Underuse / Overuse Excel Add-In (V1.2)

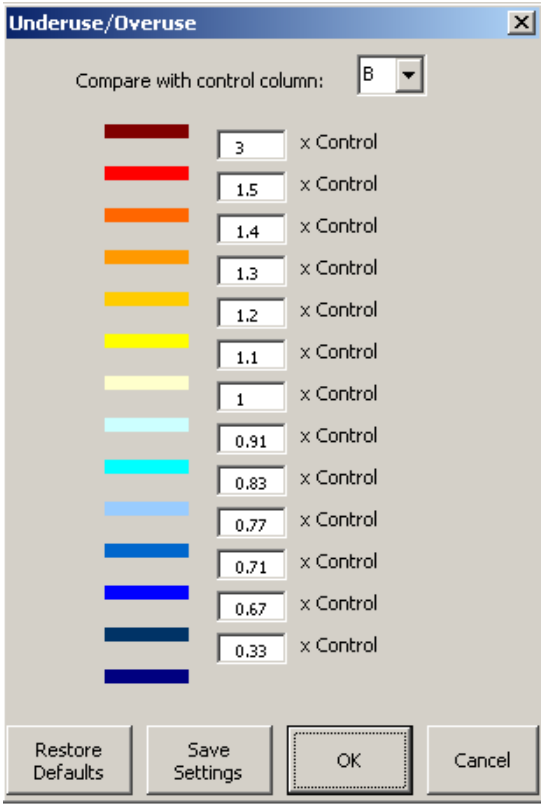
Author: Amir Zeldes

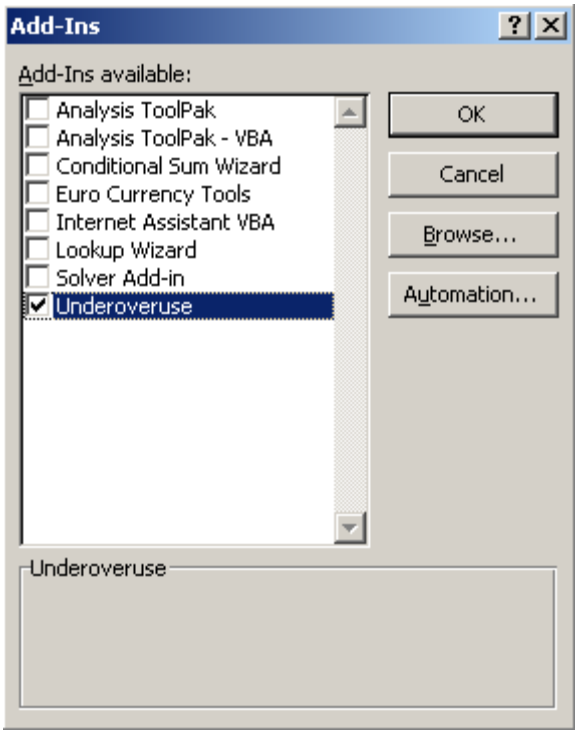
This add-in was written to allow linguists to visualize comparisons of word form or annotation frequencies between different subcorpora (e.g. according to metadata such as native language of the speaker/writer, genre of the text etc.), but it can probably be used to visualize comparisons between columns of numbers in Excel in general. Some of this functionality can also be reproduced using conditional formatting in newer versions of Excel.

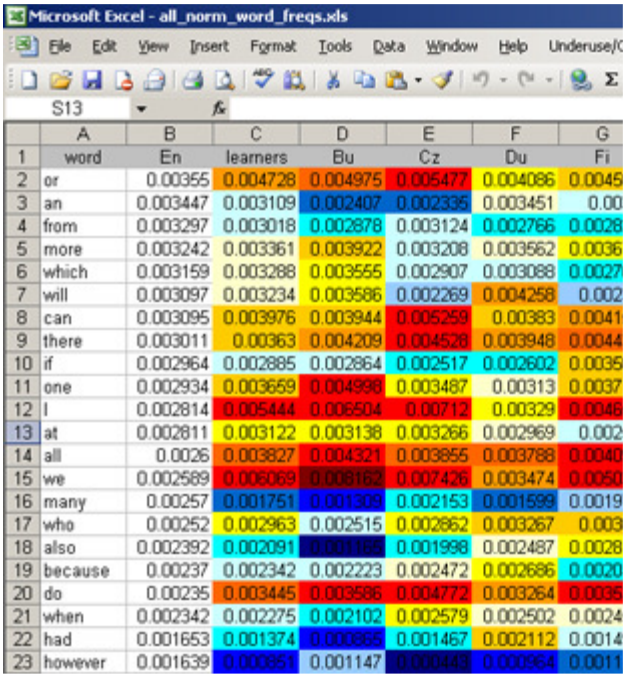
The add-in compares the numbers in each selected cell with the corresponding cell of a control column. Selected cells are then colored red if they have a higher value or blue for a lower value, with a deeper color the stronger the deviation from the control value. Below are some example images of the add-in at work.

To install the add-in, download it from https://github.com/amir-zeldes/xladdins and put the file in your add-in directory (e.g. `C:\documents and settings\YOUR_USERNAME\Application Data\Microsoft\AddIns`), then open Excel and go to Tools > Add-Ins and browse to select the file. Finally, check the add-in 'underoveruse'. The add-in can be uninstalled by unchecking this checkbox.

To use the add-in go to the new menu Underuse/Overuse and select 'Run analysis...'. This opens up a form allowing you to change the control column and the ranges of values corresponding to each color.







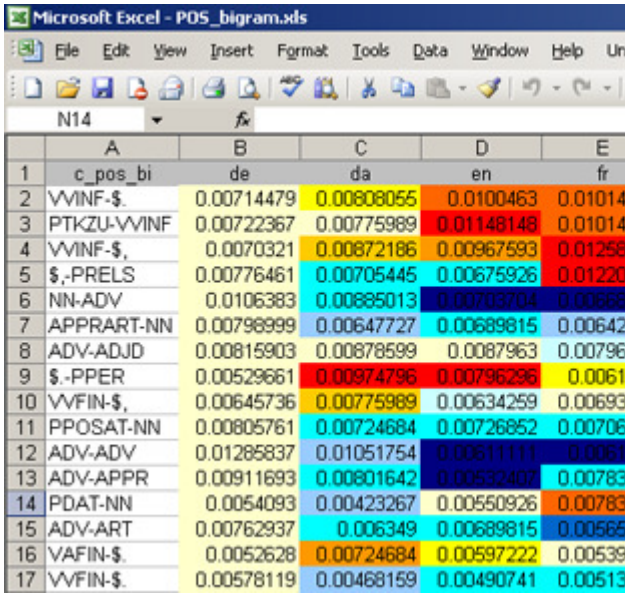
Microsoft Excel - all_norm_word_freqs.xls

File Edit View Insert Format Tools Data Window Help Underuse/C

S13

	A	B	C	D	E	F	G
1	word	En	learners	Bu	Cz	Du	Fi
2	or	0.00355	0.004728	0.004975	0.005477	0.004086	0.0045
3	an	0.003447	0.003109	0.002407	0.002335	0.003451	0.00
4	from	0.003297	0.003018	0.002878	0.003124	0.002766	0.0028
5	more	0.003242	0.003361	0.003922	0.003208	0.003562	0.0036
6	which	0.003159	0.003288	0.003555	0.002907	0.003088	0.0027
7	will	0.003097	0.003234	0.003586	0.002269	0.004258	0.002
8	can	0.003095	0.003976	0.003944	0.005259	0.00383	0.0041
9	there	0.003011	0.00363	0.004209	0.004528	0.003948	0.0044
10	if	0.002964	0.002885	0.002864	0.002517	0.002602	0.0035
11	one	0.002934	0.003659	0.004998	0.003487	0.00313	0.0037
12	I	0.002814	0.005444	0.006504	0.00712	0.00329	0.0046
13	at	0.002811	0.003122	0.003138	0.003266	0.002969	0.002
14	all	0.0026	0.003827	0.004321	0.003855	0.003788	0.0040
15	we	0.002589	0.006069	0.008162	0.007426	0.003474	0.0050
16	many	0.00257	0.001751	0.001309	0.002153	0.001593	0.0019
17	who	0.00252	0.002963	0.002515	0.002862	0.003267	0.003
18	also	0.002392	0.002091	0.001185	0.001998	0.002487	0.0028
19	because	0.00237	0.002342	0.002223	0.002472	0.002686	0.0020
20	do	0.00235	0.003445	0.003586	0.004772	0.003264	0.0035
21	when	0.002342	0.002275	0.002102	0.002579	0.002502	0.0024
22	had	0.001653	0.001374	0.000835	0.001467	0.002112	0.0014
23	however	0.001639	0.000651	0.001147	0.000443	0.000564	0.0011

Underuse and overuse of word forms in the International Corpus of Learner English by native language compared to the native English LOCNESS control corpus.



Microsoft Excel - POS_bigram.xls

File Edit View Insert Format Tools Data Window Help Un

N14

	A	B	C	D	E
1	c_pos_bi	de	da	en	fr
2	VVINF-$.	0.00714479	0.00808055	0.0100463	0.01014
3	PTKZU-VVINF	0.00722367	0.00775989	0.01148148	0.01014
4	VVINF-$,	0.0070321	0.00872186	0.00967593	0.01258
5	$,-PRELS	0.00776461	0.00705445	0.00675926	0.01220
6	NN-ADV	0.0106383	0.00885013	0.00703704	0.00668
7	APPRART-NN	0.00798999	0.00647727	0.00689815	0.00642
8	ADV-ADJD	0.00815903	0.00878599	0.0087963	0.00796
9	$.-PPER	0.00529661	0.00974796	0.00796296	0.0061
10	VVFIN-$,	0.00645736	0.00775989	0.00634259	0.00693
11	PPOSAT-NN	0.00805761	0.00724684	0.00726852	0.00709
12	ADV-ADV	0.01285837	0.01051754	0.00611111	0.0061
13	ADV-APPR	0.00911693	0.00801642	0.00532407	0.00783
14	PDAT-NN	0.0054093	0.00423267	0.00550926	0.00783
15	ADV-ART	0.00762937	0.006349	0.00689815	0.00565
16	VAFIN-$.	0.0052628	0.00724684	0.00597222	0.00539
17	VVFIN-$.	0.00578119	0.00468159	0.00490741	0.00513

Underuse and overuse of part-of-speech tag sequences in German learners by native language, compared to the German control group (de) in the Falko corpus . Completely blue rows show sequences learners find difficult independently of native language.