Frequency tables

1. Frequencies in texts originally written in Spanish, English and German

A. Corpus component ES						
Construction	Raw occurrences[1]	Analysed (thinned to)[2]	Positives	%[3]	Absolute frequency	Frequency pmw
ser + PP	3,251	1,236	852	68.9	2,241	537.4
se-constructions	40,231	2,000	189	9.5	3,865	926.7
3rd p. pl.	65,306	2,000	140	7.0	4,571	1,096.2
uno /una	952	952	152	16.0	152	36.4
2nd p. sg.	14,850	2,000	35	1.8	260	62.3
Total	124,590	8,188	1,368		11,089	2,659.1

B. Corpus component EN						
Construction	Raw occurrences	Analysed (thinned to)	Positives	%	Absolute frequency	Frequency pmw
to be + PP	8,162	2000	1271	63.6	5,187	4,585.9
to get + PP	97	97	29	29.9	29	25.6
middle voice	350	350	0	0.0	0	0.0
one	195	195	53	27.2	53	46.9
you	3,378	1200	55	4.6	155	136.9
Total	12,182	3,842	1,409		5,415.8	4,796.2

C. Corpus component DE (German texts)						
Construction	Raw occurrences	Analysed	Positives	%	Absolute frequency	Frequency pmw
werden + PP	578	578	521	90.1	521	1,301.8
man	758	758	758	100.0	758	1,894.0
bekommen + PP	10	10	2	20.0	2	5.0
gehören + PP	7	7	1	14.3	1	2.5
sich lassen	104	104	43	41.3	43	107.4
sein zu + INF	130	130	54	41.5	54	134.9
Gerundivum	8	8	2	25.0	2	5.0
es + Vfin	223	223	2	0.9	2	5.0
S + V + refl. pron. + adj (middle voice)	27	27	0	0.0	0	0.0
	1,845	1,845	1,383		1,383	3,455.7

[1] Number of raw occurrences produced by a search in COVALT using the search strings listed in Appendix 4.

[2] If the number of raw occurrences is >1200, the data are randomly thinned according to the following rules: 1201-5000: thinning to 1200, >5000: thinning to 2000. The thinned dataset is manually analysed and the proportion of positive results is extrapolated to the total number of raw occurrences. The absolute frequency is either equal to the number of positives or, in the case of thinning, the result of the extrapolation.

[3] Percentage of positive results out of the raw data set.

2. Frequencies in Spanish translated and non-translated texts

A. Corpus component ES[4]						
Construction	Raw occurrences	Analysed (thinned to)	Positives	%	Absolute frequency	Frequency pmw
ser + PP	3,251	1,236	852	68.9	2,241	537.4
se-constructions	40,231	2,000	189	9.5	3,865	926.7
3rd p. pl.	65,306	2,000	140	7.0	4,571	1,096.2
uno /una	952	952	152	16.0	152	36.4
2nd p. sg.	14,850	2,000	35	1.8	260	62.3
Total	124,590	8,188	1,368		11,089	2,659.1

D. Corpus component EN-ES (Texts translated from English into Spanish)						
Construction	Raw occurrences	Analysed (thinned)	Positives	%	Absolute frequency	Frequency pmw
ser + PP	1,119	1,119	869	77.7	869	774.3
se-constructions	11,787	2,000	250	12.5	1,473	1,312.8
3rd p. pl.	18,847	2,000	80	4.0	754	671.7
uno /una	190	190	36	18.9	36	32.1
2nd p. sg.	2,498	1,200	28	2.3	58	51.9
Total	34,441	6,509	1,263		3,190	2,842.8

E. Corpus component DE-ES (Texts translated from German into Spanish)						
Construction	Raw occurrences	Analysed (thinned)	Positives	%	Absolute frequency	Frequency pmw
ser + PP	239	239	165	69.0	165	380.2
se-constructions	4,608	2,000	214	10.7	494	1,136.2
3rd p. pl.	6,623	2,000	69	3.5	228	526.5
uno /una	91	91	23	25.3	23	53.0
2nd p. sg.	1,608	1200	11	0.9	15	34.0
Total	13,169	5,530	482		925	2,129.9

---

[4] This is the same table A as above, repeated here simply for ease of reference.