

Some methodological questions about *significance* in empirical investigations of translations

An exploratory and informal discussion

Erich Steiner, Saarbrücken, March 2007

1. Independent Variables

1.1. Structure of corpus

1.2. Types of contrast

2. Dependent Variables

2.1. Low-level features

3. Some low-level results

4. Significance

5. Appendix: Some low-level statistics on corpus data

1. Structure of corpus

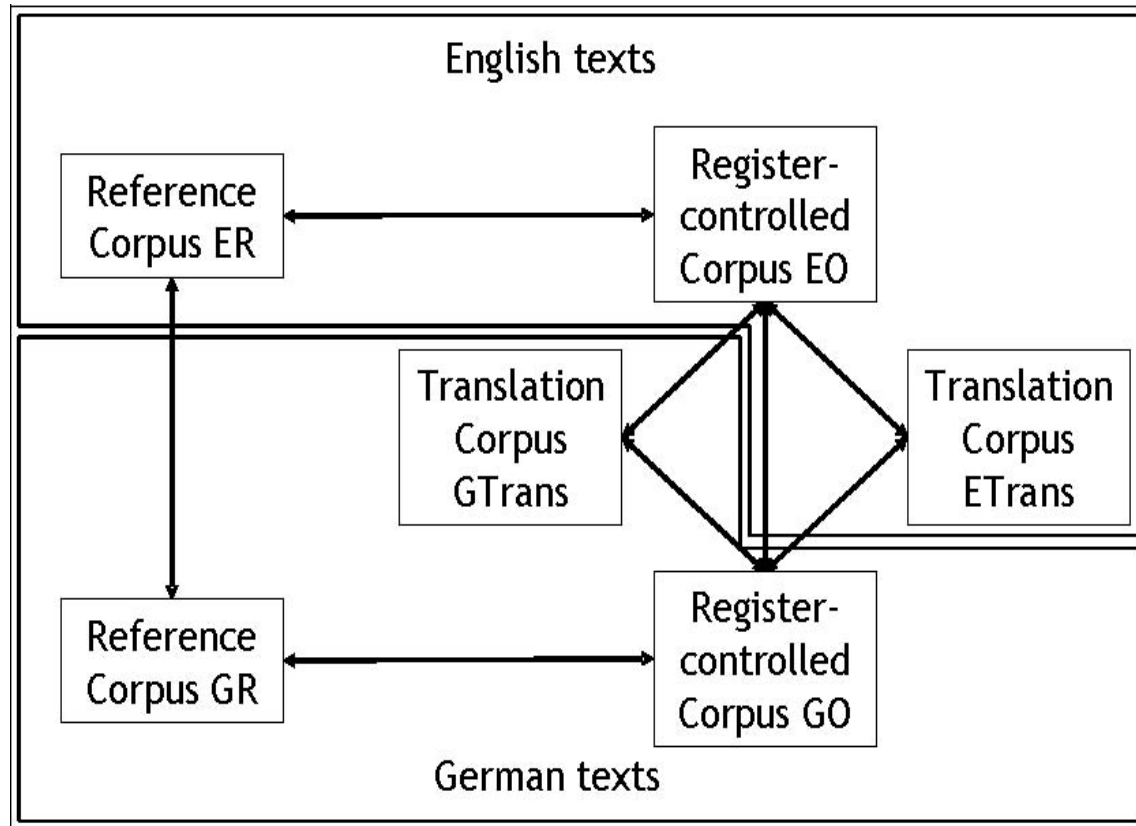


Figure 1: Structure of the CroCo-Corpus

Contrast C2 (register controlled)

C2.1. (within one register, original texts between languages differentiated into 8 sub-contrasts by register)

Lexical density

	E-Original	G-Original
Essay	56,76	57,15
Fiction	47,73	51,05
Instruction	58,77	58,36
Popsci	55,55	54,18
Share	58,02	56,16
Speech	54,02	52,79
Tourism	59,14	60,74
Web	58,21	58,03
TOTAL_REG	56,03	56,06

Table 1: Lexical density in 8 registers of original texts

Type-token relationship

	E-Original	G-Original
Essay	9,19	18,66
Fiction	12,66	19,14
Instruction	7,00	14,29
Popsci	10,98	17,55
Share	9,39	16,01
Speech	9,83	15,87
Tourism	12,58	24,17
Web	10,99	21,07

Table 2: Type token relationships in 8 registers of original texts

Appendix: Some low-level statistics on corpus data

	Lexical Density in %					
	English		German		Reference Corpora	
	Original	Translation	Original	Translation	English	German
Essay	56,76	55,99	57,15	54,39		
Fiction	47,73	48,40	51,05	50,20		
Instruction	58,77	59,56	58,36	55,04		
Popsci	55,55	51,29	54,18	56,98		
Share	58,02	56,21	56,16	56,13		
Speech	54,02	51,19	52,79	52,80		
Tourism	59,14	59,24	60,74	59,08		
Web	58,21	56,88	58,03	56,93		
Av. TOTAL_REG	56,03	54,85	56,06	55,19	57,49	56,98
Av. TOTAL_LANG	55,44		55,62			

Type Token Ratio in %						
	English		German		Reference Corpora	
	Original	Translation	Original	Translation	English	German
Essay	9,19	9,37	18,61	14,63		
Fiction	12,66	11,85	19,14	18,33		
Instruction	7,00	6,82	14,29	14,05		
Popsci	10,98	10,38	17,55	17,16		
Share	9,39	9,25	16,01	16,52		
Speech	9,83	8,56	15,87	14,73		
Tourism	12,58	13,33	24,17	19,93		
Web	10,99	11,45	21,07	18,38		
Av. TOTAL_REG	10,33	10,13	18,34	16,71	15,64	21,71
Av. TOTAL_LANG		10,23		17,53		
TOTAL	5,80	5,60	11,25	9,92		

Parts of Speech in the German Subcorpora in %

			Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct.	Others
--	--	--	-------------	----------------	-------------	------------------	---------------	--------------	----------------	-------------	-----------------	---------------	---------------	---------------

German	TRANS	Essay	25,49	8,13	12,14	11,20	3,10	8,38	11,83	5,40	1,82	0,87	11,16	0,47
		Fiction	17,16	14,64	15,12	7,52	5,67	6,91	7,99	6,18	2,44	0,48	15,60	0,30
		Instruction	25,27	8,63	12,79	7,45	2,72	8,00	10,94	5,28	2,24	0,83	14,82	1,02
		Popsci	22,85	8,81	12,78	10,65	6,00	8,55	10,03	4,71	1,79	1,07	12,43	0,33
		Share	25,08	9,15	11,40	9,47	4,32	9,39	9,35	5,28	1,56	3,10	10,90	0,99
		Speech	24,80	9,63	12,81	9,63	3,78	8,17	11,14	5,62	1,99	0,66	11,41	0,36
		Tourism	29,13	6,78	8,92	10,28	3,40	8,95	10,35	5,01	0,95	1,28	13,19	1,77
		Web	26,26	8,75	12,29	9,07	2,90	8,35	8,54	5,45	1,92	1,46	13,60	1,40
		TOTAL	23,84	8,67	12,08	9,69	4,89	8,39	10,29	4,85	1,75	1,47	13,34	0,72
ORI	Essay	24,35	7,05	11,69	10,88	5,12	8,38	11,23	4,84	1,77	1,60	12,49	0,61	
	Fiction	17,57	13,58	14,43	7,37	6,25	7,23	8,75	4,76	2,64	0,52	16,46	0,43	
	Instruction	26,22	5,94	13,41	7,69	3,43	8,22	11,01	4,50	1,71	2,08	14,83	0,96	
	Popsci	20,28	10,03	12,69	9,82	6,50	7,61	10,78	4,94	1,90	0,99	13,89	0,57	
	Share	25,30	9,26	11,64	10,69	4,30	9,27	10,30	4,44	1,52	2,19	10,45	0,64	
	Speech	21,86	10,81	13,42	9,88	5,51	7,83	10,12	5,13	1,98	0,75	12,23	0,49	
	Tourism	28,24	4,86	8,40	10,73	4,17	9,95	10,08	5,01	1,06	2,19	14,03	1,29	
	Web	26,90	7,83	10,96	10,46	3,88	8,60	10,07	5,22	1,45	1,44	12,37	0,82	
	TOTAL	24,51	9,32	12,28	9,41	3,99	8,34	10,02	5,37	1,84	1,22	12,89	0,83	
TOTAL_DE		24,17	8,99	12,18	9,55	4,44	8,36	10,16	5,11	1,80	1,34	13,12	0,78	
REF_DE		22,93	8,45	13,04	9,20	5,02	8,08	9,88	4,47	1,90	2,27	14,12	0,63	

Parts of Speech in the English Subcorpora in %

	Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct	Others
--	-------------	----------------	-------------	------------------	---------------	--------------	----------------	-------------	-----------------	---------------	--------------	---------------

English	TRANS	Essay	26,82	2,58	13,89	8,75	5,10	11,97	12,12	4,79	2,01	2,08	9,81	0,06
		Fiction	19,94	9,94	15,44	4,73	6,51	10,01	10,51	4,83	2,73	1,11	14,11	0,14
		Instruction	29,82	2,05	16,55	5,42	4,04	9,53	11,93	4,24	1,78	2,76	11,58	0,30
		Popsci	22,14	4,74	15,25	7,31	5,54	11,49	12,53	5,66	2,09	1,52	11,31	0,41
		Share	27,55	5,24	14,05	7,90	4,12	12,32	10,48	4,35	2,26	2,46	9,19	0,09
		Speech	23,22	5,37	15,84	8,40	5,53	11,19	11,85	5,17	2,74	1,23	9,44	0,03
		Tourism	31,75	2,19	10,60	7,61	3,65	11,94	11,71	4,48	1,42	2,46	12,02	0,18
		Web	27,87	4,22	13,08	8,71	4,03	11,92	11,12	4,85	1,77	2,00	10,25	0,16
		TOTAL	26,14	4,54	14,34	7,35	4,82	11,30	11,53	4,80	2,10	1,95	10,96	0,17
	ORI	Essay	27,81	3,61	15,36	8,29	3,66	10,95	10,53	5,55	2,92	1,57	9,71	0,04
		Fiction	19,11	10,06	16,66	5,30	5,81	9,65	10,15	5,73	2,40	0,99	13,97	0,15
		Instruction	29,67	3,61	15,89	5,20	3,18	9,15	11,35	4,78	2,75	2,01	12,18	0,22
		Popsci	25,57	3,14	16,17	7,76	4,98	11,31	11,53	5,53	2,20	1,48	10,13	0,19
		Share	29,14	5,34	13,83	6,97	3,15	11,01	9,02	5,49	2,18	2,95	10,73	0,19
		Speech	25,58	5,34	15,92	8,60	3,55	10,94	10,22	5,87	3,03	1,23	9,62	0,09
		Tourism	31,65	2,45	10,88	8,38	3,63	11,50	10,81	5,12	1,62	1,79	12,13	0,04
		Web	29,11	4,34	14,47	6,74	3,46	10,48	9,57	5,45	2,35	2,15	11,81	0,08
		TOTAL	27,21	4,73	14,90	7,16	3,93	10,62	10,40	5,44	2,43	1,77	11,29	0,12
		TOTAL_EN	26,67	4,64	14,62	7,26	4,37	10,96	10,96	5,12	2,27	1,86	11,12	0,15
		REF_EN	24,60	5,46	15,72	6,24	4,63	10,55	10,32	5,12	2,64	2,41	11,97	0,34

Parts of Speech per Subcorpus in %

		Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct	Others
--	--	-------------	----------------	-------------	------------------	---------------	--------------	----------------	-------------	-----------------	---------------	--------------	---------------

English	Reference	24,60	5,46	15,72	6,24	4,63	10,55	10,32	5,12	2,64	2,41	11,97	0,34
	Original	27,21	4,73	14,90	7,16	3,93	10,62	10,40	5,44	2,43	1,77	11,29	0,12
	Translation	26,14	4,54	14,34	7,35	4,82	11,30	11,53	4,80	2,10	1,95	10,96	0,17
German	Original	24,51	9,32	12,28	9,41	3,99	8,34	10,02	5,37	1,84	1,22	12,89	0,83
	Translation	23,84	8,67	12,08	9,69	4,89	8,39	10,29	4,85	1,75	1,47	13,34	0,72
	Reference	22,93	8,45	13,04	9,20	5,02	8,08	9,88	4,47	1,90	2,27	14,12	0,63