

Open research group

ESRC Centre for Corpus Approaches to Social Science

Lancaster University

Open Advanced Methods Research Group



12:00pm–12:50pm (UK time): R Scripts and



- Open space for ideas
- Corpus linguistics and statistics
- Research community

Topics



Wednesday 16 October 12.00pm - 12.50pm UK time, **Statistics and language analysis - #LancsBox KWIC**



Wednesday 30 October 12.00pm - 12.50pm UK time, **Collocations - #LancsBox GraphColl**



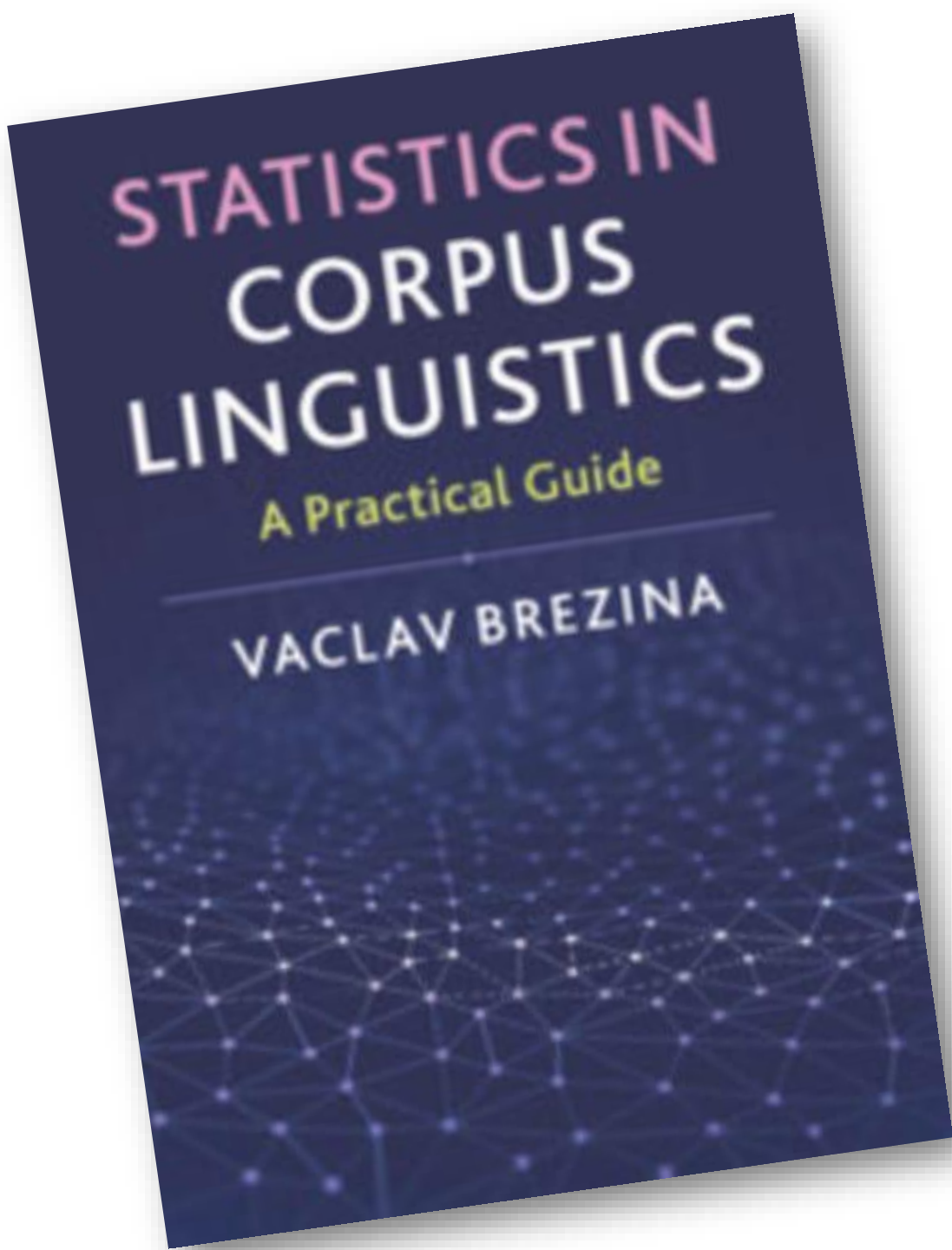
Wednesday 13 November 12.00pm - 12.50pm UK time, **Group comparison – Text tool**



Wednesday 27 November 12.00pm - 12.50pm UK time, **Wordlists and keywords - Words**

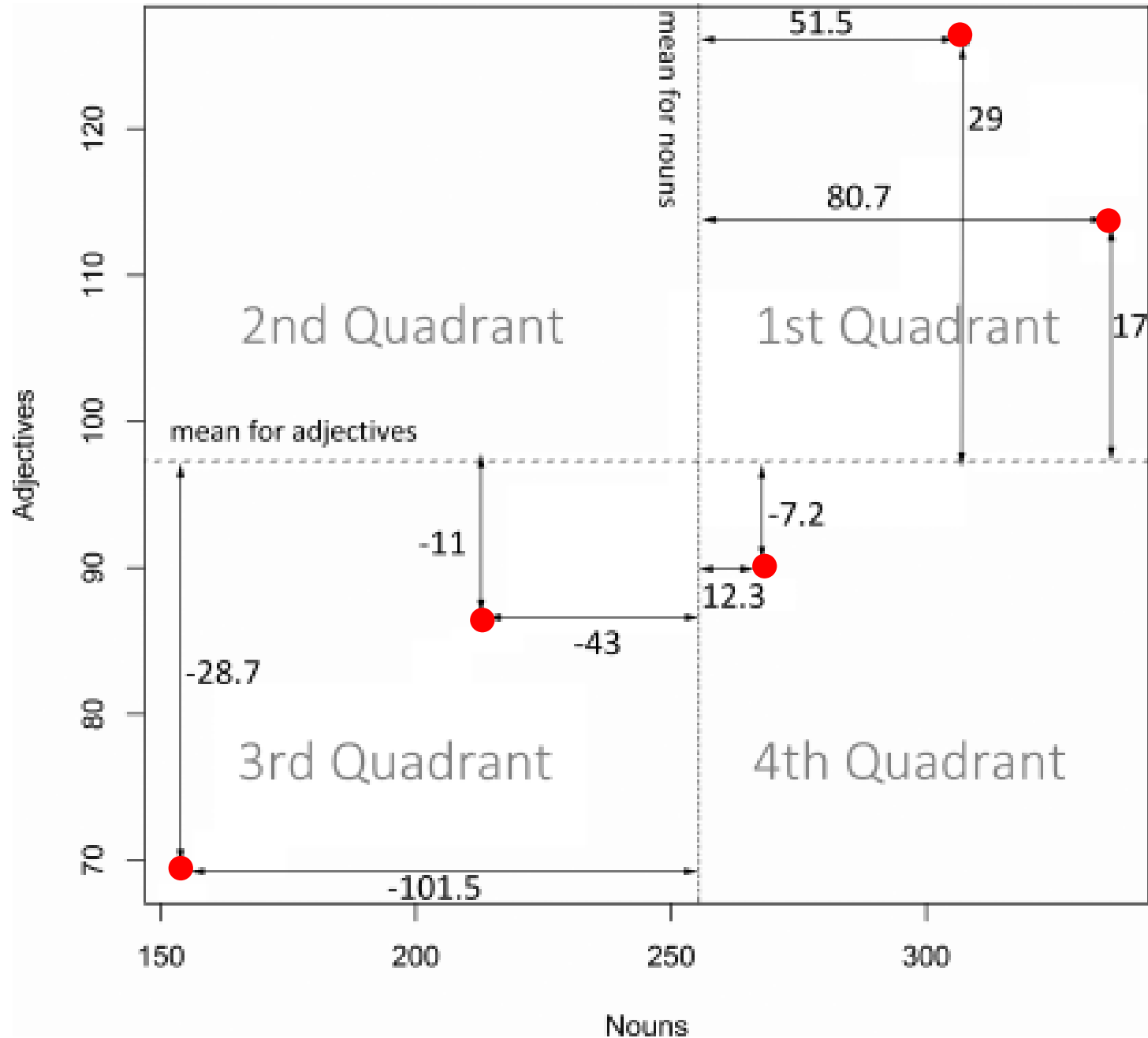


Wednesday 11 December 12.00pm - 12.50pm UK time, **R scripts and #LancsBox Wizard**



Brezina (2018)

“ By **virtue of being in a text together**, many linguistic variables **are related** in some way. For example, in English the relative frequency of adjectives in texts is related to the relative frequency of nouns because adjectives modify nouns and therefore typically occur together with them; at the same time, this doesn't mean that nouns do not occur without adjectives. [...] We can measure the relationships between two linguistic variables using a **technique called correlation**. Correlation measures whether two typically **ordinal or scale variables** are related by looking at the **extent to which they covary**. In other words, we are looking at whether, if one variable increases, the other variable increases, decreases or stays the same.



$$r = \frac{\text{covariance}}{SD_1 \times SD_2}$$

Correlation

BNC2014 2.0 CLAWS7

whole corpus

102M

new

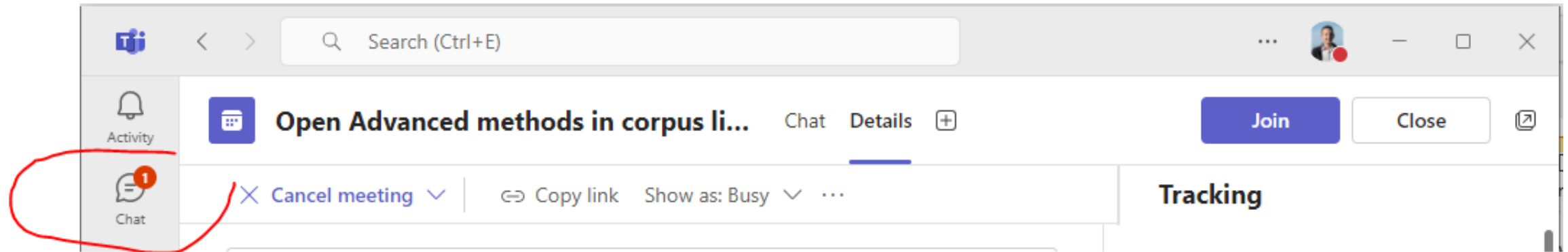
Hits: 129,581 (1,266.61)

Texts: 40,052/88,171




File	Left ▲	Node	Right	Text: genre
MagInv3149.xml	are forecast to reach \$18.7bn.	New	York-listed LiveNation (LYV) dominates t	magazines
NewSeGua2052.xml	\$21m - new 3. Paddington \$19.2m -	new	4. Taken 3 \$14m, \$62m	newspapers
MagPc2015.xml	form on Amazon for \$20 .	New!	It's Amazon Prime, even!	magazines
NewSeGua2052.xml	2. The Wedding Ringer \$21m -	new	3. Paddington \$19.2m - new 4.	newspapers
MagInv1033.xml	petroleum assets in a \$22bn	New	York listing, and has since	magazines
ElanSocFac33.xml	8 pm tickets are: \$29.50	new	york city up coming show ...!!!	elanguage
NewSeFit1591.xml	in currency rigging scandal \$2bn	New	York deal opens floodgates Global	newspapers
NewSeGua2332.xml	7. Woman in Gold: \$2m -	NEW	8. Kingsman: The Secret Service:	newspapers
NewSeTim24.xml	\$3.2bn Ebola nations seek \$3.2bn	New	York The presidents of Guinea,	newspapers
MagInv1294.xml	to raise a further \$3m.	New	shares have been listed at	magazines
NewSeGua2052.xml	\$4.2m, \$108.6m 10. Blackhat \$4m -	new	Figures from Box Office Mojo	newspapers

Feedback in chat – indicate if you'd like to be part of the group in Term 2 (January onwards)



Think about and discuss

1. Why does correlation not imply causation?
2. What is the role of correlation in linguistic research?