

**You don't need to be a  
specialist to translate  
specialized language.  
Corpus Linguistics can  
come to your rescue.**

**Stella E. O. Tagnin**

University of São Paulo

**CATS Conference**

Victoria, Canada

June 3-5, 2013

# Outline

- Translator training
  - “deliberate practice” (Shreve 2006)
- Actual practice
  - Case studies
- The “real world”
  - From submission to publication

# Translator training

- Translators cannot be specialists in all areas they translate.
- Adequate translator training can prepare students to cope with different specialized areas.
- Same goes for professional translators.
- Actual “real world” task:
  - **Série 1001 Termos** – what is of interest to the publishing house – market needs
  - Defining the domain: Chemistry

# “Deliberate practice” (Shreve 2006)

- Motivating task: well designed and appropriately difficult
  - a “real world” task
  - best glossaries would be submitted for publication
- Feedback by teacher
- Chance to redo / correct task

# Actual practice: the task

- To build a **corpus-driven** technical glossary for publication
  - 2005: various groups – various domains
  - 2008: various groups – one domain: Photography



# Actual practice: Methodology - Corpus Linguistics

- Introduction to Corpus Linguistics
  - Compiling a specialized bilingual corpus
    - English – Portuguese
    - Comparable corpus: same topics, same genres etc
- Balancing
  - Text types
  - Number of texts
  - Number of words

# Actual practice: Methodology - Corpus Linguistics

- Corpus investigation tools
  - WordSmith Tools (Scott 2012)
  - AntConc (Laurence 2006)
- Term extraction
  - WordLists
  - KeyWord Lists

WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	% Lemmas	Set
1	THE	89,881	6.98	128	99.22	
2	#	68,650	5.33	128	99.22	
3	TO	32,824	2.55	128	99.22	
4	AND	25,422	1.97	128	99.22	
5	A	24,471	1.90	128	99.22	
6	OF	22,229	1.73	126	97.67	
7	IN	20,314	1.58	128	99.22	
8	IS	17,456	1.36	128	99.22	
9	OR	14,885	1.16	126	97.67	
10	YOU	12,614	0.98	114	88.37	
11	CAMERA	2,189	0.95	101	78.29	
12	FOR	10,711	0.83	126	97.67	
13	ON	10,506	0.82	127	98.45	
14	IMAGE	9,946	0.77	111	86.05	
15	WITH	9,498	0.74	128	99.22	
16	BE	8,755	0.68	123	95.35	
17	ARE	7,518	0.58	119	92.25	
18	YOUR	7,478	0.58	111	86.05	
19	CAN	7,388	0.57	121	93.80	
20	THAT	7,009	0.54	120	93.02	
21	IT	6,205	0.48	126	97.67	
22	IMAGES	6,099	0.47	97	75.19	
23	NOT	6,012	0.47	116	89.92	
24	THIS	5,937	0.46	120	93.02	
25	WHEN	5,785	0.45	120	93.02	

WordList

Photography corpus





**VS**

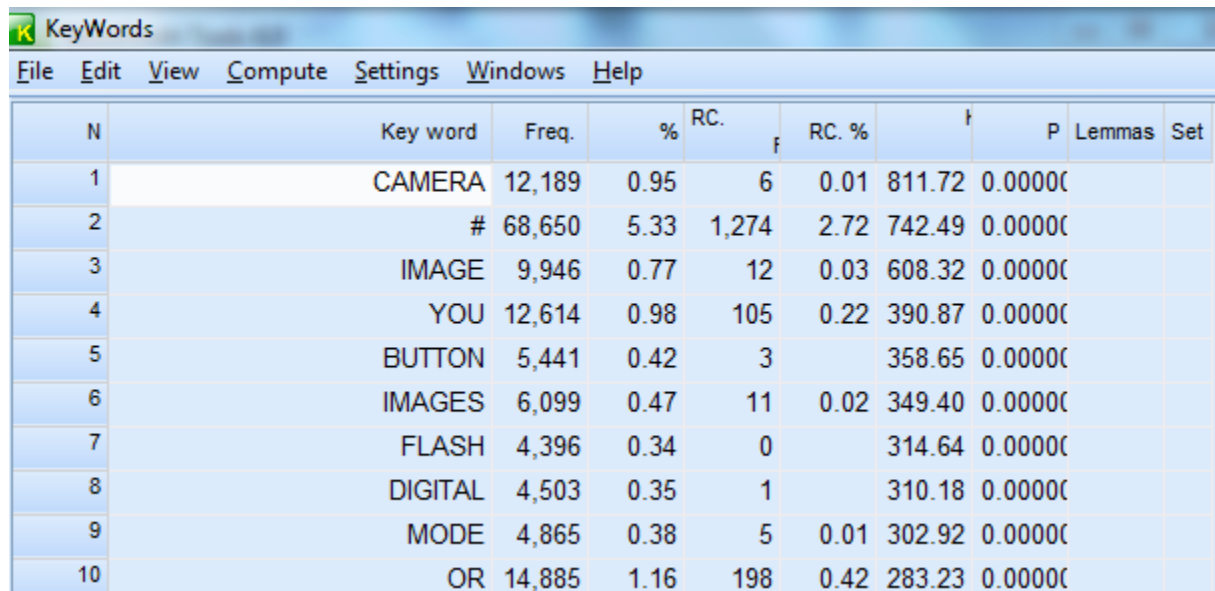


**Study corpus WL**

**Reference/comparison corpus WL**

**Keywords = term candidates**

# Keywords - Photography



N	Key word	Freq.	%	RC.	f	RC. %	t	P	Lemmas	Set
1	CAMERA	12,189	0.95	6	0.01	811.72	0.00000			
2	#	68,650	5.33	1,274	2.72	742.49	0.00000			
3	IMAGE	9,946	0.77	12	0.03	608.32	0.00000			
4	YOU	12,614	0.98	105	0.22	390.87	0.00000			
5	BUTTON	5,441	0.42	3		358.65	0.00000			
6	IMAGES	6,099	0.47	11	0.02	349.40	0.00000			
7	FLASH	4,396	0.34	0		314.64	0.00000			
8	DIGITAL	4,503	0.35	1		310.18	0.00000			
9	MODE	4,865	0.38	5	0.01	302.92	0.00000			
10	OR	14,885	1.16	198	0.42	283.23	0.00000			

# Actual practice: Methodology - Corpus Linguistics

- Identification of equivalents
  - Prima facie translations
  - Collocates
  - Context

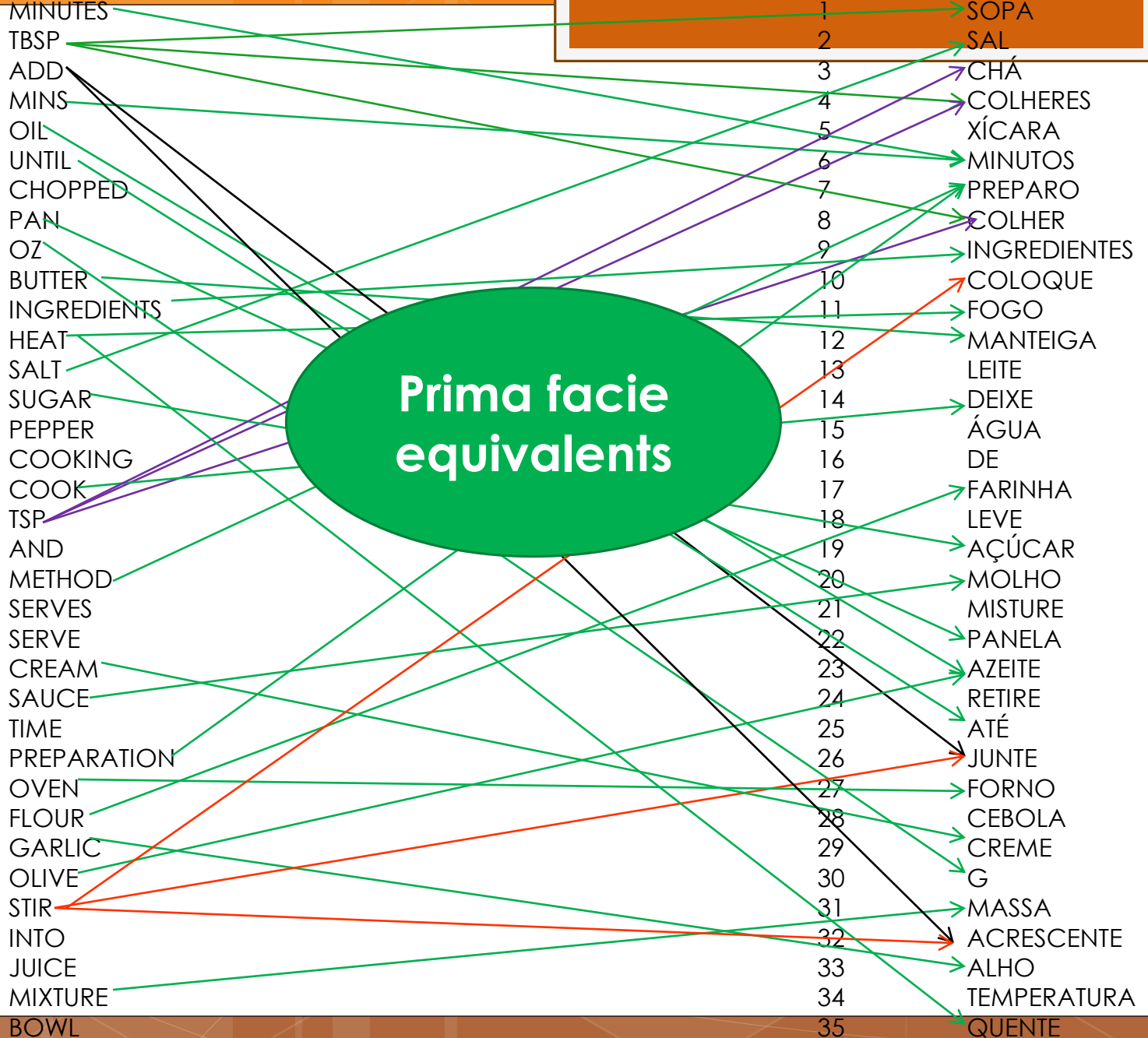
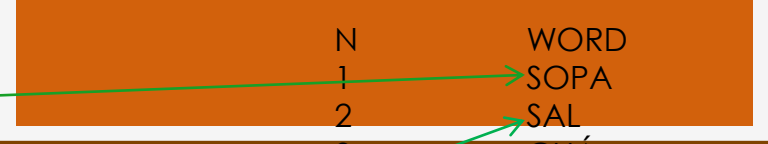
N  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

WORD  
MINUTES  
TBSP  
ADD  
MINS  
OIL  
UNTIL  
CHOPPED  
PAN  
OZ  
BUTTER  
INGREDIENTS  
HEAT  
SALT  
SUGAR  
PEPPER  
COOKING  
COOK  
TSP  
AND  
METHOD  
SERVES  
SERVE  
CREAM  
SAUCE  
TIME  
PREPARATION  
OVEN  
FLOUR  
GARLIC  
OLIVE  
STIR  
INTO  
JUICE  
MIXTURE  
BOWL

**Prima facie  
equivalents**

N  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

WORD  
SOPA  
SAL  
CHÁ  
COLHERES  
XÍCARA  
MINUTOS  
PREPARO  
COLHER  
INGREDIENTES  
COLOQUE  
FOGO  
MANTEIGA  
LEITE  
DEIXE  
ÁGUA  
DE  
FARINHA  
LEVE  
AÇÚCAR  
MOLHO  
MISTURE  
PANELA  
AZEITE  
RETIRE  
ATÉ  
JUNTE  
FORNO  
CEBOLA  
CREME  
G  
MASSA  
ACRESCENTE  
ALHO  
TEMPERATURA  
QUENTE



N	Concordance	Word No	File	%
1		650	_1810.txt	79
2	proximately 35 minutes. 10. For angel hair potatoes, shred the potatoes as finely as possible (thinner than a match stick), using a food processor or m	295	_3272.txt	72
3	oz tomatoes, peeled and halved 1 fresh red chilli, de-seeded and sliced as finely as possible 1 small yellow pepper, de-seeded and thinly sliced 4-6	71	_0116.txt	30
4	d remove the outer layer of the spring onions. 2. Chop all the vegetables as finely as possible without totally crushing them. 3. Mix in all the other ingr	77	_1478.txt	80
5	ves, peppercorns and cardamom pods in a clean coffee grinder and grind as finely as possible. Put the 10-12 garlic cloves and the ginger in the contain	197	_3257.txt	47
6	ing a sharp knife slice it lengthways into 4, then slice the lengths across as finely as possible. After 10 minutes, add this to the pan to join the onions a	222	_0033.txt	36
7	ximately 30 minutes. 3. Remove the seeds and scoop out the baked pulp. Finely chop and place into a bowl. 4. Heat the olive oil and sauté; the onio	183	_0598.txt	54
8	quite cold. water. 3. Finely chop and crush the potatoes. 4. In a frying pan melt the butter and f	85	_0585.txt	58
9	h to hold the stalks and finely chop any stalk tender enough to use, then add the parsley to the pan	119	_4677.txt	61
10	un-dried tom en, or just finely chop before adding to the pan. Season well	247	_4117.txt	62
11	body flaps -seed and finely chop chillies, garlic and coriander	7	_4308.txt	50
12	after shelling 4 c. celery finely chopped 1 tsp. minced garlic	76	_064.txt	53
13	/ 1lb skinle shed, and finely chopped 1 tsp curry powder	76	_064.txt	16
14	For the po garlic cloves finely chopped 1 lb of pork on t	1	_0000.txt	11
15	(optional) s tum onion, finely chopped 1 clove garlic, pe	1	_0000.txt	19
16	sed 2 tsp s en chilies, finely chopped 1 tsp chopped fr	1	_0000.txt	46
17	hutney 6 to ve garlic, finely chopped 1 medium-hot red c	1	_0000.txt	43
18	Ingredient small onion, finely chopped 1 Fresh lemon juice	1	_0000.txt	49
19	e: 1 tsp oliv beeled and finely chopped 1 sprig fresh thyme 2 bay	1	_3800.txt	30
20	ients 4 x red onion, finely chopped 1 red finger chilli, seeded and finely	35	_2188.txt	21
21	chopped 1 s ves garlic, finely chopped 1 large carrot, peeled and finely chopped 30g / 1oz dried c	65	_0270.txt	15
22	open-cup m ove garlic, finely chopped 1 tsp Worcestershire sauce 2 tbsp parsley, chopped 2 tbs	59	_1434.txt	22
23	ients 1 bu s 1 onion, finely chopped 1 large red chilli, finely chopped 1 litre / 13/4 pints vegetabl	39	_0037.txt	17
24	romage frais ng onions, finely diced salt and freshly ground black pepper (to taste) For the pears i	84	_5085.txt	17
25	cream 1 tb 1 tomato, finely diced salt and freshly ground black pepper, to taste Method 1. P	141	_3656.txt	36
26	nts 250g / beeled and finely diced salt freshly ground black pepper 4 tbsp olive oil 2 tbsp fresh	38	_1060.txt	32
27	from the he place the finely diced shallots into a bowl and crumble the cheese over the top. Stir we	109	_3295.txt	59
28	our Ingre ped 1 tsp finely diced shallots 1/2 tsp crushed garlic 1 tomato, blanched and diced	41	_2949.txt	16
29	lsa 1/4 man eeded and finely diced small piece fennel head, finely diced handful soft herbs (eg co	117	_1581.txt	41
30	ring onion, finely chopped 1/4 beef tomato, finely diced 1/4 yellow pepper, finely diced small handful of fresh basil and chives, chopped dash of Taba	93	_5002.txt	34
31	/ 50g mature Cheddar, grated 1oz / 25g Parmesan (Parmigiano Reggiano), finely grated 1 pint / 570ml milk 11/2oz / 40g plain flour 11/2oz / 40g butt	42	_0045.txt	30
32	finely grated zest and juice 1 orange, finely grated zest and juice 1 carrot, finely grated 1 medium cooking apple, peeled and grated 2 tbsp brandy or	103	_1953.txt	23
33	mon juice For the icing: 175g / 6oz icing sugar, sifted 1 lemon, zest only, finely grated 1 tsp lemon juice Method 1. Preheat the oven to 150C /	93	_1872.txt	35
34	7oz smoked salmon 2 tbsp roughly chopped fresh dill 1 lemon, zest only, finely grated 1 tsp horseradish sauce 1 tsp drained capers, chopped (opt	66	_2667.txt	29
35	(optional): 20g / 3/4oz unsalted butter 30g / 13/4oz extra bitter chocolate, finely grated 1 tsp cocoa To Serve: 200ml / 7fl oz whipping cream 50g /	75	_0851.txt	14
36	pint natural yoghurt 4 large cloves garlic, crushed 5cm piece fresh ginger, finely grated 1 red chilli, finely chopped (optional) 2 lemons, juice only Sp	44	_4280.txt	44
37	(25 g) parmesan (Parmigiano Reggiano), finely grated 1 oz (25 g) gruyère, finely grated 1 heaped tablespoon finely snipped chives 1/2 oz (10 g) butte	55	_0137.txt	17
38	ature cheddar, finely grated 1 oz (25 g) parmesan (Parmigiano Reggiano), finely grated 1 oz (25 g) gruyère, finely grated 1 heaped tablespoon finely	48	_0137.txt	16

Concordance Sort

Main Sort ...

- 1R
- 2R
- 3R
- 4R
- 5R

then by ...

- 2R
- 3R
- 4R
- 5R
- Context

finally by

- 1R
- 2R
- 3R
- 4R
- 5R

Sort

all

Case Sensitive

Ascending

A with A, Ç with C etc.

OK Cancel Help



# Looking for equivalents

- *finely* (2.987 occ.) = “finamente” (337 occ.)
- **Not equivalent**  
(*finely* ≠ *finamente*)
- Indirect method: *chop* = “picar”
- collocates of “pique/picad\*”  
in Portuguese corpus

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3
1	DE	11121	5940	5181	783	00	939
2	PICADA	3727	239	230	155	46	32
3	E	3253	1647	1606	226	230	432
4	SOPA	2661	1503	1158	135	231	19
5	PICADO	2261	172	162	75	58	28
6	A	2066	884	1182	201	111	138
7	CHÁ	1737	849	888	78	182	579
8	CEBOLA	1681	1403	277	40	68	69
9	XÍCARA	1590	831	759	181	447	202
10	COLHERES	1446	675	771	217	449	8
11	O	1439	659	780	175	84	54
12	COLHER	1372	656	716	198	438	19
13	ALHO	1228	833	395	83	79	39
14	PICADOS	1215	107	130	35	45	25
15	PICADAS	1205	95	69	44	33	14
16	EM	1062	270	792	87	69	35
17	SALSINHA	1036	892	144	19	28	33
18	G	957	511	446	179	132	195
19	SAL	770	141	629	39	48	33
20	PIQUE	740	5	13	1	1	0
21	AS	707	378	329	78	68	73
22	BEM	669	413	256	32	21	15
23	GOSTO	646	146	500	41	41	16
24	COM	643	357	286	80	72	119
25	SEM	628	403	225	126	30	47
26	DENTES	551	316	235	20	30	266
27	MODO	550	93	457	3	4	86
28	OS	545	229	316	51	46	37
29	CEBOLINHA	536	427	109	8	8	13
30	PREPARO	512	147	365	18	24	6
31	TOMATES	450	262	188	23	44	32
32	SALSA	448	404	44	5	14	7
33	COL	429	228	201	37	190	1
34	VERDE	381	306	75	18	10	6
35	MANTEIGA	376	228	148	59	81	46
36	FOLHAS	374	258	116	29	46	143
37	PIMENTA-DO-REINO	351	74	277	16	33	7

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
38	AZEITE	338	108	230	53	39	15	1	0	0	10	43	13	38	126
39	PIMENTÃO	320	194	126	9	12	41	105	27	0	0	68	22	15	21
40	MÉDIA	319	256	63	12	5	8	10	221	0	0	0	38	18	7
41	GROSSEIRAMENTE	311	74	237	7	2	4	1	60	0	182	29	22	0	4
42	PIMENTA	305	132	173	13	44	21	42	12	0	1	21	80	42	29
43	DENTE	299	187	112	6	25	155	0	1	0	0	56	48	5	3
44	MANJERICÃO	288	213	75	7	5	14	19	168	0	2	8	3	17	45
45	SEMENTES	284	225	59	30	23	10	24	138	0	0	4	7	16	32
46	TOMATE	279	184	95	13	26	31	46	68	0	0	28	21	24	22
47	PICADINHO	260	54	15	13	13	11	10	7	191	1	0	3	5	6
48	PELE	259	184	75	13	106	16	21	28	0	0	16	8	20	31
49	COENTRO	251	209	42	4	12	11	15	167	0	2	8	2	9	21
50	PEDAÇOS	246	50	196	21	12	12	4	1	0	0	103	41	40	12
51	PICADINHA	244	21	7	18	0	2	1	0	216	1	0	3	1	2
52	SALSÃO	241	178	63	14	18	9	15	122	0	0	8	6	28	21
53	ACRESCENTE	236	167	69	18	22	127	0	0	0	1	24	9	12	23
54	FINAMENTE	236	39	197	4	4	2	1	28	0	149	28	12	4	4
55	PEQUENOS	233	48	185	14	10	14	4	6	0	0	6	94	45	40
56	COLOQUE	228	55	173	15	6	33	1	0	0	0	29	26	48	70
57	CEBOLAS	222	176	46	15	6	11	84	60	0	0	14	18	7	7
58	NOZES	221	189	32	5	6	6	33	139	0	0	2	7	7	16
59	PEQUENA	212	163	49	7	7	4	13	132	0	1	2	22	18	6
60	ÓLEO	207	108	99	34	38	34	1	1	0	2	13	7	25	52
61	PARA	206	49	157	20	10	15	3	1	0	32	22	28	33	42
62	QUEIJO	202	76	126	15	16	7	30	8	0	0	13	7	40	66
63	OU	198	110	88	21	11	20	43	15	0	44	3	9	14	18
64	CENOURA	196	114	82	7	15	13	30	49	0	0	37	12	13	20
65	CHOCOLATE	196	155	41	4	4	86	28	33	0	0	20	11	5	5
66	XÍCARAS	193	74	119	19	37	18	0	0	0	0	63	33	14	9
67	AZEITONAS	188	122	66	9	16	9	50	38	0	0	25	5	11	25
68	RESERVE	188	90	98	7	14	11	7	51	0	4	16	21	25	32
69	UMA	188	82	106	32	25	11	12	2	0	5	22	20	24	35
70	CUBOS	184	52	132	24	13	8	4	3	0	0	69	23	29	11
71	XÍC	184	72	112	29	43	0	0	0	0	0	27	70	11	4
72	HORTELÃ	182	149	33	5	8	6	10	120	0	2	3	8	7	13
73	DO	181	78	103	24	11	12	31	0	0	1	2	24	48	28
74	GRANDE	181	131	50	7	9	7	11	97	0	0	0	23	19	8

# Results for picad\*

Most frequent collocates:

- **“bem”** (669 occurrences): “bem picada”, “bem picado”, “bem picadinho” etc.
- **“picadinh\*”** (504 occurrences, while 69 are of “bem picadinh\*”)



Best equivalents for *finely chopped*  
**“bem picad\*” ou “picadinh\*”**



# Actual practice: Building the glossary

- A translator-oriented glossary
  - Equivalents
  - Authentic examples
  - Collocates and phraseology in general
  - Cultural information where relevant
  - Suggestion of substitutions (cooking)

# Actual practice: Building the glossary

- Hand-in glossary
- Review by teacher
- New submission (optional)

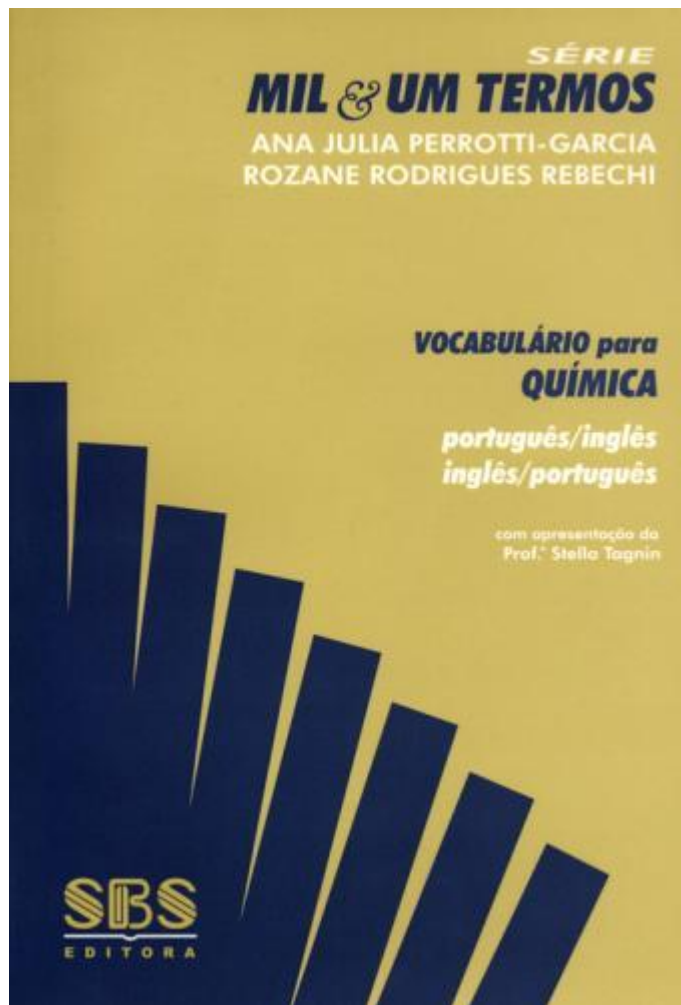
# Photography

- Life is not a bowl of cherries...
- Domain subdivided:
  - Camera
  - Studio
  - History
  - Light
  - Preservation
  - Digital photography
- Not all glossaries met publishing quality
- New group to correct and improve material produced

# The real world: from submission to publication

- Submission to publisher
- Review by specialist
- Back to authors for corrections and changes
- New submission
- Wait...
- Wait...
- Wait...
- **Publication!!!**

# Publications



**Vocabulário  
para Química**  
Ana Julia Perrotti-  
Garcia  
Rozane Rodrigues  
Rebechi  
(2007)

# Publications

## Vocabulário para Culinária

Elisa Duarte  
Teixeira  
Stella E. O.  
Tagnin  
(2008)





# Vocabulário para Fotografia

Eliana C. R. Antonopoulos  
Helena Akemi Misumi  
Moir Martins de Andrade  
Maria Angelica Royo  
Veridiana Rocha Schwenck  
(2013)

# Final remarks

- Not all glossaries were publishable
- Nevertheless
  - students became familiar with Corpus Linguistics
  - learned how to build a glossary
  - and, by extension, learned how to mine for lexical information



# Says one student

- “ I knew nothing about Chemistry so that it was difficult to select the texts for the corpus. Consequently, some had to be discarded later on.”



○ = difficulty of task

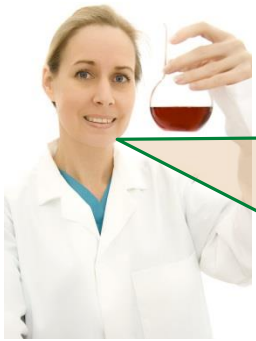
- : “the process really worked and using the corpus made it possible to overcome the problems you experience when working with an area in which you are not a specialist.”

- → becoming familiar with the area



# Says the reviser

- herself a translator and Chemical Engineer:



“I would have never been able to come up with the terms as they did”

- → how efficient CL is
- → language actually used by that professional community

# Most importantly

**“the learner’s authentic experience allows him/her to apply the knowledge acquired to any other area in which he/she may come to work in”** (Alves & Tagnin, 2010)

**Thank you**

**Stella**  
**seotagni@usp.br**

# References

- Alves, F. & S.E.O. Tagnin. 2010. *Corpora e Ensino de Tradução. In Corpora no Ensino de Línguas Estrangeiras*, São Paulo: HUB, pp. 189-203.
- Anthony, L. 2006. ConCORDancing with AntConc - An Introduction to Tools and Techniques in Corpus Linguistics - Proceedings of the JACET 45th Annual Convention pp. 218-219.
- Scott, M., 2012, WordSmith Tools version 6, Liverpool: Lexical Analysis Software.
- Shreve, G. 2006. The deliberate practice: translation and expertise. *Journal of Translation Studies*, v. 9, n. 1, p. 27-42.