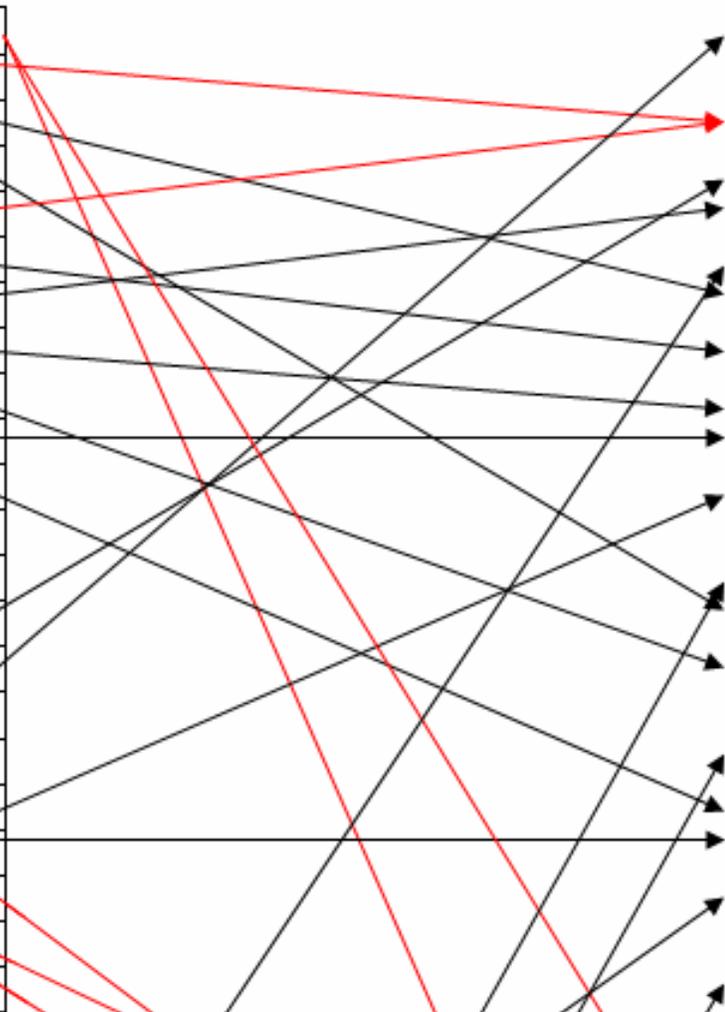
- Comparação da lista de palavras-chave da tradução (somente receitas) e do corpus de receitas originais
 - Keywords de receitas vertidas para o inglês X Brown
 - Keywords de receitas originais em inglês X Brown
 - Comparação entre as 50 primeiras palavras-chave
 - Linhas de concordância dos itens que não foram encontrados ou estavam numa posição muito diferente na lista de palavras-chave do outro corpus (mais de 100)
 - Levantamento da padronização dessas palavras – clusters e concordâncias
- 

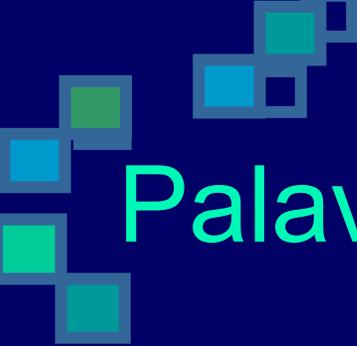
Corpus de referência culinária x Brown

1	TBSP
2	MINUTES
3	ADD
4	OZ
5	MINS
6	OIL
7	UNTIL
8	PAN
9	CHOPPED
10	BUTTER
11	HEAT
12	INGREDIENTS
13	COOKING
14	PEPPER
15	SALT
16	COOK
17	METHOD
18	SUGAR
19	AND
20	SERVES
21	TSP
22	SERVE

Livro traduzido x Brown

SALT
CUP
MINUTES
PEPPER
UNTIL
OLIVE
ADD
OIL
PAN
BUTTER
SUGAR
CUPS
BOWL
OZ
CHOPPED
BAKING
FINELY
HEAT
AND
FLOUR
REFRIGERATE
CHEESE





Palavras “suspeitas” - corpus original

- cooking
 - cook
 - method
 - preparation
 - time
 - into
 - fresh
 - lemon
 - remove
 - place
 - mixture
 - less
 - freshly
 - chicken
 - pour
 - milk
- 

Palavras “suspeitas” - corpus original

- **cooking** (não aparece entre as 1as. 500 palavras-chave do livro traduzido!)
- **cook** (posição 73 do livro traduzido)
- **method**
- **preparation**
- **time**
- **into** (posição 83)
- **fresh** (posição 91)
- **lemon** (posição 165)
- **remove** (posição 67)
- **place** (posição 97)
- **mixture** (posição 59)
- **less** (não aparece entre as 1as. 500 palavras-chave do livro traduzido!) característica do corpus (“preparation time: less than X minutes” = 5.658 oc. x 40 oc. de outros tipos)
- **freshly** (posição 446!)
- **chicken** (posição 125)
- **pour** (posição 63)
- **milk** (posição 86)

Clusters de *cooking* (min. 5 oc.)

■ 15 clusters duplos

N	cluster	Freq.
1	cooking time	5.271
2	overnight cooking	145
3	cooking liquid	77
4	continue cooking	76
5	cooking apples	74
6	cooking liquor	53
7	cooking juices	48
8	finish cooking	34
9	cooking apple	24
10	cooking water	13
11	cooking fat	12
12	cooking process	11
13	cooking oil	10
14	cooking gently	7
15	cooking chocolate	6

10 clusters triplos

N	cluster	Freq.
1	no cooking required	437
2	continue cooking for	37
3	end of cooking	30
4	to finish cooking	23
5	finish cooking for	15
6	continue cooking until	10
7	oil for cooking	9
8	reserved cooking liquid	8
9	continue cooking gently	6
10	reserved cooking liquor	5

Clusters de *cooking* (min. 5 oc.)

■ 10 clusters quadruplos

N cluster	Freq.
1 calculate the cooking time	9
2 cooking to wilt for	9
3 add the cooking liquid	6
4 continue cooking until the	6
5 halfway through the cooking	6
6 reserving the cooking liquid	6
7 stop the cooking process	6
8 collect the cooking juices	5
9 cooking for # minutes	5
10 during the cooking time	5

■ Outros: cooking

- wine (2 oc)
- tomatoes (2 oc.)
- thermometer (2 oc.)
- ring(s) (6 oc.)
- rack (2 oc)
- pot (2 oc)
- milk (2 oc)
- lamb (2 oc)
- foil (2 oc)
- dark chocolate (2 oc)
- broth (4 oc)

Outros padrões (Concordancer)

- halfway through cooking (12 oc – 3 half way)
- strain the cooking liquid/liquor/milk (6 oc)
- collect (all) the cooking liquor/liquid (8 oc)
- the cooking liquid/liquor from the FOOD (5 oc)
- quick(-)cooking (como adjetivo – 3 oc)
- towards the end of cooking (8 oc)
- near the end of cooking (2 oc)
- at the end of cooking (6 oc), X minutes/hours before the end of cooking (time) (14 oc)
- Bramley cooking apple(s) (6 oc)
- just before cooking (4 oc)
- (to) ensure(s/ing) even cooking (6 oc)
- while INGREDIENT is/are cooking, VERB-IMPERATIVE (66 oc)
- whilst INGREDIENT is/are cooking, VERB-IMPERATIVE (15 oc)
- plunge into (ice) cold water to stop the cooking (6 oc)

Palavras “suspeitas” - corpus original

- **cooking** (não aparece entre as 1as. 500 palavras-chave do livro traduzido!)
- **cook** (posição 73 do livro traduzido)
- **method**
- **preparation**
- **time**
- **into** (posição 83)
- **fresh** (posição 91)
- **lemon** (posição 165)
- **remove** (posição 67)
- **place** (posição 97)
- **mixture** (posição 59)
- **less** (não aparece entre as 1as. 500 palavras-chave do livro traduzido!) característica do corpus (“preparation time: less than X minutes” = 5.658 oc. x 40 oc. de outros tipos)
- **freshly** (posição 446!)
- **chicken** (posição 125)
- **pour** (posição 63)
- **milk** (posição 86)

Clusters de *freshly* (min. 5 oc.)

■ Clusters duplos de *freshly* (todos)

N	cluster	Freq.
1	freshly ground	2.730
2	freshly grated	208
3	freshly chopped	84
4	freshly milled	39
5	freshly squeezed	15
6	freshly cooked	13
7	freshly made	10

■ 10 clusters quádruplos de *freshly* (após limpeza)

N	cluster	Freq.
1	freshly ground black pepper	2.417
2	freshly ground white pepper	47
3	freshly milled black pepper	36
4	freshly grated parmesan cheese	33
5	pinch of freshly grated	10
6	plenty of freshly ground	10
7	freshly chopped flat leaf	8
8	freshly squeezed lemon juice	7
9	season with freshly ground	6
10	freshly ground sea salt	5

Clusters de *freshly* (min. 5 oc.)

■ 15 clusters triplos de *freshly*

N	cluster	Freq.
1	freshly ground pepper	218
2	freshly grated nutmeg	68
3	freshly grated parmesan	60
4	freshly chopped parsley	23
5	freshly chopped coriander	8
6	freshly squeezed lemon	8
7	freshly chopped herbs	7
8	freshly chopped chives	6
9	freshly chopped mint	6
10	freshly grated horseradish	6
11	freshly grated salt	6
12	freshly cooked rice	5
13	freshly grated ginger	5
14	freshly ground salt	5
15	sprinkle with freshly	5



Palavras “suspeitas” - livro traduzido

- cup
 - cups
 - baking
 - refrigerate
 - g
 - combine
 - leaves
 - Transfer
 - diced
 - parsley
 - medium
 - preheat
 - brown
 - bake
 - drizzle
 - cm
 - cut
 - slices
- 

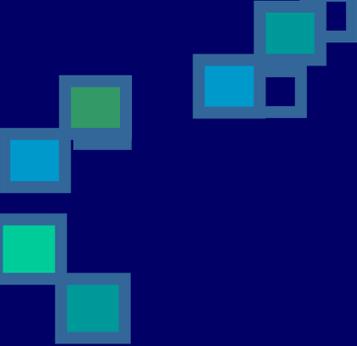
Palavras “suspeitas” - livro traduzido

- **cup / cups** (263 oc. de cup* no corpus de compar. – medidas mais precisas)
- **baking** (posição 99 do corpus de referência)
- **refrigerate** (não aparece entre as 1as. 500 palavras-chave do corpus de referência!)
- **g** (no corpus de compar. a medida está colada no “)”) – WSTools considera 1 palavra)
- **combine** (posição 234)
- **leaves** (posição 84)
- **transfer** (posição 154)
- **diced** (posição 102)
- **parsley** (posição 60)
- **medium** (posição 93)
- **preheat** (posição 87)
- **brown** (posição 146)
- **bake** (posição 90)
- **drizzle** (posição 147)
- **cm** (posição 259) – as medidas anglo-saxãs são as mais usadas (inch)
- **cut** (posição 57)
- **slices** (posição 74)

O que fazer nesses casos?

■ Possíveis soluções:

- Observar a lista de palavras-chave (ou a wordlist) do corpus de receitas originais e procurar sinônimos – testar no corpus (concordâncias):
 - refrigerate: (allow – 204 – to) cool (118), chill (283), ice? (286)
 - combine: add (3), stir (28), mix (46), whisk (129), fold (155), blend (166)
 - transfer: remove (36), place (37), put (189),
 - brown: golden? (91), sauté (171), roast (203), grill (205)
 - drizzle: brush (249), pour (47)
- Procurar equivalentes em contextos semelhantes
 - EX: para encontrar sinônimos de drizzle, pesquisar os contextos de oil, vinegar, sauce, dressing, etc.

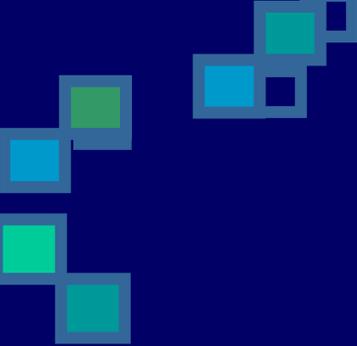


Palavras-chave negativas

- Outra possibilidade de identificação de usos “suspeitos”
 - Fazer a wordlist do corpus e estudo (livro traduzido)
 - Fazer wordlist do corpus de comparação (receitas originalmente escritas em inglês)
 - Na ferramenta Wordlist, usar a opção “compare 2 wordlists”
 - o programa usa o cálculo e as estatísticas de palavras-chave, mas retorna, além dessas, as **palavras-chave negativas** – palavras do corpus de comparação (receitas originais) cuja frequência está abaixo do esperado no corpus de estudo (livro traduzido)
- 

N	WORD	FREQ.	HELOSREN.TXT %	FREQ.	HELENBBC.LST %	KEYNESS	P
1	CUP	772	0,99	182	0,01	3.782,7	0,000000
2	CUPS	368	0,47	81		1.820,2	0,000000
3	G	366	0,47	177	0,01	1.555,7	0,000000
4	LET	370	0,47	388	0,03	1.235,9	0,000000
5	REFRIGERATE	264	0,34	180	0,01	1.022,0	0,000000
6	TEASPOON	208	0,27	195	0,01	724,4	0,000000
7	TABLESPOONS	230	0,29	356	0,02	647,1	0,000000
8	TABLESPOON	195	0,25	259	0,02	589,6	0,000000
9	L	101	0,13	8		557,0	0,000000
10	QT	91	0,12	0		552,6	0,000000
11	PURPOSE	96	0,12	11		513,2	0,000000
12	ADJUST	131	0,17	108		477,1	0,000000
13	COMBINE	234	0,30	732	0,05	423,7	0,000000
14	OBTAINED	80	0,10	11		419,8	0,000000
15	PLASTIC	109	0,14	94		390,9	0,000000
16	FRAGRANT	89	0,11	39		386,9	0,000000
17	SALT	778	0,99	6.759	0,44	385,4	0,000000
18	DAYS	117	0,15	136		374,6	0,000000
19	ABOUT	409	0,52	2.607	0,17	346,8	0,000000
20	WAIT	72	0,09	19		345,9	0,000000
21	BAKING	316	0,40	1.674	0,11	342,3	0,000000
22	CM	198	0,25	732	0,05	311,6	0,000000
23	BELL	53	0,07	3		298,7	0,000000
24	AHEAD	65	0,08	25		290,8	0,000000
25	TRANSFER	252	0,32	1.276	0,08	287,9	0,000000
26	PREPARED	124	0,16	275	0,02	285,6	0,000000
27	TEASPOONS	70	0,09	43		279,2	0,000000
28	WRAP	129	0,16	327	0,02	272,4	0,000000
29	BROWN	318	0,41	2.031	0,13	268,7	0,000000
30	HOURS	270	0,34	1.615	0,10	250,1	0,000000

N	WORD	FREQ.	HELOSREN.TXT %	FREQ.	HELENBBC.LST %	KEYNESS	P
336	WARM	33	0,04	1.397	0,09	23,9	0,000001
337	NEW	1		326	0,02	24,6	0,000001
338	NAIRN	0		258	0,02	25,4	0,000000
339	BEEN	0		259	0,02	25,5	0,000000
340	HADDOCK	0		261	0,02	25,7	0,000000
341	HOUSECALL	0		261	0,02	25,7	0,000000
342	STUFFING	0		261	0,02	25,7	0,000000
343	NICK	0		261	0,02	25,7	0,000000
344	CABBAGE	4		486	0,03	25,7	0,000000
345	THESE	3		445	0,03	26,0	0,000000
346	SCATTER	0		269	0,02	26,5	0,000000
347	FIVE	1		348	0,02	26,6	0,000000
348	EITHER	0		273	0,02	26,9	0,000000
349	FOUR	3		458	0,03	27,1	0,000000
350	FRIED	3		458	0,03	27,1	0,000000
351	FILO	0		278	0,02	27,4	0,000000
352	PEELED	77	0,10	2.652	0,17	27,4	0,000000
353	FL	12	0,02	812	0,05	27,5	0,000000
354	SALMON	20	0,03	1.080	0,07	27,8	0,000000
355	RANKIN	0		284	0,02	28,0	0,000000
356	TROUT	0		287	0,02	28,3	0,000000
357	SQUASH	1		366	0,02	28,3	0,000000
358	THROUGH	44	0,06	1.792	0,12	28,3	0,000000
359	TURMERIC	2		423	0,03	28,4	0,000000
360	REDUCE	13	0,02	861	0,06	28,5	0,000000
361	HERE	6		605	0,04	28,6	0,000000
362	BASE	7		646	0,04	28,7	0,000000
363	WHISK	23	0,03	1.202	0,08	29,6	0,000000
364	BRANDY	1		380	0,02	29,6	0,000000
365	FRY	83	0,11	2.863	0,18	29,7	0,000000



Conclusões preliminares

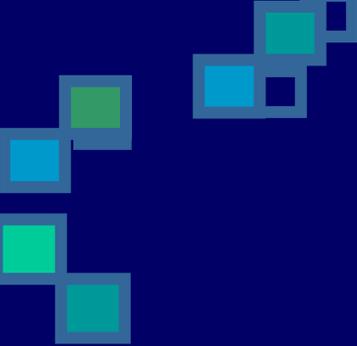
- O uso do programa WordSmith Tools parece facilitar quantitativamente a identificação de elementos caracterizadores da naturalidade em textos originais e traduzidos
 - Isso pode ser feito de várias formas, dentre elas:
 - Contrastando a wordlist do corpus de tradução com a wordlist de um corpus comparável
 - Contrastando a lista de palavras-chave do corpus de tradução com a do corpus comparável, ambas obtidas da comparação com um corpus de língua geral
- 

Conclusões preliminares

- Observando a lista de palavras-chave positivas e negativas obtidas da comparação (Compare 2 wordlists) da wordlist do corpus de tradução com a wordlist de um corpus comparável
- Observando as linhas de concordância e clusters de palavras cujo comportamento se mostre atípico – identificação de padrões lexicais e sintáticos



possíveis elementos garantidores da naturalidade
(ou falta dela)



Futuros desenvolvimentos

- Fazer uma pesquisa mais aprofundada das palavras de comportamento atípico (especialmente as palavras-chave negativas)
 - Testar a metodologia com traduções e corpora comparáveis de outras áreas
 - Criar e aplicar exercícios práticos junto a alunos de tradução e observar sua eficácia para a conscientização da naturalidade
 - Testar a aplicação da metodologia na avaliação de tradução
- 



Obrigada!



elisadut@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.