Lexical preferences in the Perfect construction

Martijn van der Klis
UiL OTS, Utrecht University

The PERFECT (e.g. "Mary has read *Alice*") forms an interesting category both within languages and cross-linguistically. In this study, we aim to find verbs appearing more or less frequently than expected in the PERFECT construction. We use a collostructional analysis following Stefanowitsch and Gries (2003).

As Goldberg et al. (2004) state, high-frequency collexemes could facilitate language learners to note correlations between meaning of the word and the construction itself. We set out to find whether these collexemes exist for English, and whether there is any variation in genre or across languages.

Genre variation (BNC)

data: BNC, >100 occurrences, -be, have, got

Spoken

lemma	#perfects	#total	logpvF
hear	1078	5198	∞
see	2547	32843	251,5
finish	516	2589	209,0
lose	511	2659	199,6
change	489	3364	139,7
happen	553	5967	78,8
decide	242	1751	65,3
mention	232	1747	59,4
make	1211	19656	58,1
notice	160	1008	51,4
live	118	3693	-1,3
play	121	3986	-2,0
talk	214	7950	-7,3
go	2397	71338	-10,3
keep	108	6822	-26,4
look	325	20496	-78,7
say	1341	63330	-136,8
do	3996	171766	-317,4
think	203	52349	−∞
know	237	61719	−∞

Fiction

lemma	#perfects	#total	logpvF
hear	966	14127	247,2
see	1911	47716	197,9
happen	564	8625	135,3
change	343	4348	103,9
finish	287	3167	101,2
meet	455	7453	99,1
lose	327	4959	79,4
forget	278	4650	59,0
tell	963	29214	56,4
learn	197	2952	48,9
keep	136	12614	-13,0
take	498	35941	-14,2
ask	263	21724	-15,5
turn	137	17916	-36,8
feel	125	24245	-79,5
want	128	27022	-95,5
know	426	55008	-112,5
think	322	46886	-113,3
do	1501	141078	-160,8
say	563	109254	−∞

Newspapers

lemma	#perfects	#total	logpvF
become	679	5880	67,4
appoint	165	722	52,9
agree	330	2304	52,2
decide	350	2661	46,9
change	295	2141	43,5
launch	231	1484	42,8
sign	235	1535	42,3
criticise	125	583	37,3
prove	263	2031	34,3
lose	495	4930	33,4
claim	127	3764	-10,1
hold	147	4652	-14,8
use	191	5662	-14,9
go	763	17900	-16,8
keep	139	4708	-17,7
run	133	4974	-23,2
know	150	8009	-63,9
look	100	7320	-79,5
do	797	27762	-112,5
say	435	53576	−∞

Academic

lemma	#perfects	#total	logpvF
show	2001	16202	290,8
report	554	3588	117,5
develop	771	6799	92,8
argue	684	6019	82,7
prove	409	2799	79,7
become	1120	12701	68,8
suggest	877	9763	57,1
change	437	3990	48,9
demonst.	283	2297	40,8
evolve	117	541	39,6
base	105	5780	-37,7
give	678	21020	-39,5
allow	103	6258	-46,0
involve	126	7018	-46,3
include	128	7340	-50,3
mean	119	7087	-50,8
follow	132	8474	-65,9
provide	300	14316	-75,6
know	129	10438	-100,9
do	629	44497	−∞

Variation between languages (EuroParl)

data: Europarl (year 2000), >50 occurrences, -be, have, got

English

lemma	#perfects	#total	logpvF
hear	303	965	84,6
mention	360	1568	60,3
table	247	928	53,9
receive	269	1075	52,9
show	300	1358	46,8
prove	157	577	35,8
say	936	7260	27,4
become	317	1852	27,2
reach	216	1125	25,0
lose	120	541	19,4
continue	53	1949	-28,9
apply	54	1974	-29,1
use	104	2880	-29,8
go	153	3789	-32,6
provide	96	2865	-33,1
include	84	2760	-35,9
consider	52	2191	-36,8
support	84	3052	-44,5
do	734	15177	-91,9
like	89	6813	-169,0

French

lemma	#perfects	#total	logpvF
voter	803	2030	261,9
dire	1632	8220	147,6
décider	442	1202	130,3
recevoir	341	905	104,1
déclarer	332	919	95,6
déposer	260	645	87,1
évoquer	306	941	75,5
souligner	444	1943	56,6
traire	99	170	51,9
accomplir	205	632	50,6
considérer	58	1811	-30,0
souhaiter	70	2518	-47,7
constituer	66	2608	-53,3
croire	55	2462	-55,4
penser	56	4627	-143,5
savoir	78	5488	-159,4
falloir	86	6083	-177,3
vouloir	164	9820	-266,1
pouvoir	641	18449	-285,1
devoir	229	20524	−∞

Dutch

lemma	#perfects	#total	logpvF
gebeuren	197	207	127,9
overtuig.	258	349	110,5
leiden	184	222	94,4
indienen	394	759	93,4
baseren	251	431	73,2
besluiten	246	473	58,7
bekennen	188	333	52,2
bereiden	205	417	44,1
verrichten	167	312	42,3
aantonen	117	178	42,1
geven	299	2652	-25,6
zien	156	1887	-37,4
zorgen	72	1252	-40,6
stellen	332	3283	-43,0
komen	458	4157	-43,3
maken	475	4280	-43,6
houden	195	2456	-52,2
blijven	63	1730	-80,8
vinden	108	2539	-107,2
gaan	137	5086	-284,4

Legend

lemma: canonical form
perfects: number of occurrences of the lemma in the PERFECT construction
total: total occurrences of the lemma (as a verb) in the corpus
logpvF: log-transformed *p*-value on the Fischer exact test:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{a+b+c+d}{a+c}}$$

	PERFECT	¬PERFECT
lemma	a	b
¬lemma	c	d

Conclusion

Across genres, we find inchoative verbs like "change", "become" and "evolve" to appear more frequent. Also, telic verbs like "win" and "lose" are attracted. Verbs of perception ("see", "hear") are more frequent in the spoken and fiction register, while we find reporting verbs ("show", "report") typically in the PERFECT construction in academic prose. Cognitive state verbs like "think" and "know" are typically repelled.

Across languages, the picture is not clear. We do find cognitive state verbs to be repelled in French as well, but we find "say" and "dire" as attracted in both French and English, which goes against our monolingual analysis.
All in all, the role of the lexicon in the PERFECT is not to be dismissed: not all verbs are equally willing to appear in a PERFECT, and these dissimilarities could well prove to be of interest for getting closer to a semantics of the PERFECT.

Netherlands Organisation for Scientific Research

This research is part of the "Time in Translation" project, funded by NWO under grant nr. 360-80-70.
Visit time-in-translation.hum.uu.nl for the latest on our project.