On association

Stefan Th. Gries
UC Santa Barbara & JLU Giessen
http://tinyurl.com/stgries

What is the relevance of association/contingency?

- Frequency of form & their dispersion are important, but so is association/contingency (w/ function) – especially for learning, recall Ellis (2006):
- "'[l]anguage learning can be viewed as a statistical process requiring the learner to acquire a set of likelihood-weighted associations between constructions & their functional/semantic interpretations"
- association quantifies what-if relations: what [happens] if [the context is like this]?
- "Learning, memory and perception are all affected by frequency, recency, and context of usage: […] The more times we experience conjunctions of features, the more they become associated in our minds and the more these subsequently affect perception and categorization" (Ellis, Römer, & O'Donnell 2016:45f.)
- in other words, association → correlation, → how much does knowing X help you predict Y?
- that's why "human learning is to all intents and purposes perfectly calibrated with normative statistical measures of contingency like r, $\chi 2$ and ΔP" (Ellis 2006:7)

How is association usually measured?

- For every, say, word co-occurring w/ cx 1, a 2x2 table is created, from which many association measures (AMs) can be computed easily
- then, the words can be ranked according to their association to cx1

	c 1	other	Sum
w1	**80**	*200*	**280**
other	1000	...	...
Sum	**1080**	...	sum

	c 1	other	Sum
w2	**60**	*310*	**370**
other	1020	...	...
Sum	**1080**	...	sum

	c 1	other	Sum
w3	**40**	*420*	**460**
other	1040	...	...
Sum	**1080**	...	sum

- there has been a lot of discussion about which AM is 'best' but some of this is purely academic – most widely-used measures can be derived from logistic regression
  - G^2, odds ratio, log odds ratio, *MI*, *t*, *z*, …
  - other can't but are still very highly correlated with some of the above: p_{FYE}, X^2, …

How should association be measured?

- The following considerations are relevant to choosing an AM
  - symmetry: is the AM supposed to be symmetric or not?
    - nearly all AMs are: p_{FYE}, *LLR*, X^2, *MI*, *t*, *z*, log odds ratio …
    - some are not: $p(y|x)$, ΔP, …
  - metric type: +effect −freq. vs +effect +freq
    - the former: log odds ratio, the asymmetric ones above, …
    - the latter: p_{FYE}, *LLR*, X^2, …
  - frequency information: token vs token+type frequency
    - the former: all but one
    - the latter: lexical gravity *G*
- probably best settings in an ideal world:
  - symmetry: no
  - metric type: +effect
  - (frequency: token+type)
- ideally dispersion would be included in some way
- let me suggest two measures for your consideration
  - log odds ratio
  - ΔP

Why am I suggesting these two & how is the log odds ratio computed?

- The log odds ratio
  - symmetric, +effect -frequency
  - you compute
    - the odds of one outcome in one condition/context
    - the odds of the same outcome in the other condition/context
    - you divide them and log
    - maybe add 0.5 to all cells first to help w/ 0s

G^2=762.2		*as*-predicat		
		yes		
regard	yes	80	19	99
	no	607	137958	138565
	Totals	687	137977	138664

odds of *r* when *a*	0.1318
odds of *r* when not *a*	0.0001
odds ratio	956.9618
log odds ratio	6.8638

G^2=7622		*as*-predicative		
		yes	no	Totals
regard	yes	800	190	990
	no	6070	1379580	1385650
	Totals	6870	1379770	1386640

odds of *r* when *a*	0.1318		
odds of *r* when not *a*		0.0001	
odds ratio			956.9618
log odds ratio			6.8638

Why am I suggesting these two & how is $\Delta P_{\mathrm{c \to r}}$ computed?

- ΔP
  - asymmetric, +effect – frequency
  - you compute
    - the % of the outcome of interest in one condition/context
    - the % of the outcome of interest in the other cond./context
    - you subtract the former from the latter
    - obviously, this can then be done in either direction

G^2=762.2		*as*-predica		
		yes		
regard	yes	80	19	99
	no	607	137958	138565
	Totals	687	137977	138664
% of *r* when *a*		0.1164		
% of *r* when not *a*			0.0001	
Delta *P*				0.1163

G^2=7622		*as*-predicative		
		yes	no	Totals
regard	yes	800	190	990
	no	6070	1379580	1385650
	Totals	6870	1379770	1386640
% of *r* when *a*		0.1164		
% of *r* when not *a*			0.0001	
Delta *P*				0.1163

Why am I suggesting these two & how is $\Delta P_{r \to c}$ computed?

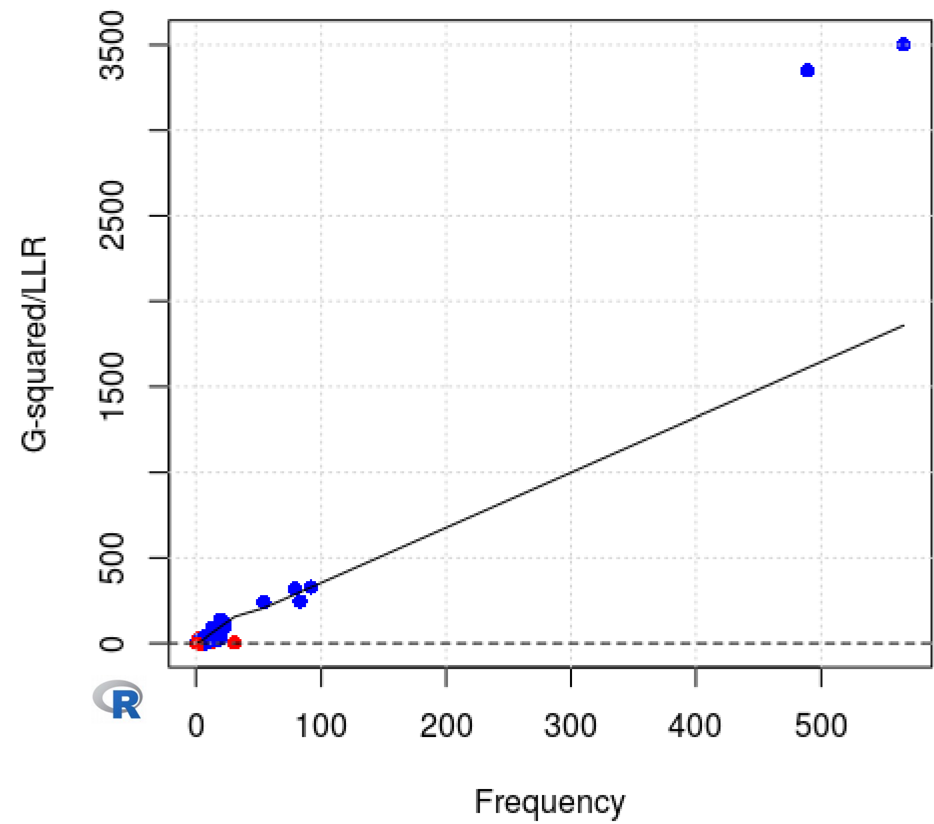
- ΔP
  - asymmetric, +effect – frequency
  - you compute
    - the % of the outcome of interest in one condition/context
    - the % of the outcome of interest in the other cond./context
    - you subtract the former from the latter
    - obviously, this can then be done in either direction

G^2=762.2		*as*-predicat		
		yes		
regard	yes	80	19	99
	no	607	137958	138565
	Totals	687	137977	138664
% of *a* when *r*		0.8081		
% of *a* when not *r*			0.0044	
Delta *P*				0.8037

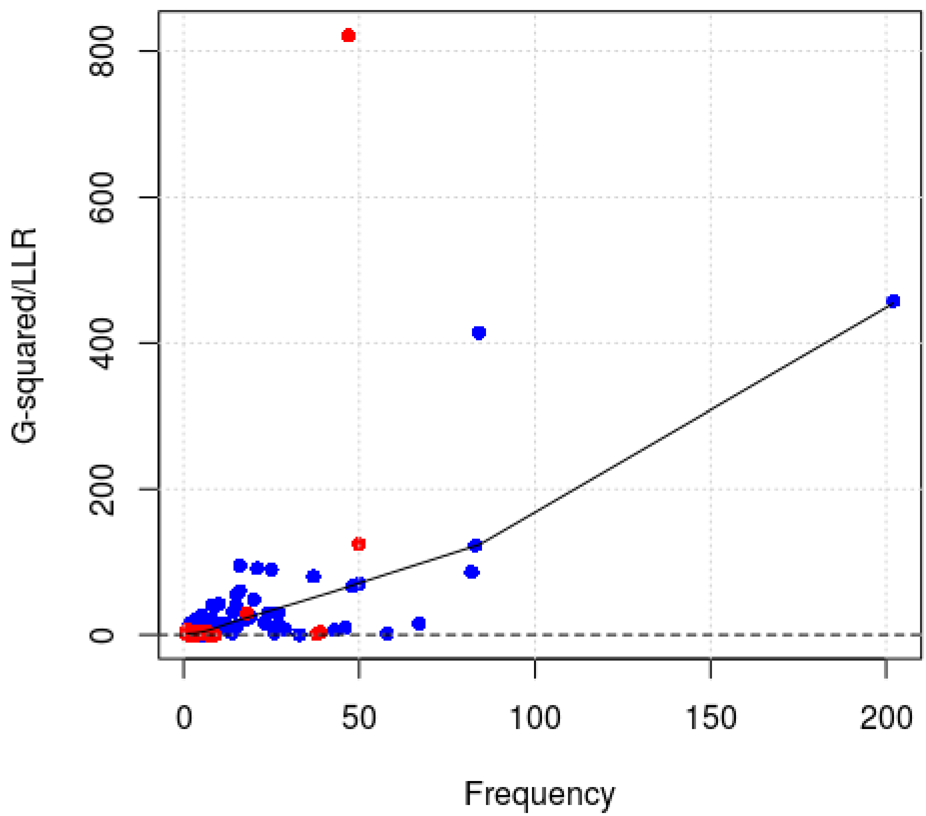
G^2=7622		*as*-predicative		
		yes	no	Totals
regard	yes	800	190	990
	no	6070	1379580	1385650
	Totals	6870	1379770	1386640
% of *a* when *r*		0.8081		
% of *a* when not *r*			0.0044	
Delta *P*				0.8037

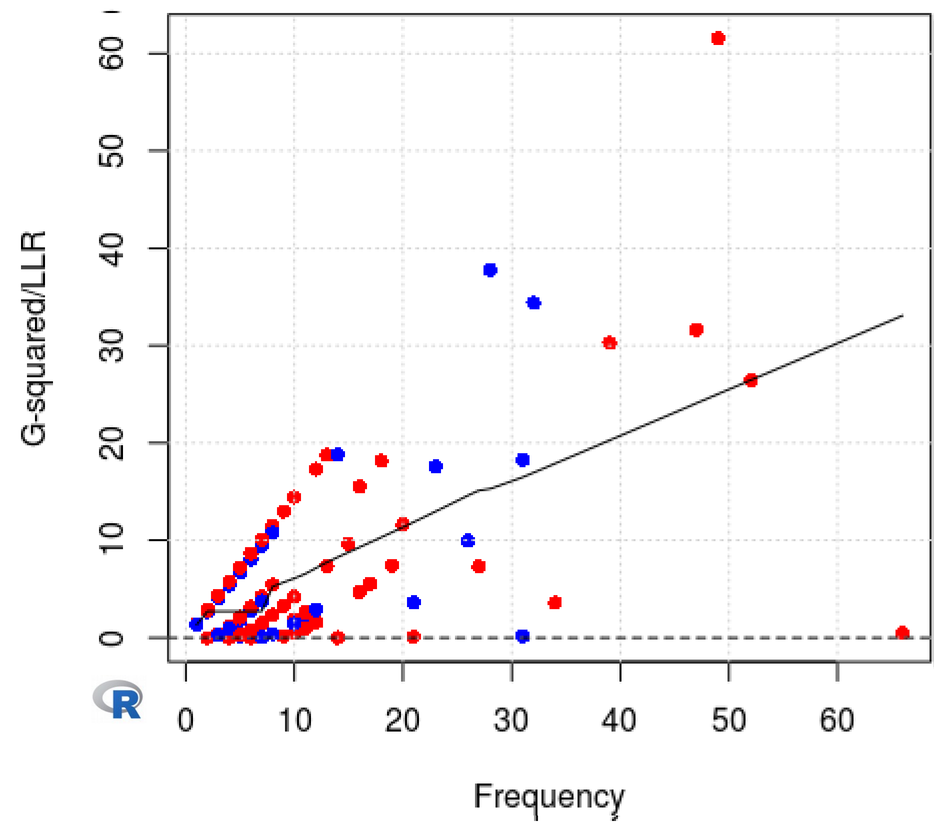
Note what they share

- The log odds ratio and Δ*P* are not affected if the frequencies go up much (eg by an order of magnitude, as exemplified above)
- to reiterate: many other AMs do not behave that way: they react to effect size & frequency
- here's the most widely-used one: G^2
  - in ditransitives
    - The cat brought her a mouse
  - in imperatives
    - Kill the mouse!
  - in verb-particle constructions
    - He picked up the book vs He picked the book up



Note what they share





Note what they don't share

- The log odds ratio is symmetric, Δ*P* is not, ie
  - the former cannot distinguish these collocations,
  - the latter can
    - *of←course, at←least, for←instance, in←vitro, de←facto, …*
    - *according→to, upside→down, instead→of, ipso→facto, …*
    - *Sinn↔Fein, bona↔fide, …*
- in the spoken part of the BNC, all of these have
  - G^2>178
  - log odds ratio>5
- but why would such learned connections would be (as) symmetric? (Trautschold 1883, Cattell 1887)
- in fact, mismatches between corpus and psycho-linguistic data might be in part due to overlooking the directionality of collocations

But is Δ*P* really worth it?

- Given how Δ*P* is computed, it is
  - correlated much w/ transitional probability *p*(x|y)
  - only natural to ask whether it's different enough from *p*(x|y) to even make a difference
- Schneider (to appear): yes
  - data: Switchboard NXT 2008 (642 phone conversations)
  - dependent variable: hesitation placement in PPs
  - predictors: *a*, Δ*P*→, TP→, Δ*P*←, TP←, *MI*, lex. grav. *G*
  - statistical analysis: party::cforest
  - results: many different results for the three kinds of PPs, but
    - "it is mostly Δ*P* which outperforms transitional probability"
    - this is true for both forward-directed measures and backwards at phrase boundaries
    - ← measures are good predictors of collocation status when w1 = function word & w2 = content word
    - other major finding: lexical gravity *G* does very well!
- Dunn (2018): tuples of different Δ*P*s are useful

Collostructional analysis

- Collostructional analysis (CA) is an method based on the maybe most fundamental corpus linguistic assumptions: the distributional hypothesis
  - "[i]f we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution" (Harris 1970:785f.)
- CA is a straightforward extension of …
  - of collocations: co-occurrence of words/lexical units
  - to (one sense of) colligation: co-occurrence
    - of words
    - and patterns/constructions

Collostructional analysis

- CA is a 'family' of 3 methods
  - collexeme analysis
    - co-occurrence of each of *n* words
    - in/with 1 construction
  - distinctive collexeme analysis
    - co-occurrence of each of *n* words
    - in/with 2 (or more) constructions
  - co-varying collexeme analysis
    - co-occurrence of words in 2 slots of 1 construction
- for each 2x2 table, one computes an assoc. measure to see
  - which words like cx 1
  - which words prefer which cx
  - which words go together in cx
  - based on the (ranks of) sorted association measures (AMs)
- AMs most widely used:
  - $p_{\text{Fisher-Yates exact}}$ & G^2/LLR

	cx 1: y	cx 1: n	Σ
word 1: y	**80**	200	**280**
word 1: n	1000	…	…
Σ	**1080**	…	Σ

&

	cx 1: y	cx 1: n	Σ
word 2: y	**60**	310	**370**
word 2: n	1020	…	…
Σ	**1080**	…	**Σ**

	cx 1	cx 2	Σ
word 1: y	**150**	**80**	230
word 1: n	930	720	1650
Σ	**1080**	**800**	1880

&

	cx 1	cx 2	Σ
word 2: y	**60**	**310**	370
word 2: n	1020	490	1510
Σ	**1080**	**800**	1880

	word 2: y	word2: n	Σ
word 1: y	**40**	240	**280**
word 1: n	330	470	800
Σ	**370**	**710**	**1080**

&

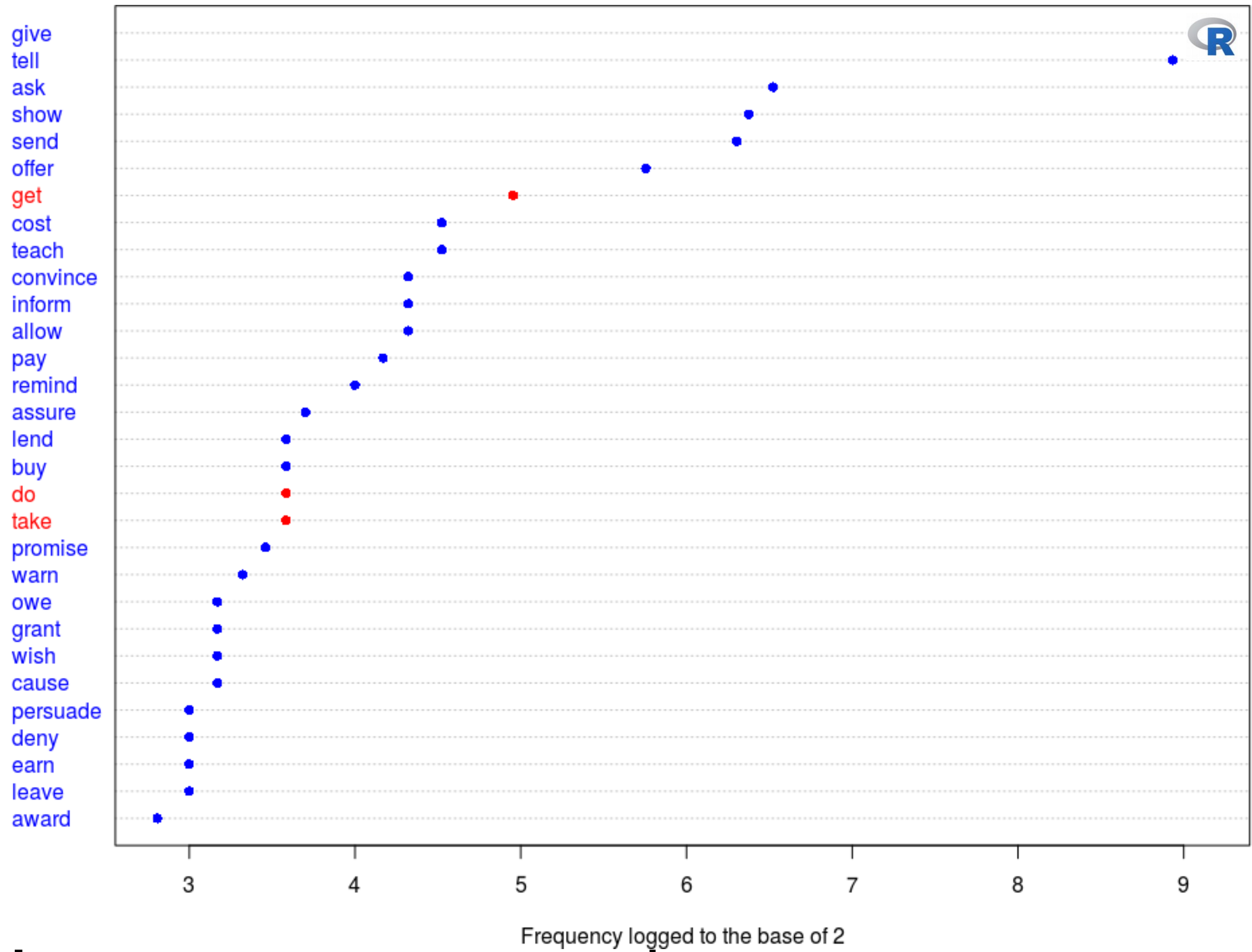
	word 3: y	word3: n	Σ
word 1: y	**20**	260	**280**
word 1: n	180	620	800
Σ	**200**	**880**	**1080**

Addressing at least some of the above-mentioned problems

- Let's look at a few examples of CA, where we
  - keep frequency and contingency separate
    - using (log2) of the observed co-occurrence frequency of verbs & a construction
    - using an association measure that doesn't include the observed co-occurrence frequency
  - add dispersion to the mix
    - computing the dispersion of, say, verbs in the construction against the distribution of verbs in general
- examples
  - collexeme analysis: ditransitive
  - collexeme analysis: imperative
  - distinctive collexeme anal.: verb-particle constructions
- things not to be discussed (much) here:
  - keeping directions of association separate
    - we could use ΔPs as AMs (ΔPvfromc & ΔPcfromv)
  - no entropy
  - no polysemy

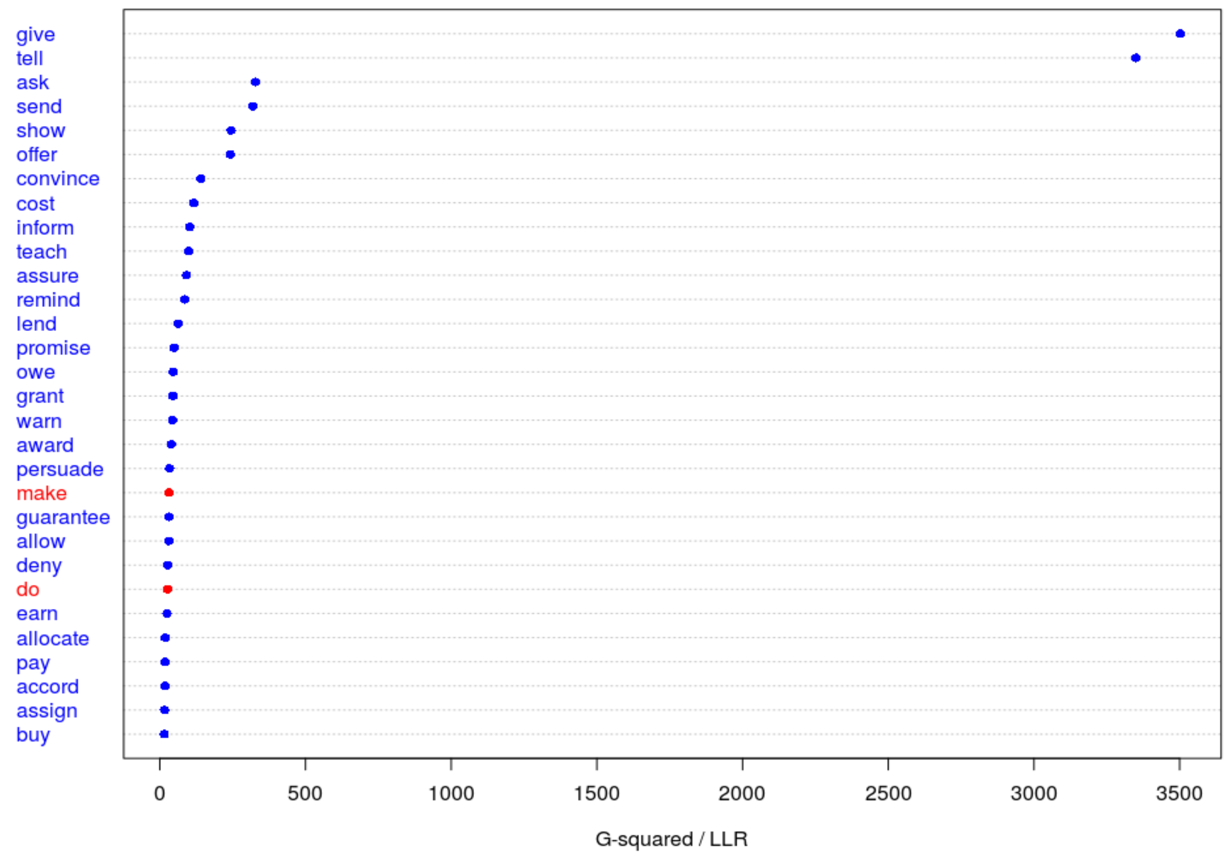
A collexeme analysis of the ditransitive: steps 1-3

- Data: ICE-GB
  - 1820 ditransitives
  - 88 different verbs
    - with freqs between 1 and 566
- step 1: frequency
  - surprisingly good: *give*, *tell*, *ask*, *show*, *send*, *offer*, but then *get* …
- step 2: association w/ G^2/LLR, confla-ting freq & effect
  - also good but not that distinctive
- step 3: keeping frequency & association separate: much more informative (esp. if you want to be cogn. realistic)



A collexeme analysis of the ditransitive: steps 1-3

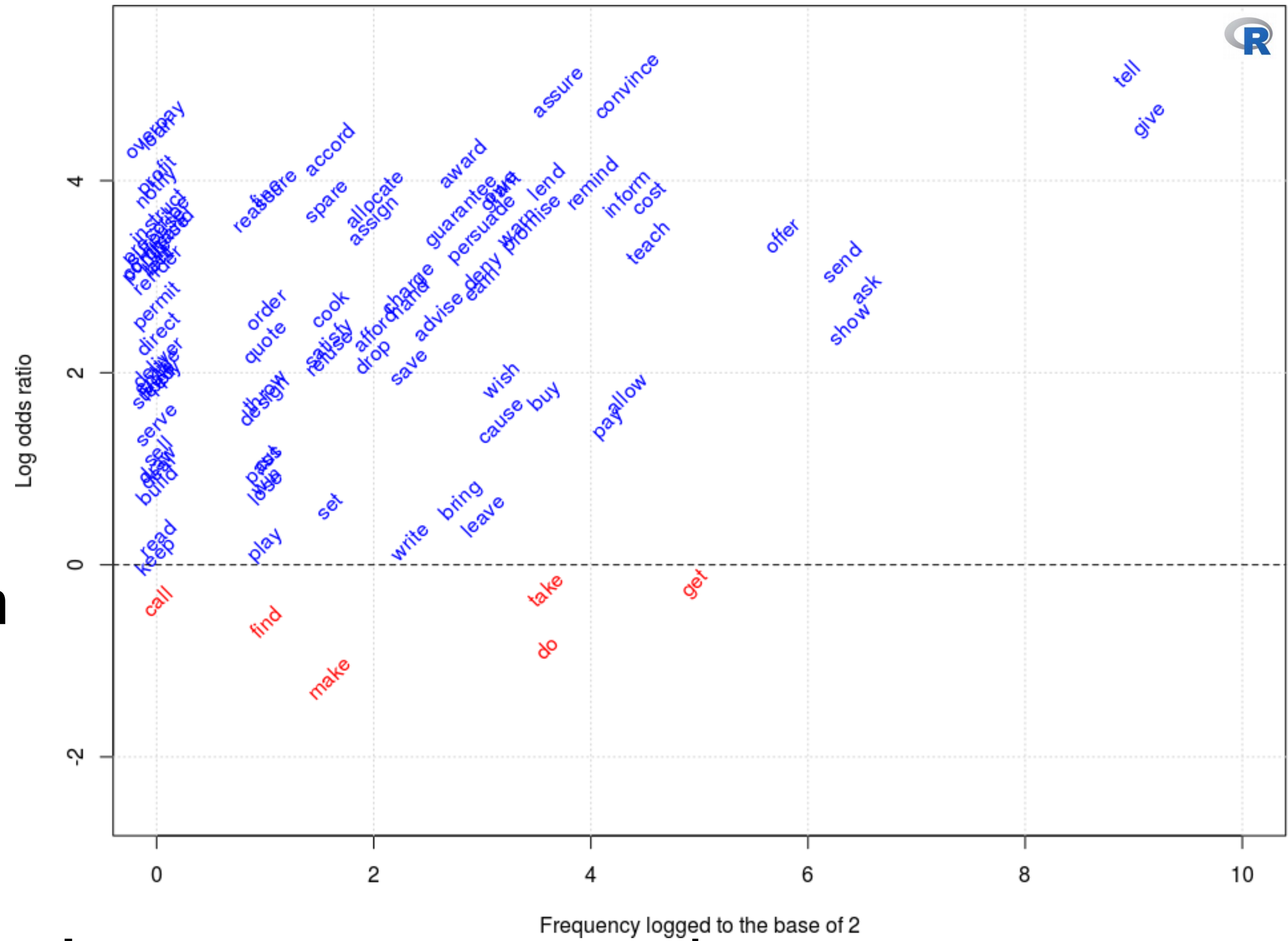
- Data: ICE-GB
  - 1820 ditransitives
  - 88 different verbs
    - with freqs between 1 and 566
- step 1: frequency
  - surprisingly good: *give*, *tell*, *ask*, *show*, *send*, *offer*, but then *get* …
- step 2: association w/ G^2/LLR, confla-ting freq & effect
  - also good but not that distinctive



- step 3: keeping frequency & association separate: much more informative (esp. if you want to be cogn. realistic)

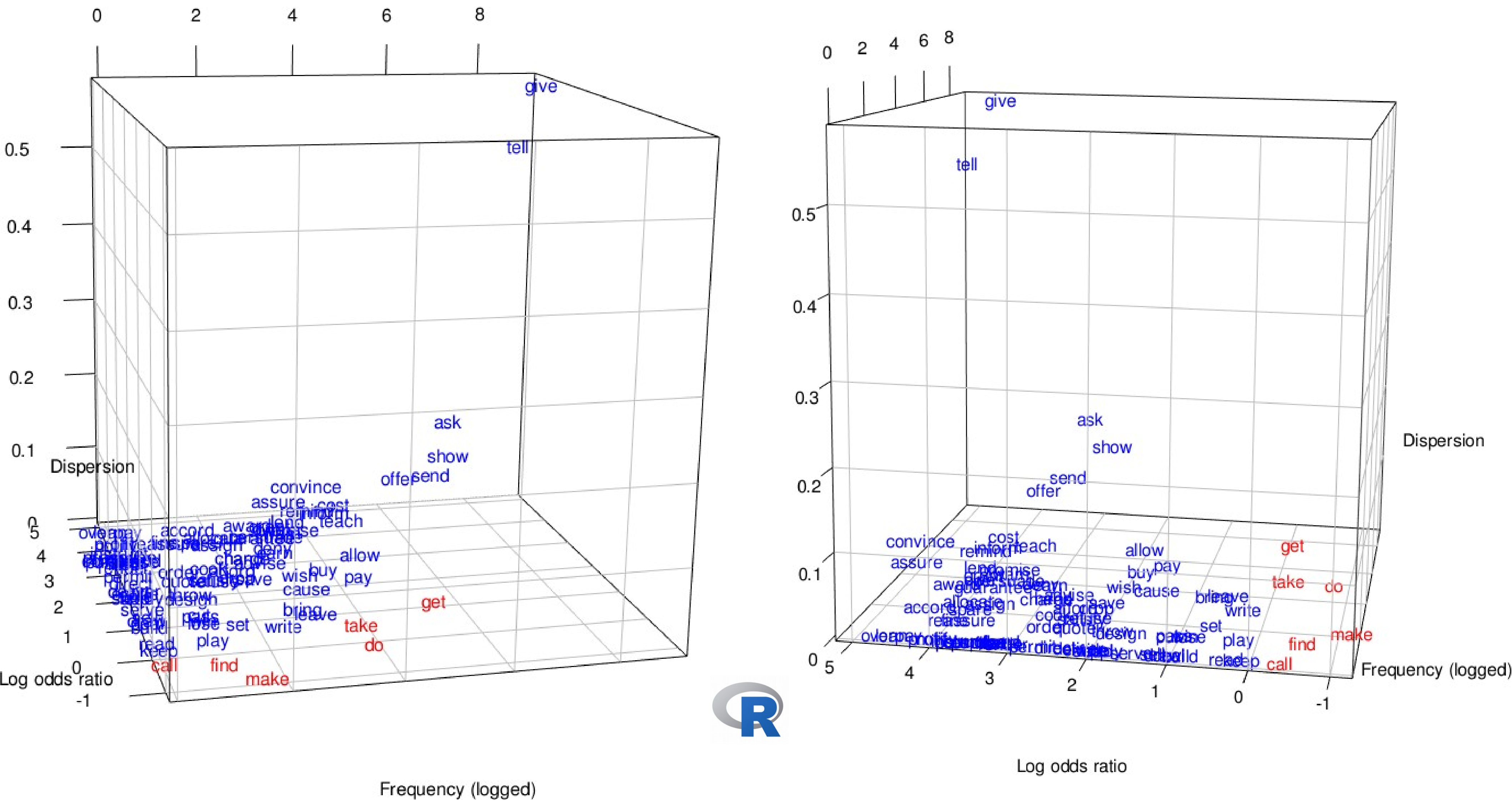
A collexeme analysis of the ditransitive: steps 1-3

- Data: ICE-GB
  - 1820 ditransitives
  - 88 different verbs
    - with freqs between 1 and 566
- step 1: frequency
  - surprisingly good: *give*, *tell*, *ask*, *show*, *send*, *offer*, but then *get* …

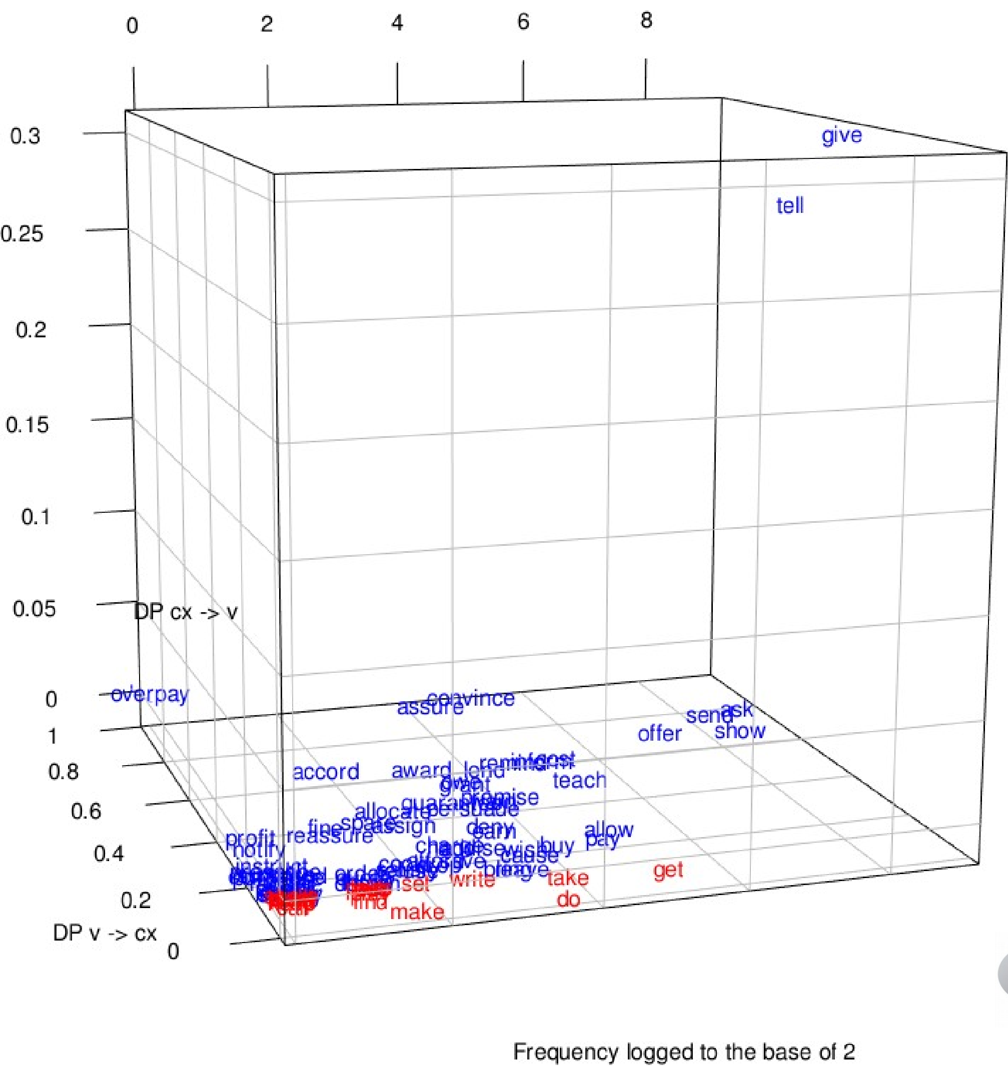


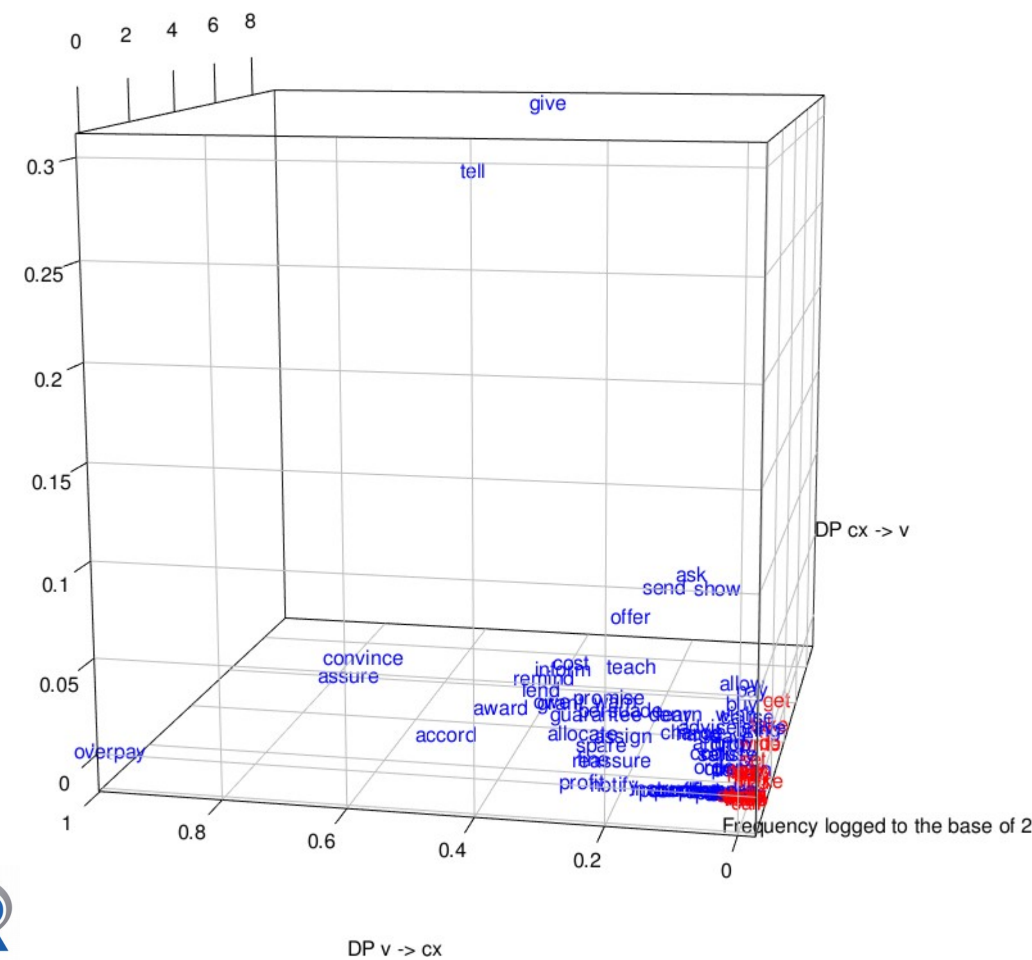
- step 2: association w/ G^2/LLR, conflating freq & effect
  - also good but not that distinctive
- step 3: keeping frequency & association separate: much more informative (esp. if you want to be cogn. realistic)

A collexeme analysis of the ditransitive: step 4



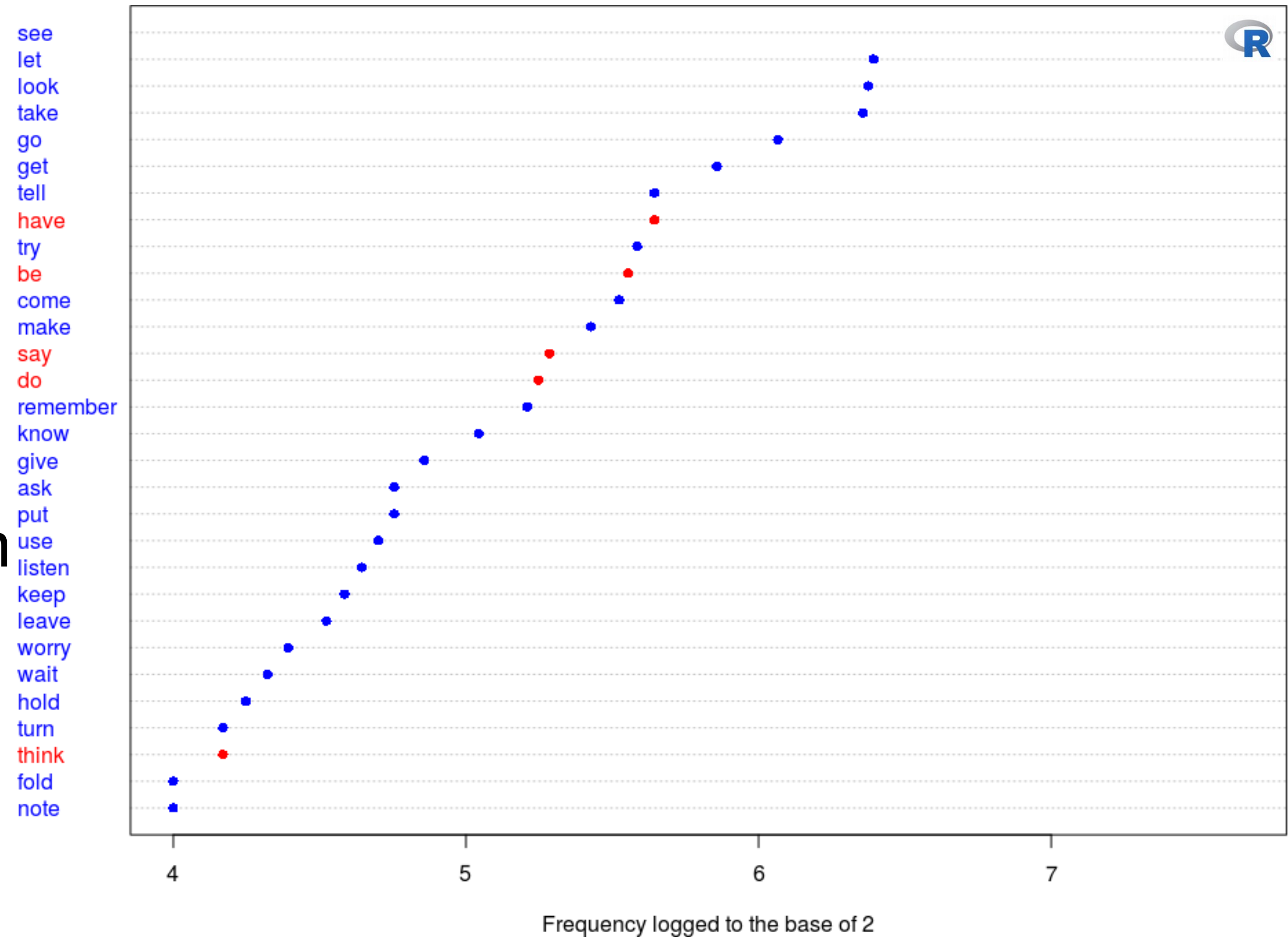
A collexeme analysis of the ditransitive: step 5





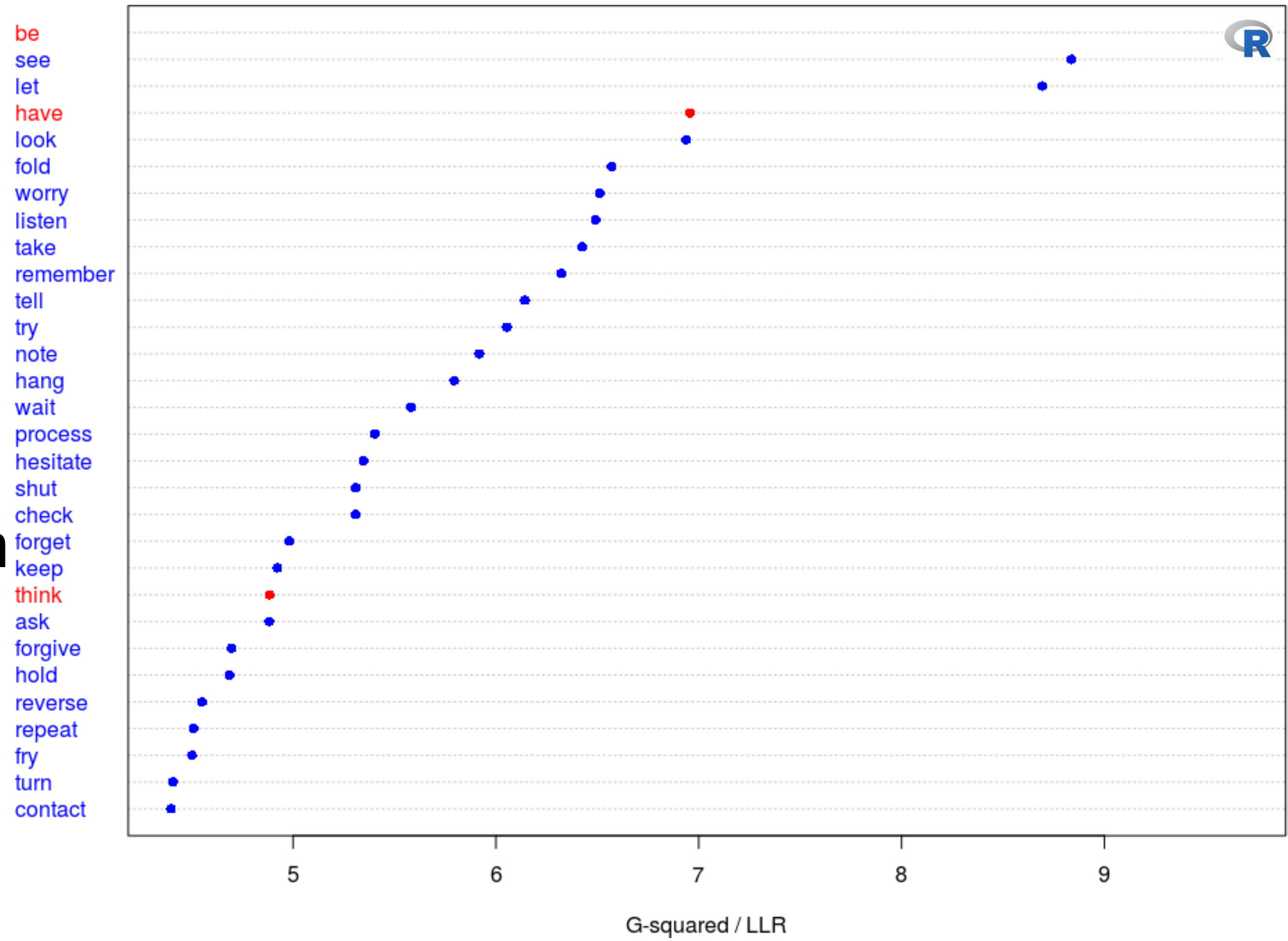
A collexeme analysis of the imperative: steps 1–3

- Data: ICE-GB
  - 2083 imperatives
  - 314 different verbs
    - with freqs between 1 and 202
- step 1: frequency
  - surprisingly good: *see*, *let*, *look*, *take*, … *have*, …, *be*
- step 2: association w/ G^2/LLR, confla-ting freq & effect
  - also good, but
    - huge impact of freq on repelled verbs
    - *fold* & *process*?
- step 3: keeping frequency & association separate: much more informative



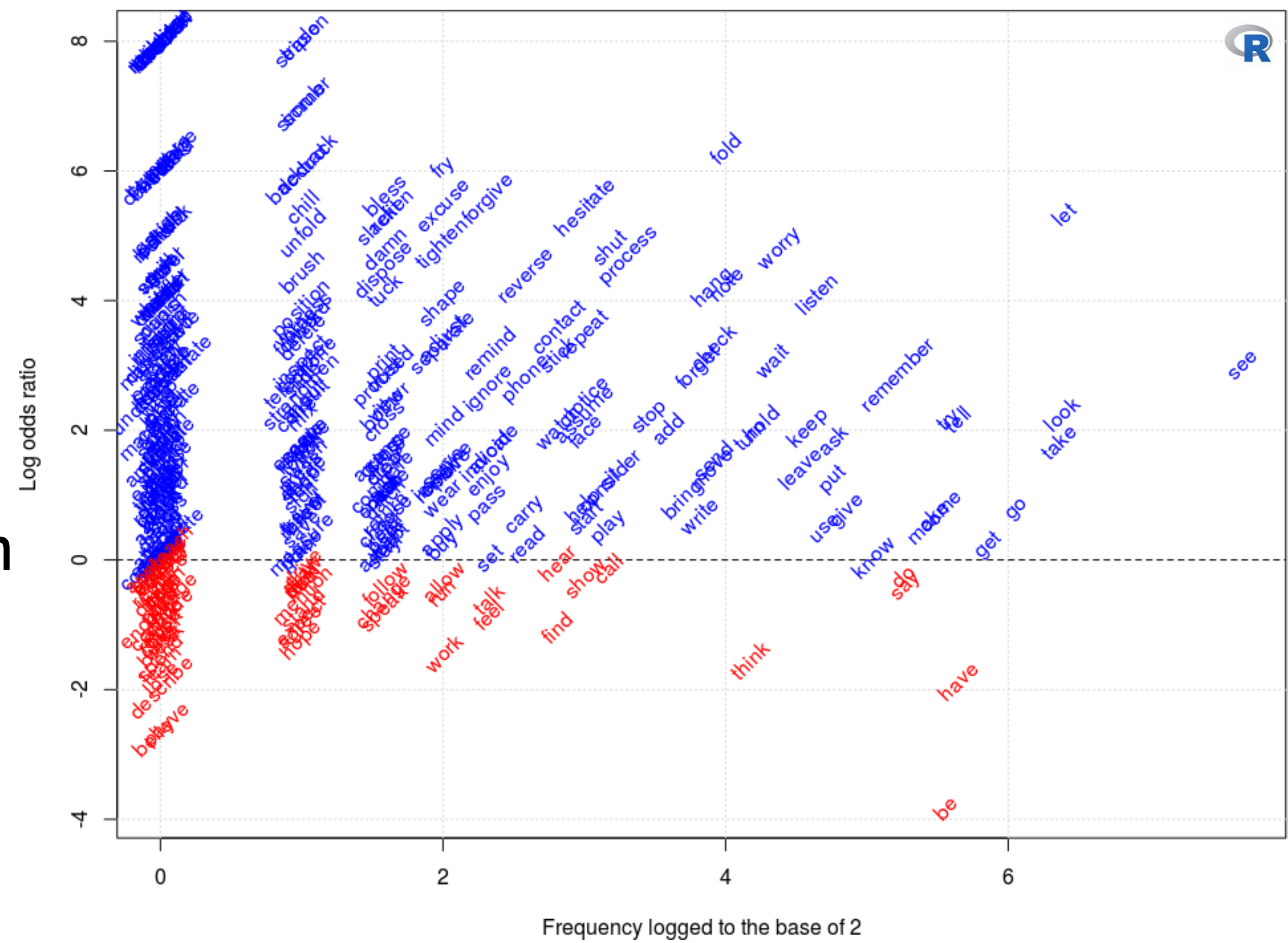
A collexeme analysis of the imperative: steps 1-3

- Data: ICE-GB
  - 2083 imperatives
  - 314 different verbs
    - with freqs between 1 and 202
- step 1: frequency
  - surprisingly good: *see*, *let*, *look*, *take*, … *have*, …, *be*
- step 2: association w/ G^2/LLR, confla-ting freq & effect
  - also good, but
    - huge impact of freq on repelled verbs
    - *fold* & *process*?
- step 3: keeping frequency & association separate: much more informative



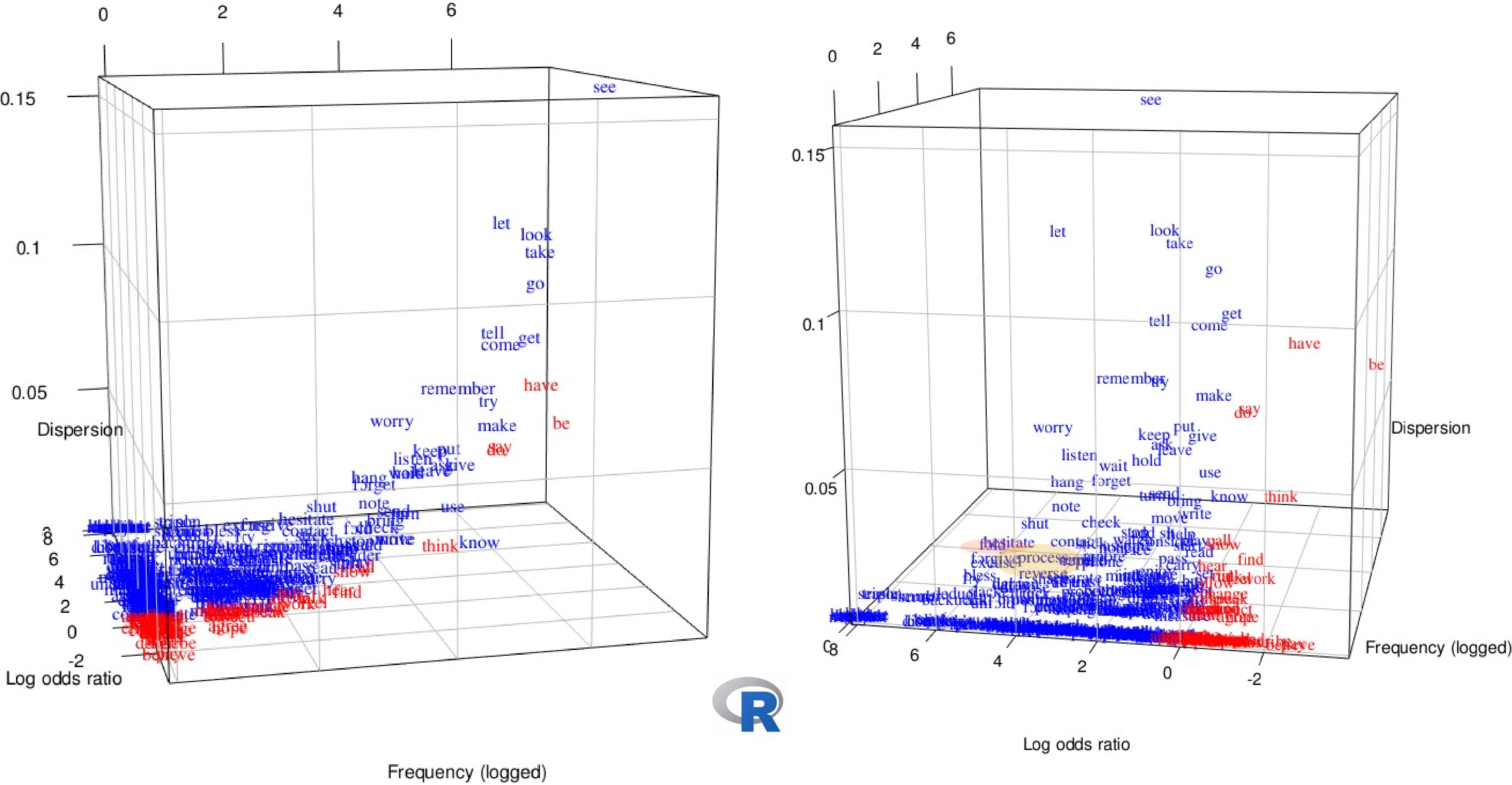
A collexeme analysis of the imperative: steps 1–3

- Data: ICE-GB
  - 2083 imperatives
  - 314 different verbs
    - with freqs between 1 and 202
- step 1: frequency
  - surprisingly good: *see*, *let*, *look*, *take*, … *have*, …, *be*
- step 2: association w/ G^2/LLR, conflating freq & effect
  - also good, but
    - huge impact of freq on repelled verbs
    - *fold* & *process*?

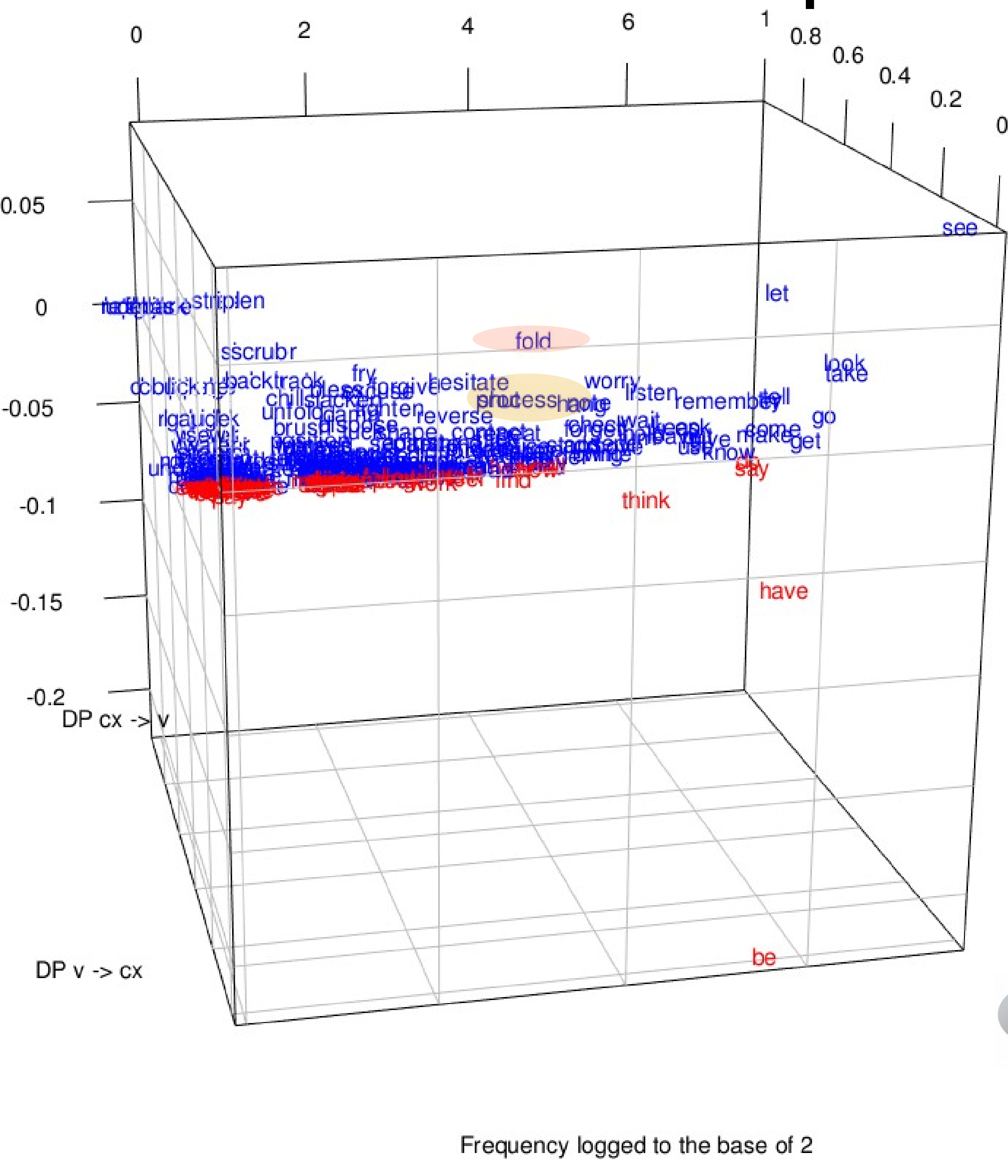


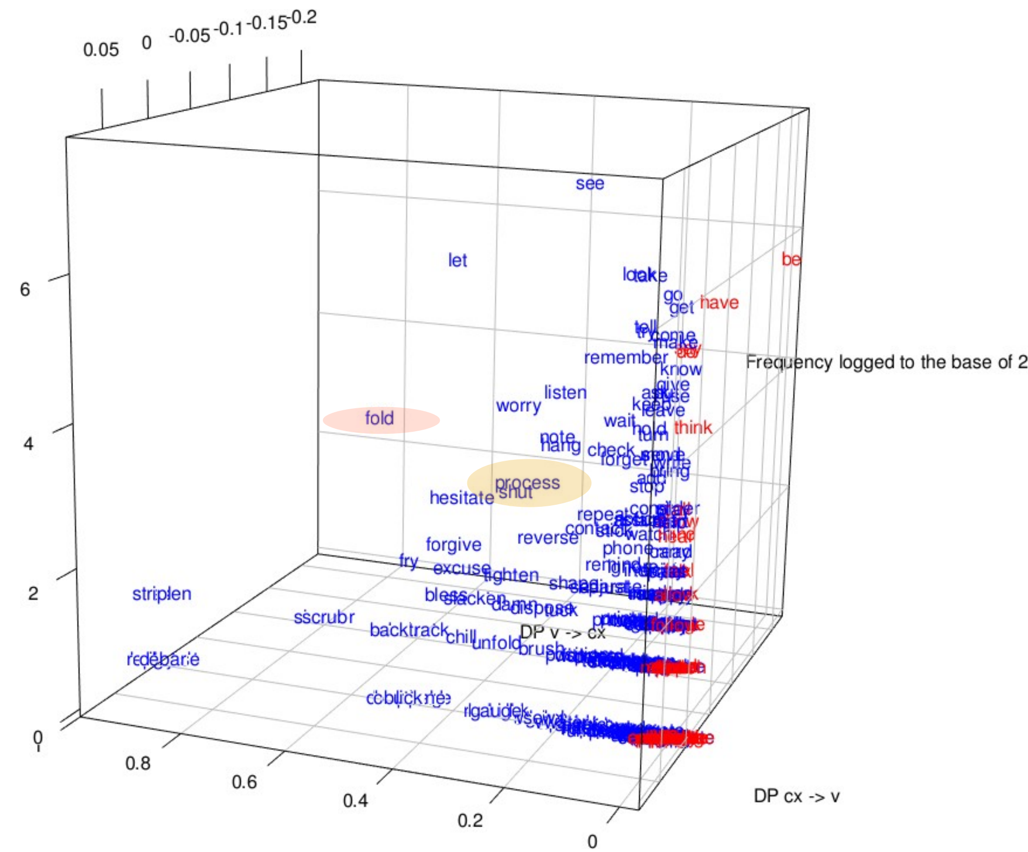
- step 3: keeping frequency & association separate: much more informative

A collexeme analysis of the imperative: step 4



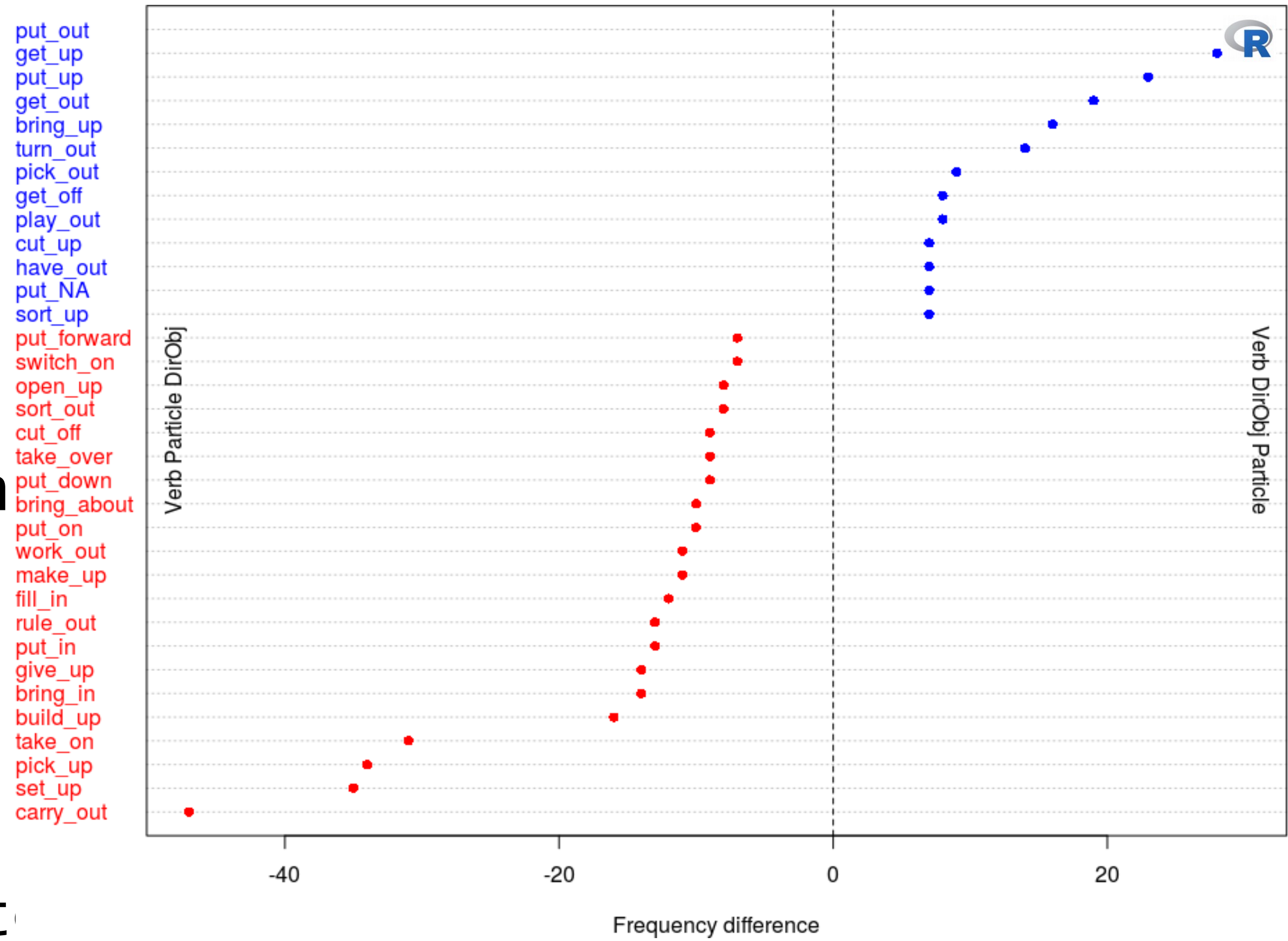
A collexeme analysis of the imperative: step 5





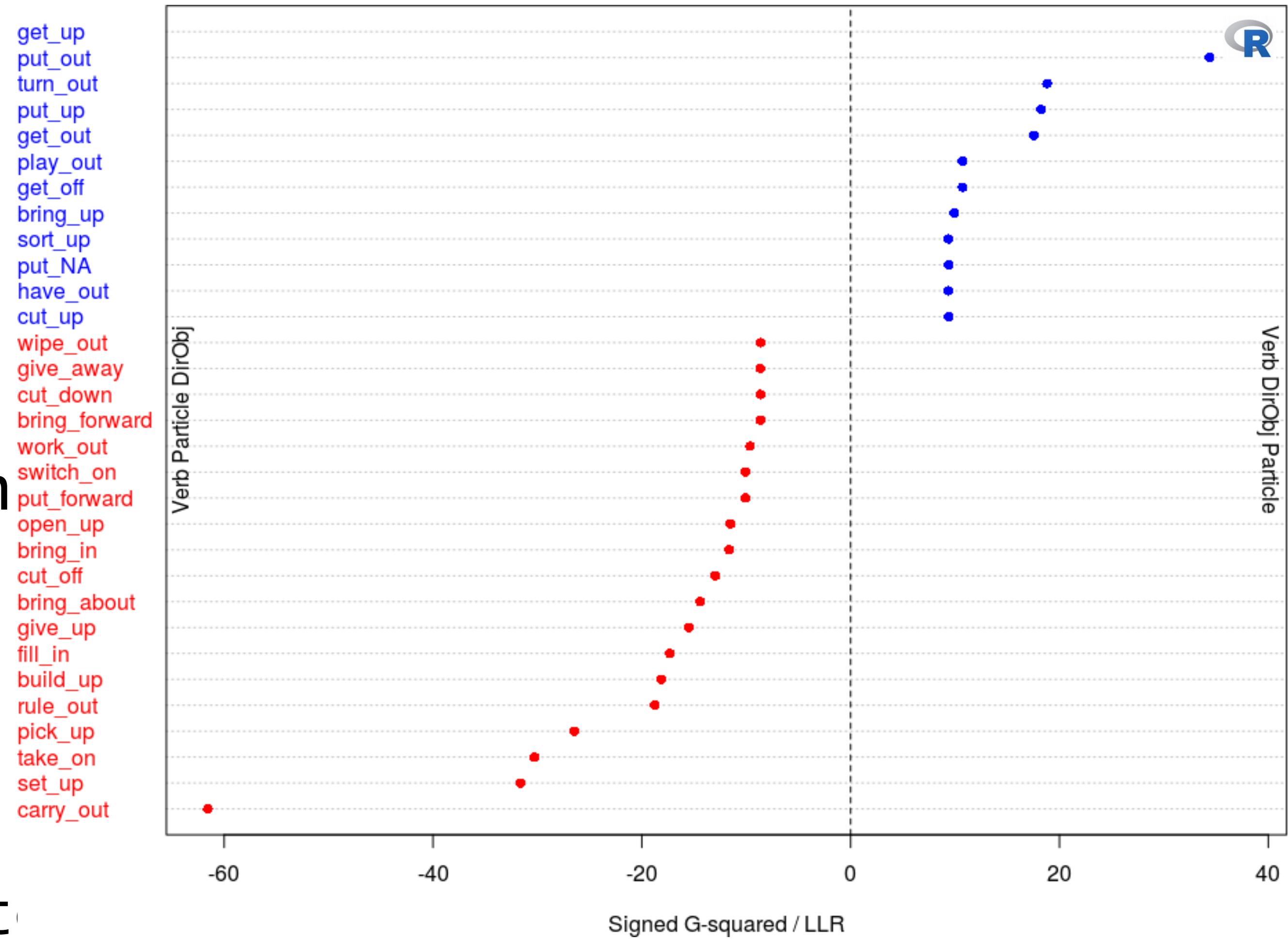
A distinctive collexeme analysis of verb-particle constructions: steps 1-3

- Data: ICE-GB
  - 1164 VPCs
  - 835 different verbs
    - with freqs between 1 and 31
- step 1: frequency
  - hard to evaluate, seems reasonable(?)
- step 2: association w/ G^2/LLR, conflating freq & effect
  - seems very similar w/ ranking changes
- step 3: keeping frequency & association separat much more informative



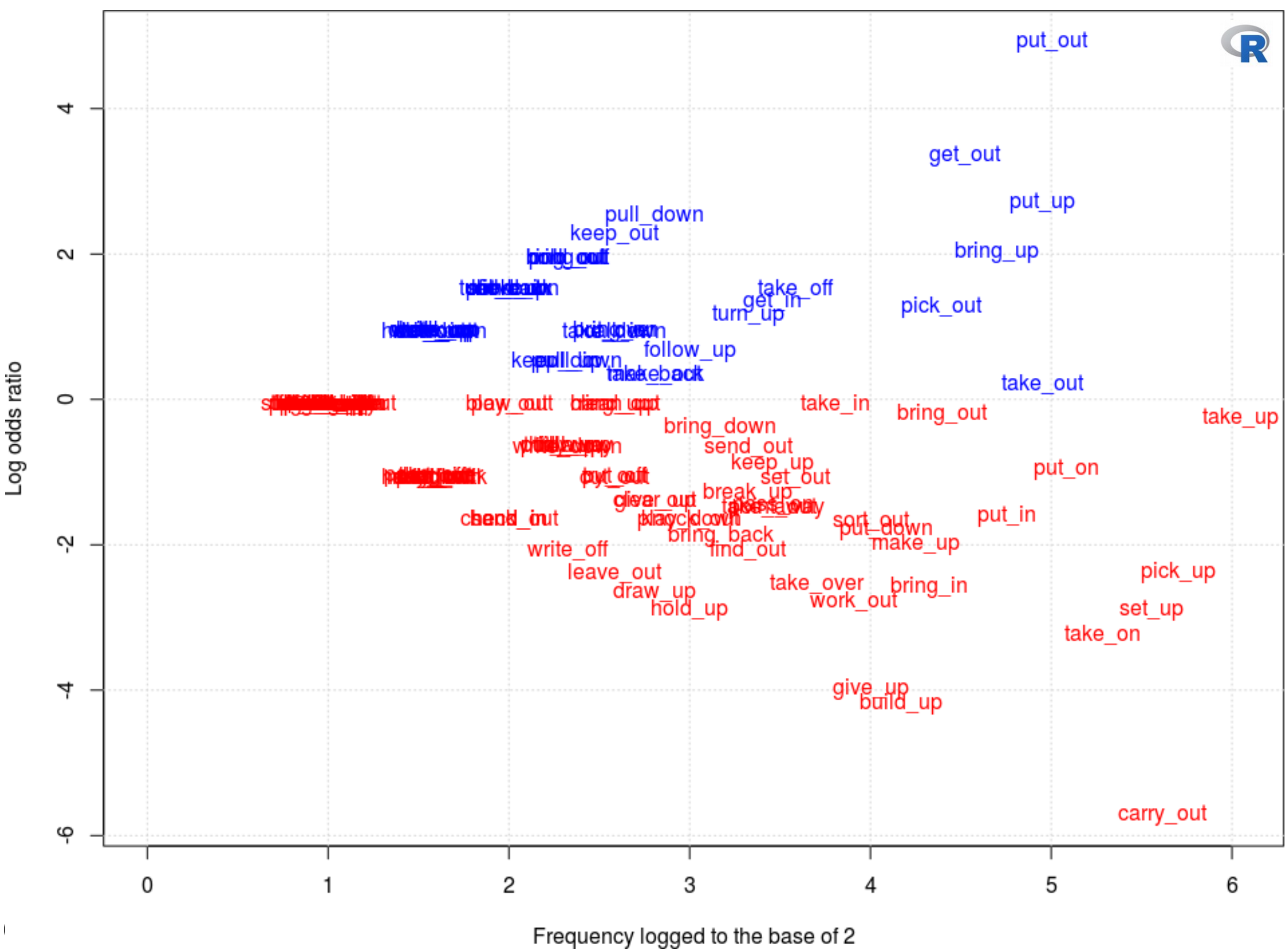
A distinctive collexeme analysis of verb-particle constructions: steps 1-3

- Data: ICE-GB
  - 1164 VPCs
  - 835 different verbs
    - with freqs between 1 and 31
- step 1: frequency
  - hard to evaluate, seems reasonable(?)
- step 2: association w/ G^2/LLR, conflating freq & effect
  - seems very similar w/ ranking changes
- step 3: keeping frequency & association separat[e] much more informative

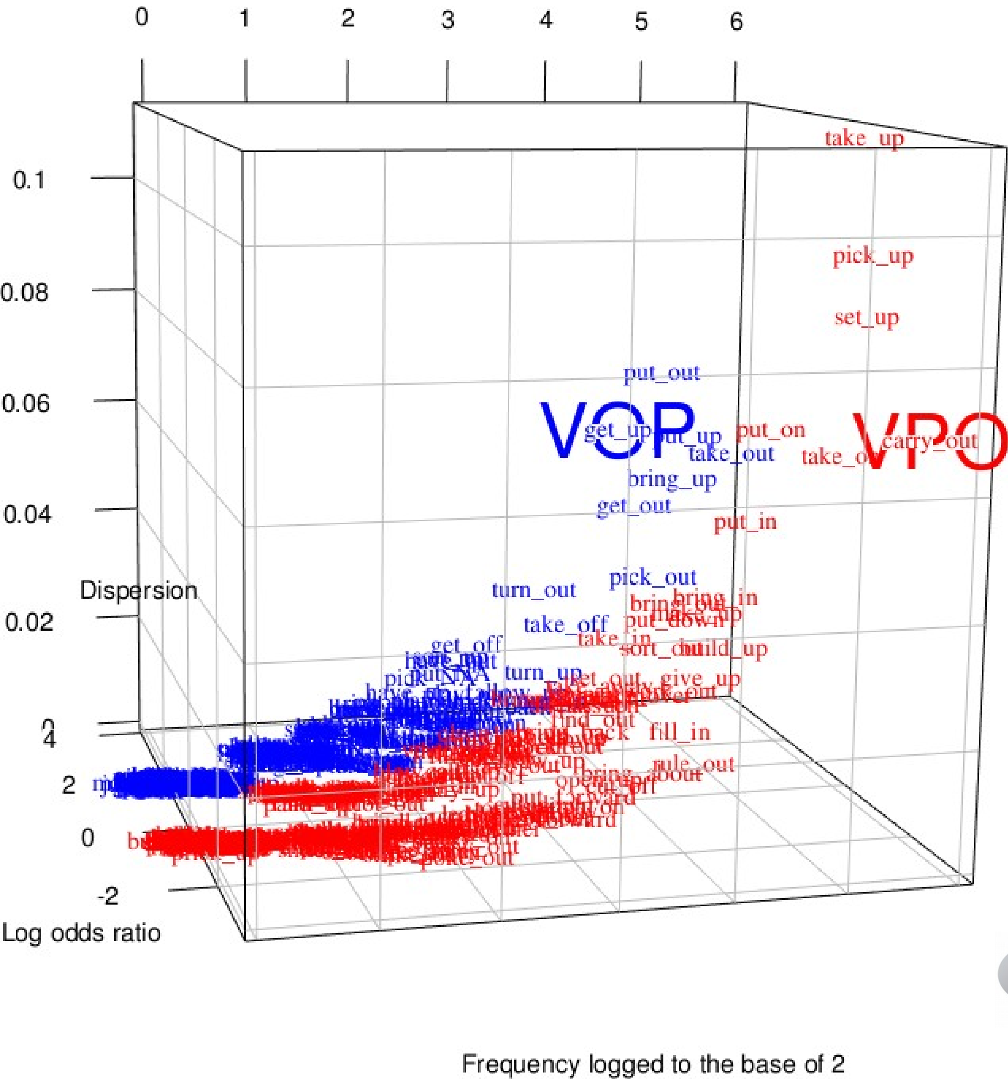


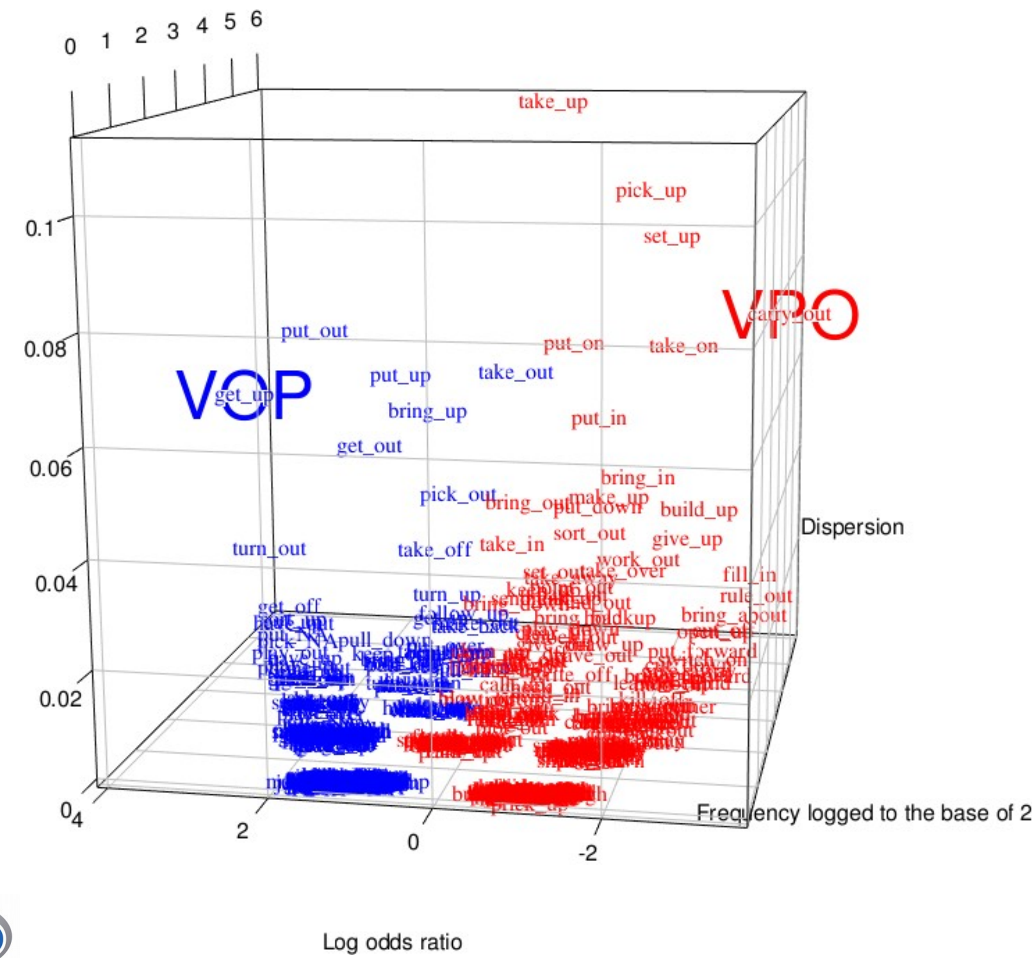
A distinctive collexeme analysis of verb-particle constructions: steps 1-3

- Data: ICE-GB
  - 1164 VPCs
  - 835 different verbs
    - with freqs between 1 and 31
- step 1: frequency
  - hard to evaluate, seems reasonable(?)
- step 2: association w/ G^2/LLR, confla-ting freq & effect
  - seems very similar w/ ranking changes
- step 3: keeping frequency & association separat much more informative



A distinctive collexeme analysis of verb-particle constructions: step 4





Concluding remarks re collostructions (from Gries 2012, 2015)

- Collostructional analysis has been widely applied
  - diachronic & synchronic construction studies
  - first & second/foreign language acquisition
  - psycholinguistic studies of priming, …
- while its implementation may need to vary between applications, the association logic per se is sound
- so don't believe all sorts of nonsense about it
  - no, the use of AMs – *p*-based or otherwise – is not a big significance testing problem but maybe a conflation one
    - conflation of effect & frequency
    - conflation of direction of association
  - no, the other-other cell (d) is not a huge problem – you estimate it reasonably
  - no, semantics doesn't go *into* it, but it might *emerge from* it
  - so, if you criticize it for something
    - you better understand it first
    - provide alternative measures that are as good or better
    - *then* we can talk …

		as-predicative	
		yes	no
regard	yes	80 (a)	19 (b)
	no	607 (c)	137958 (d)

Concluding remarks re association

- In terms of learning, acquisition, & processing, there's little that's more important than association
- association measures quantify
  - what-if?
  - if …, then …
- different measures are available,
  - all based on frequency of occurrence and co-occurrence
  - but differing in terms of implementation & implications
  - which shows that 'frequency' per se is versatile, if used properly and non-anxiously
- thus and not forgetting all previous 'lessons'
  - include frequencies of occurrence & co-occurrence
  - be aware of direction of association
  - be aware of dispersion
  - be aware of whether you can or cannot tolerate the information loss resulting from conflation – if yes, conflate properly

Thank you!

http://tinyurl.com/stgries