On frequency in corpora 2: the broader picture (dispersion, entropies, Zipf, …)

Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara
http://tinyurl.com/stgries

1: the frequency of A in X

frequency of A in X	A: *give*	112

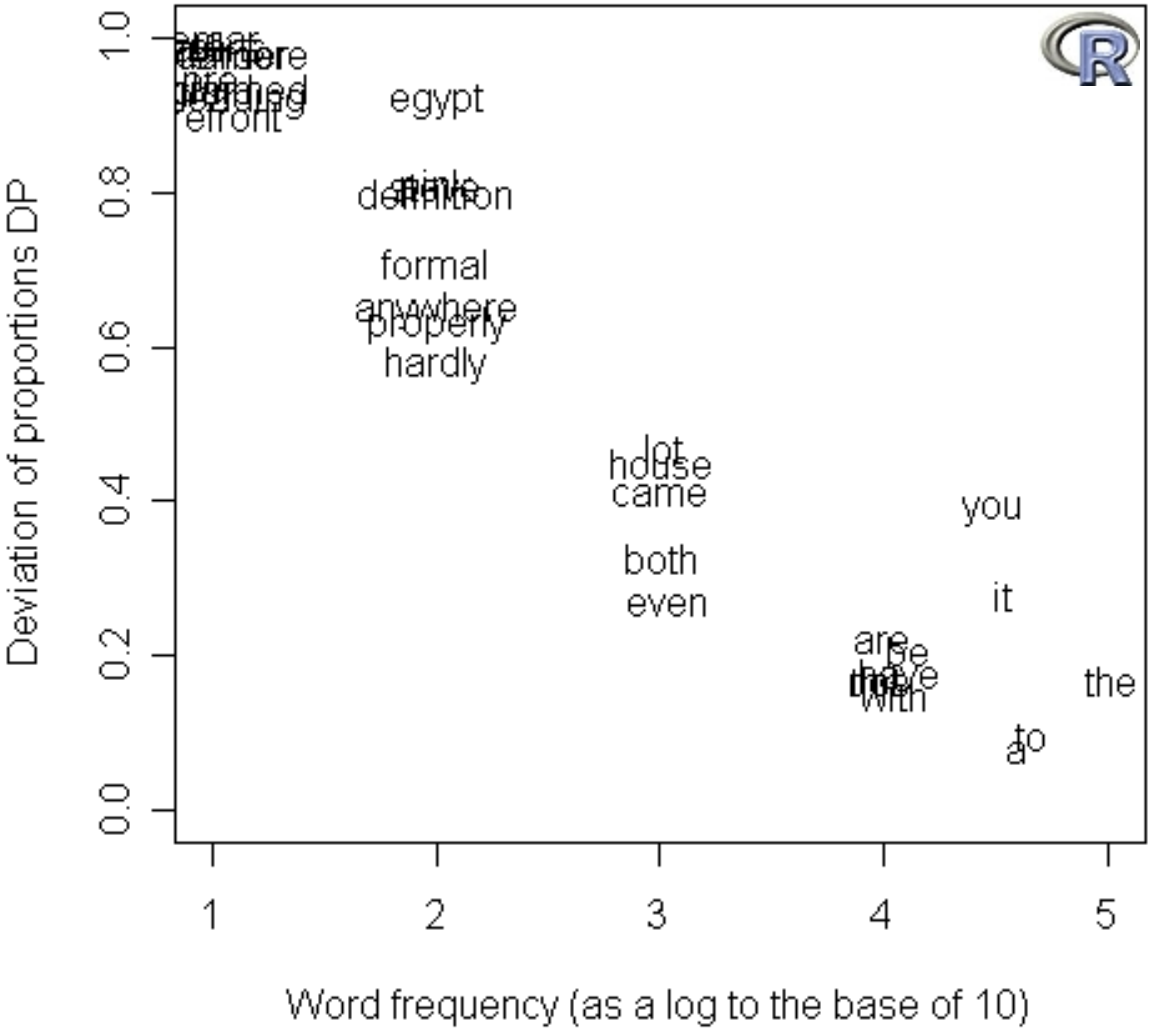
- The simplest corpus-linguistic method involves raw counts of A, as found in frequency lists
- while a crude tool, token frequencies are important for many linguistic areas: they correlate with
  - cognitive entrenchment (Schmid 2000)
  - phonetic reduction and development of new forms (Schuchardt 1885, Fidelholtz 1975, …)
  - resistance to language change (Bybee & Thompson 1997)
  - ease/earliness of acquisition (Casenhiser & Goldberg 2005)
  - reaction times in lexical decision tasks, word naming, picture naming (Howes & Solomon 1951, Forster & Chambers 1973, but cf. below …)

However, frequencies are risky when not coupled with dispersion measures

- Frequencies per se can be very misleading
  - Leech, Rayson, & Wilson (2001) show that *HIV*, *keeper*, and *lively* have about the same relative frequency of occurrence in the BNC (16 p.m.), but …
    - *HIV* occurs only in 62 of 100 equally-sized corpus parts whereas *keeper* and *lively* occur in 97 of these parts
    - Juilland et al.'s *D* for *HIV*: 0.62
    - Juilland et al.'s *D* for *keeper*: 0.87
    - Juilland et al.'s *D* for *lively*: 0.92
- many measures of dispersion have been proposed
  - Juilland et al.s'*D*, Carroll's D_2, Rosengren's *S*, inverse document frequency, Distributional Consistency, …
- but many come with problems
  - needing equally large corpus parts or neglecting sizes
  - depending on the order of corpus parts considered
  - too sensitive (to zeroes or outliers)
  - too insensitive (returning maximal values too easily)
  - incomparable ranges

A dispersion measure to be added when frequencies are considered

- A measure that doesn't suffer from these issues: *DP* (Gries 2008, 2010)
  - compute the size of each corpus part in %
  - compute the frequencies of the word in each part in %
  - compute absolute differences between the %s, sum, and divide by 2
- result: ≈0 ≤ *DP* ≤ 1
  - *DP* is small: the word is distributed evenly
  - *DP* is large: the word is distributed clumpily

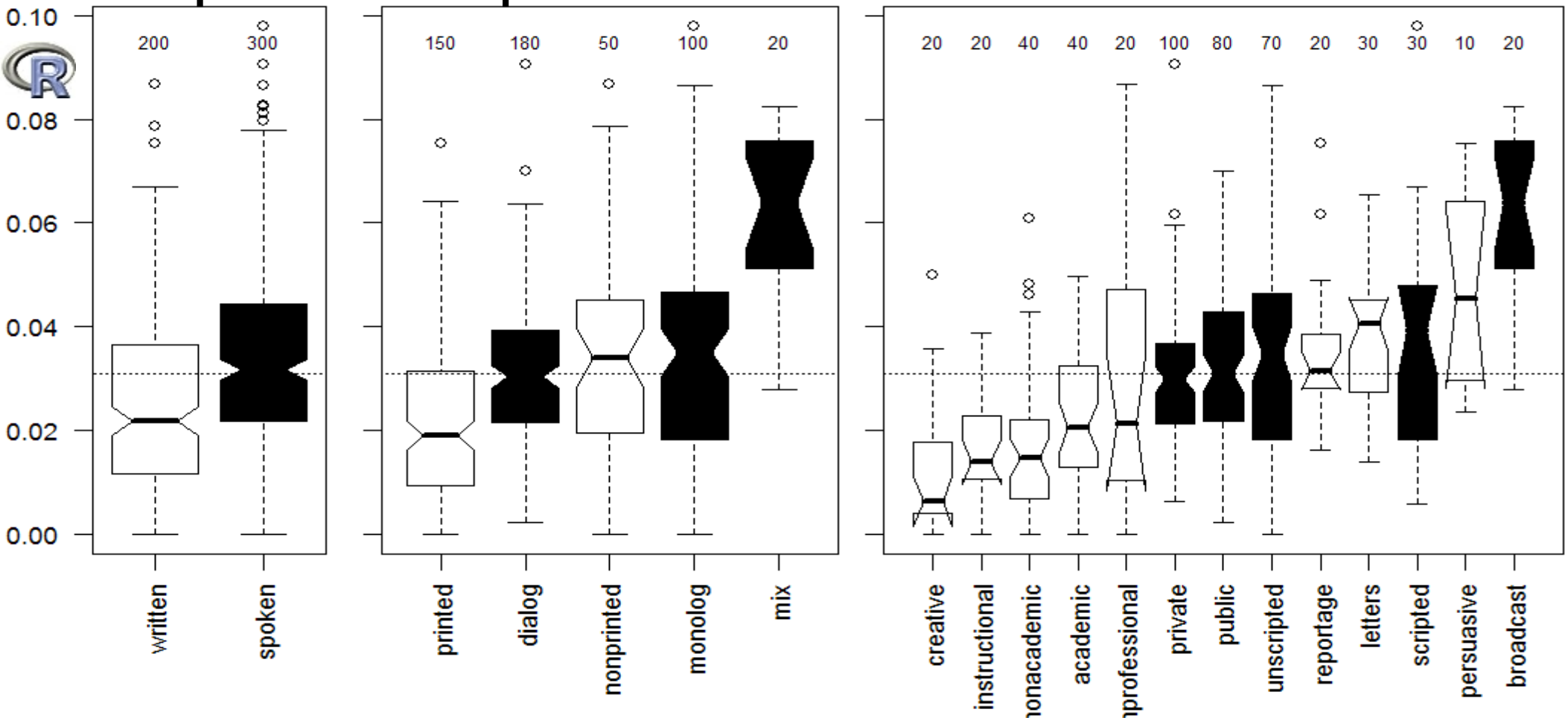


But dispersion is usually computed across linguistically meaningless parts

- Most (of the too few) applications of dispersion measures are based on dividing the corpus into parts that are linguistically irrelevant, typically files
- but corpora usually come with a linguistically meaningful substructure …
- … which means any statement about what's in a corpus is
  - a generalization over parts of a corpus that may be valid, but …
  - a generalization over parts of a corpus that
    - hopes that the H_0 of equal distributions is right
    - may be terribly wrong or oversimplified if said H_0 is wrong

Mode	Register	Sub-register
spoken	dialog	private
		public
	monolog	scripted
		unscripted
	mix	broadcast
written	print	academic
		creative
		instructional
		non-academic
		persuasive
		reportage
	non-printed	letters
		non-professional

Overall frequencies of present perfects in the ICE-GB



- frequencies should be checked regarding within-corpus homogeneity and maybe explored bottom-up with bootstrapping/resampling methods (cf. Gries 2006)

However, frequency effects are amplified by / emergent from context effects …

- Often, raw decontextualized frequency counts are not as relevant anyway
  - Raymond & Brown (2012) find that reduction effects are less due to overall frequency and more due to cumulative exposure and contextual predictability
- more radically, Baayen (2010) suggests that frequency effects might be epiphenomenal
  - word frequency is correlated with many other lexical properties; in fact, …
  - … most of the variance in lexical space is carried by a principal component on which contextual measures load highest: syntactic family size, syntactic entropy (!), BNC dispersion (!), morphological family size, adjective relative entropy, variety of contexts) – by contrast, …
  - frequency only explains a modest proportion of lexical variability
- so why don't we add context …

2: the frequency of A in P in X

frequency of A in P in X		P: ditrans.	in corpus (part)
	A: *give*	66	X

- A method that takes contexts of A more into consideration involves the notions of collocates or colligations/collostructions, where, say, a word is seen in its context (e.g., the words or patterns/constructions with/in which it co-occurs)
- frequencies of occurrence in a context have many applications: they correlate with
  - phonological reduction phenomena (cf. above, Bybee 2002)
  - grammaticalization, emergence of prefabs (Bybee 2010), verb islands in L1 acquisition (Tomasello 2005), …
- however, as we have seen when discussing collostructions, the frequency of A in P should be augmented by, or at least compared with, frequencies of A in, say, a competing context

3: the frequency of A in P & Q in X

frequency of A in P and Q in X		P: ditrans.	Q: prep. dat.	in corpus (part)
	A: *give*	66	23	X
	A: *!give*	400	500	X

- An approach taking a bit more comprehensive view at the distribution of A uses the above kinds of association measures computed from 2x2 tables
  - percentages or conditional probabilities (e.g., *MinSen*)
  - bi-directional assoc. measures (e.g., *MI*, *t*, *LL*, p_{FYE})
  - uni-directional assoc. measures (e.g., Δ*P*)
- what do the words ranked highest by association measures show/reflect?
  - for argument structure constructions: the core verb and the senses of the construction (Stefanowitsch & Gries)
  - lexically-specific sizes of priming effects (Szmrecsanyi 2005, 2006; Gries 2005;)
  - high correlations with learners' completions and acceptability ratings (Gries & Wulff 2005, 2009)
  - predict reduction effects (Gregory et al. 1999)

$-\log_{10} p_{FYE}$ yields good results, but can we do better than that?

frequency of *(!)x* and *(!)y* in some corpus		2 *y*	*!y*	Totals
	1: *x*	a	b	a+b
	1: *!x*	c	d	c+d

- Nearly all measures are bidirectional – but (associative) learning is not
- I will explore a measure called ΔP, which
  - arose out of the associative learning literature
  - was first discussed in linguistics by Ellis (2007)
- $\Delta P = p(\text{outcome}|\text{cue=present}) - p(\text{outcome}|\text{cue=absent})$
- $\Delta P = 0$ when the two p's are the same and the cue doesn't affect the probability of the outcome
- $\Delta P > 0$ / $\Delta P < 0$ when the presence of the cue increases / decreases the probability of the outcome
- thus, ΔP
  - normalizes conditional probabilities
  - is computationally easy to obtain
  - may be cognitively more realistic

Towards a different directional AM … Δ*P*

- The formulae

	$word_2$: present	$word_2$: absent	Totals
$word_1$: present	a	b	a+b
$word_1$: absent	c	d	c+d
Totals	a+c	b+d	a+b+c+d=N

$$\Delta P_{2|1} = p(word_2|word_1 = present) - p(word_2|word_1 = absent) = \frac{a}{a+b} - \frac{c}{c+d}$$

$$\Delta P_{1|2} = p(word_1|word_2 = present) - p(word_1|word_2 = absent) = \frac{a}{a+c} - \frac{b}{b+d}$$

- e.g., *of course*

	course: pres.	*course*: abs.	Totals
of: present	5610	168938	174548
of: absent	2257	10233063	10235320
Totals	7867	10402001	10409868

$$\Delta P_{2|1} = p(course|word_2 = of) - p(course|word_2 \neq of) = \frac{5610}{174548} - \frac{2257}{10235320} \approx 0.032$$

$$\Delta P_{1|2} = p(of|word_2 = course) - p(of|word_2 \neq course) = \frac{5610}{7867} - \frac{168938}{10402001} \approx 0.697$$

- compare that to
  - *MI*=5.41, *t*=476.97, *G*=36693.85, $p_{FYE}<10^{-320}$, …

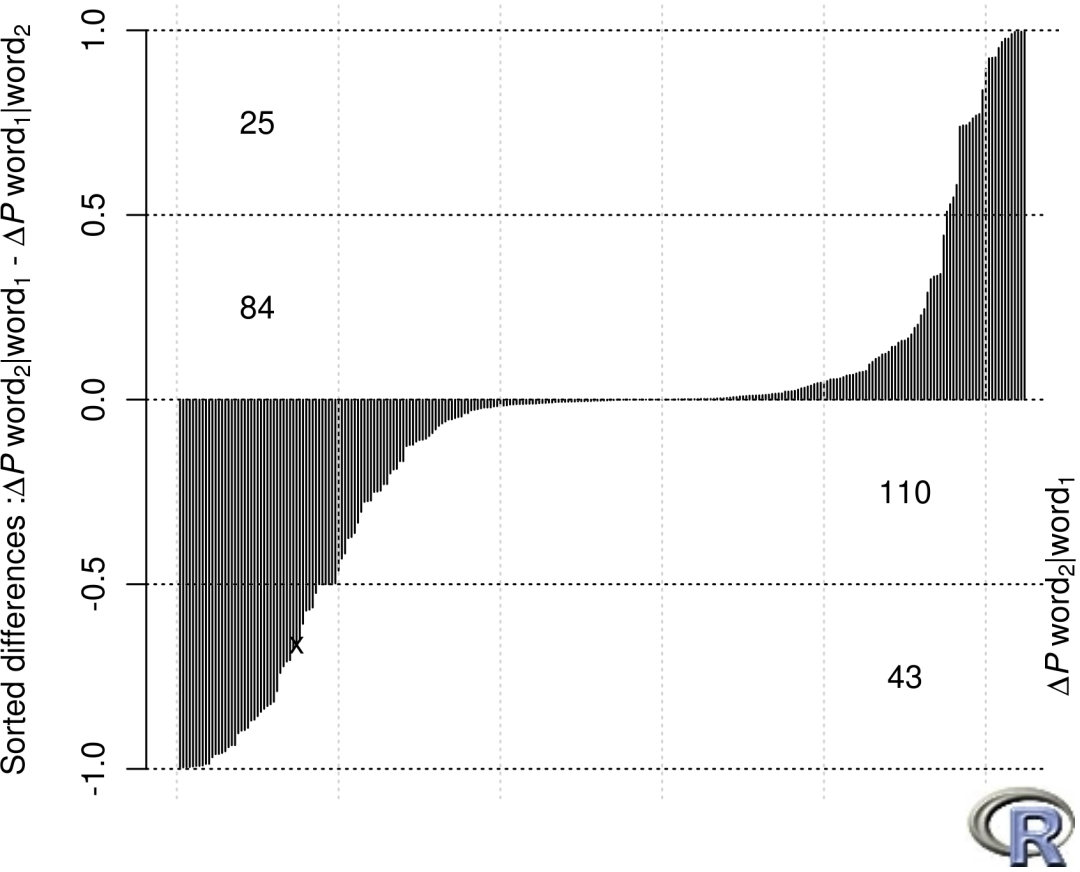
Validation 1a: strong collocations

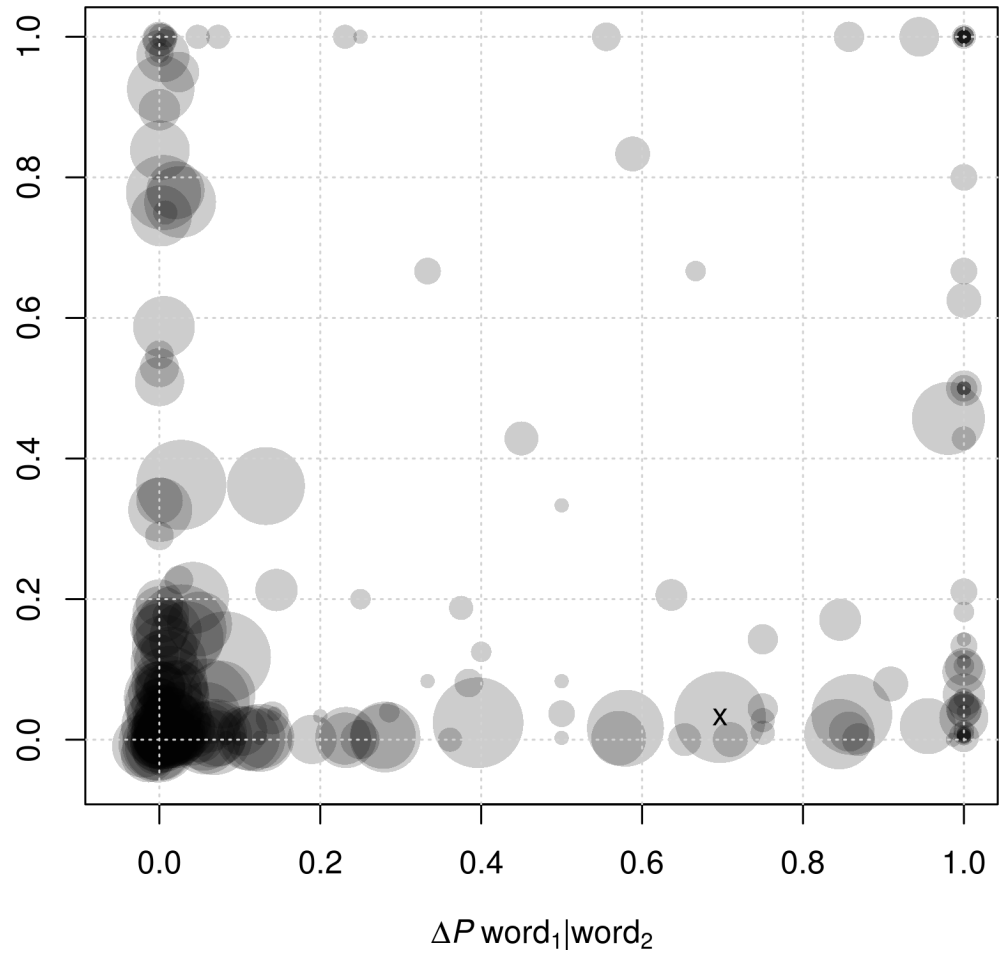
- To explore whether Δ*P* finds asymmetries in strong collocations, I computed all the above collocational statistics and both Δ*P*s for 262 two-word units annotated as such in the spoken component of the BNC
- results
  - the means of all measures support the strong association of these 2-word units – but some measures' quantiles go far into repulsion territory, but not so for Δ*P*

	MI	*t*	G^2	$\Delta P_{1\|2}$	$\Delta P_{2\|1}$	*MinSem*
mean	7.65	466.4	1064.11	0.28	0.2	0.1
0.025 quantile	-3.68	-13.52	-287.05	-0.01	-0.01	0
0.975 quantile	22.79	3226.43	12909	1	1	1

  - more than 25% of all 2-word units are highly asymmetric, with absolute differences between Δ*P*s of more than 0.5
  - in a way that is not straightforwardly related to frequency of co-occurrence, the units in question are different in how one word attracts the other more/less

Validation 1a: strong collocations and their amounts of asymmetry

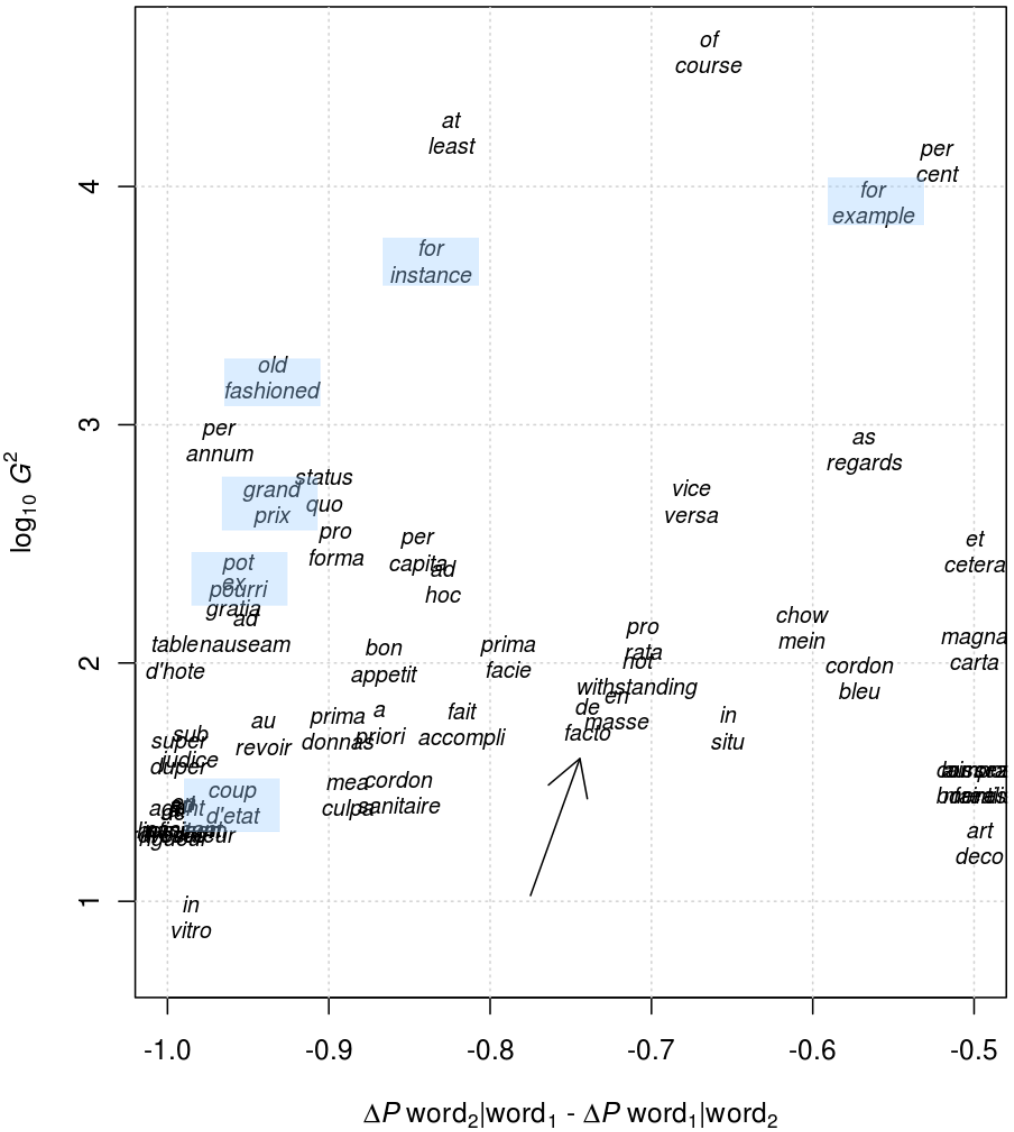




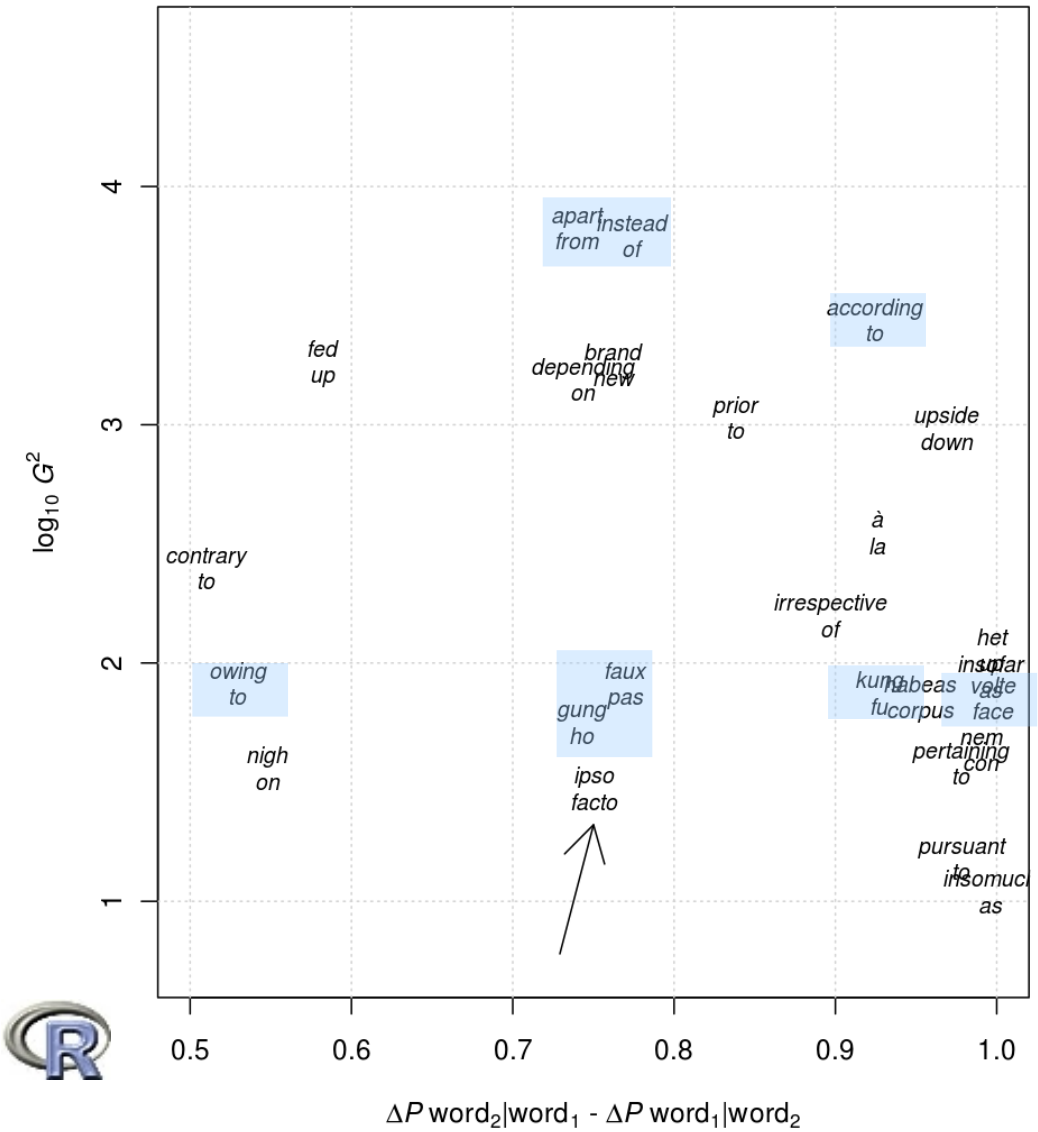
Collocations are not necessarily bidirectional:
of course (and others)

validation 1b: what are asymmetric ones?

Word$_2$ is much more predictive of word$_1$ than vice versa



Word$_1$ is much more predictive of word$_2$ than vice versa



Collocations are not necessarily bidirectional:
of course (and others)

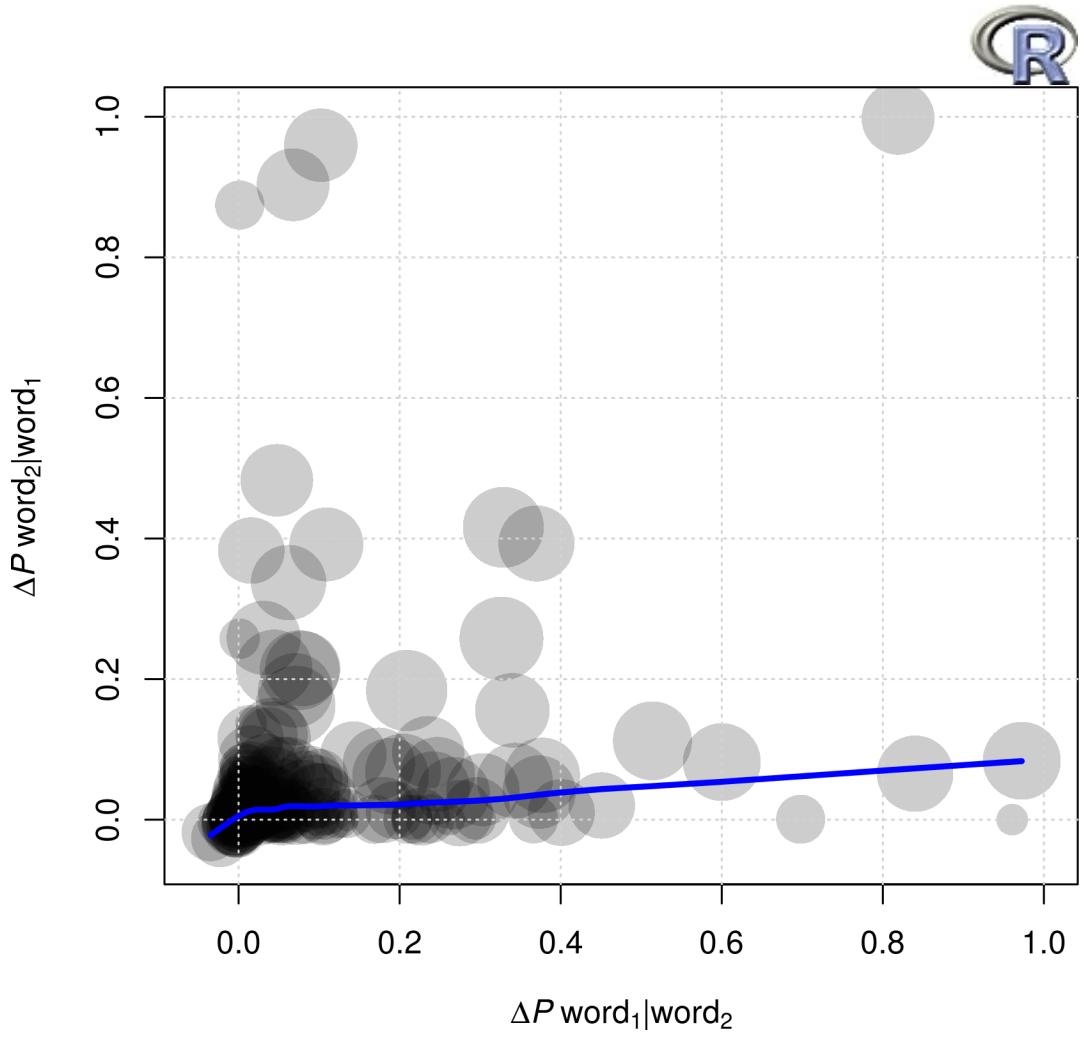
Validation 2b: pseudo-randomly chosen 2-word units

- But this just used strong collocations – is Δ*P* discriminative enough to not always return such results, e.g., with weak collocations
- I computed all the above measures for 237 pseudo-randomly chosen within-sentence 2-word collocations from 8 frequency bins
- results
  - most measures reflect the randomness of the collocations and the values for Δ*P* and the quantiles are much smaller than before and include 0

	MI	*t*	G^2	$\Delta P_{1\|2}$	$\Delta P_{2\|1}$	*MinSem*
mean	2.28	126.18	14687.63	0.08	0.05	0.03
0.025 quantile	-2.23	-15.08	0.25	-0.01	-0.01	0
0.975 quantile	6.31	802.39	147225.3	0.4	0.52	0.12

  - what are the 8 units for which high Δ*P* values are found?
    - diffs <0: *I mean*, *I think*, *I'm*, *the faintest*, *the biggest*
    - diffs >0: *sort of*, *can't*, *lack of*

validation 2b: pseudo-randomly chosen 2-word units



Interim summary

- What we have seen: Δ*P*
  - is by design more sensitive than traditional association measures since it can tease apart directionality effects
  - is very easy to understand and compute – everybody understands differences between percentages – while at the same time not as arbitrary as, say, Kilgarriff's add-*n* approach
  - makes no distributional assumptions and avoids problems of the Null Hypothesis Significance Testing paradigm (for those who care)
  - has received experimental support both in psychology and in linguistic work by Ellis and colleagues
  - maybe it can even help explore mismatches between corpus and experimental data (e.g., Mollin 2009)
- but maybe it's even more useful to not consider A in not-P a single category – a more comprehensive division of this frequency might actually be useful

4: the frequency of A in P(, Q, R, S, …) in X

frequency of A in P(, Q, R, S, …) in X	P: ditrans.	Q: prep. dat.	R: phras. V.	S: idioms	in corpus (part)
A: *give*	66	23	16	7	X
A: …	…	…	…	…	X

- A yet more comprehensive look at A's distribution involves recovering its whole range of usage types
- note how this is actually a different type of dispersion
  - not a dispersion of A across files or corpus parts
  - but a dispersion of A across co-occurrence patterns
- what is this relevant for?
  - e.g., a more precise identification of verbs' exact sub-categorization preferences (i.e., form-function contingencies) can
    - reveal the reliability of a form-function cue
    - help identify cases of preemption
  - the spread of forms of a type over different patterns has implications for learning and processing

The role of distributions (and entropies): experimental data

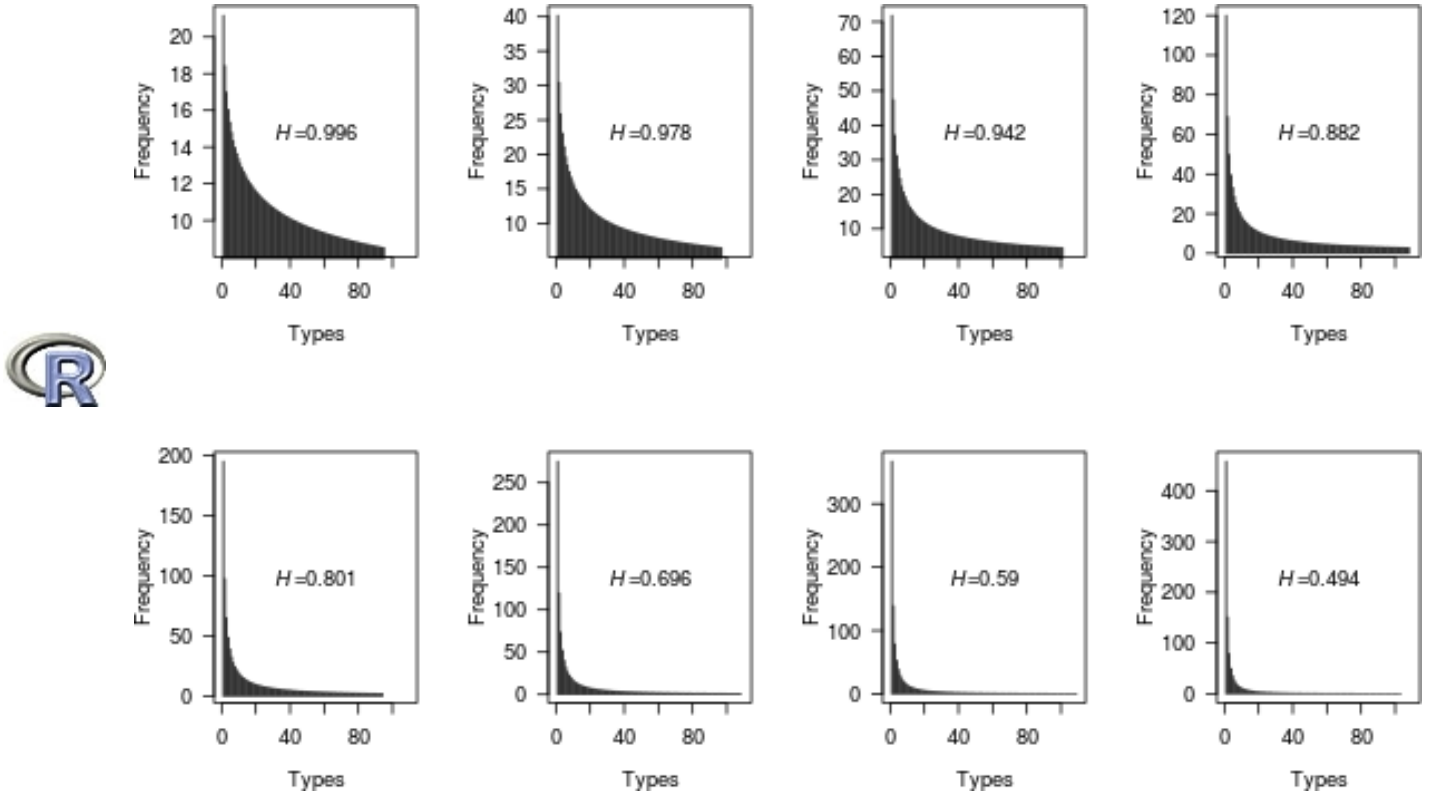
- Casenhiser & Goldberg (2005) find that children and adults learn a new construction better from skewed than from balanced exposure (5 verb types, 16 tokens)
  - skewed condition: 8-2-2-2-2 (H_{rel}=0.86)
  - balanced condition: 4-4-4-2-2 (H_{rel}=0.97)

$$H_{rel} = -\sum_{i=1}^{n} p \cdot \log p \cdot \frac{1}{\log n}$$

- Boyd & Goldberg (2011: exp. 2-3) show that speakers
  - learn to not use 4 novel *a*-adjectives prenominally from only 3 exposures to 2 of these adjectives in a preempting relative clause context
  - distinguish preemptive from pseudo-preemptive contexts
- we also know from non-linguistic categorization that categories with lower member type frequencies, lower entropies, and much exposure to the prototype are learned better than more representative categories
- "in category learning in general a centred, or low variance, category is easier to learn" (Bybee 2010:89), and words in a cx slot are a category

3: The frequency of A in P & Q in X Token AND type frequencies
4: The frequency of A in P(, Q, R, S, …) in X The notion/role of entropy (experimental data)
5: The freq. of A in P(, Q, R, S, …) in X(, Y, Z, …) The notion/role of entropy + Zipf (corpus data)
6: The similarity of As in P(, Q, R, S, …) in X

Where does *this* lead us?



The role of distributions (and entropies): corpus data

- Goldberg, Casenhiser, & Sethuraman (2004) find
  - verb types in constructions in CDS exhibit a Zipfian distribution
  - verb types in constructions in children's utterances exhibit a Zipfian distribution
- Ellis & Ferreira-Junior (2009)
  - find similar Zipfian distributions for, e.g., the caused-motion construction (V NP PP) and the ditransitive construction (V NP NP)
  - find frequency of learner uptake is predicted by frequency and association measures (ΔP and p_{FYE}), but p_{FYE} outperforms freq in $^{2}/_{3}$ constructions $\Delta P = P(O|C) - P(O|\neg C)$
  - conclude (p. 216) that for each island there is the frequency, the frequency distribution, … the reliabilities of the form-function mapping, and the degree to which the different elements in the VAC sequence are mutually informative and form predictable chunks – ideally, we'd have all of this information

5: the frequency of A in P(, Q, R, S, …) in X(, Y, Z, …)

frequency of A in P(, Q, R, S, …) in X(, Y, Z, …)	P: ditrans.	Q: prep. dat.	R: phras. V.	S: idioms	in corpus (part)
A: *give*	66	23	16	7	X
A: *give*	35	33	15	8	Y

- What is one of the most frequent PPs in corpus-linguistic papers?
  corpus]]
- but, as we saw above, relying on a frequency *in the/my corpus* is often a huge simplification
  - because of dispersion across files
  - because of dispersion across meaningful corpus parts
- now, if that's true for a simple frequency, it's gonna be even more true for more complex data, e.g., A: *give*'s behavior in different corpus parts …
- … and what if we, as we should, include data for more than just A's distribution, a bottom-up strategy becomes even more indispensable

How much verbs – not just *give* – like the ditransitive in parts of the ICE-GB

- Using the ICE-GB, I computed for each verb how much it prefers to occur in the ditransitive
  - in the whole corpus
  - in the 5 registers
  - in the 12 sub-registers (because sub-reg = 1 reg)
- that is, the result is a table
  - with 18 columns (the above corpus parts)
  - with 87 rows (one for each verb used ditransitively)
  - where each cell contains a number quantifying how much the verb (dis)likes the ditransitive in the corpus part
- a principal components analysis of this table revealed 4 principal components with *Eigenvalues*>1 (22.2% of columns account for 72.35% of the variance)
  - spoken (- private dialog)
  - spoken private dialog
  - written printed
  - written non-printed

What the bottom-up division of a corpus can look like …

- Thus, a corpus linguist interested in an analysis of X (ditransitive semantics) that takes registers / text types into account should distinguish these 4 corpus parts
- note
  - this is not just spoken vs. written
  - this is not just a division into registers or into sub-registers
  - rather, the obtained corpus division cuts across different levels of granularity, something that analysts usually don't like to do

6: the similarity of As in P(, Q, R, S, …) in X

match 1 in X	He		gave	him	the book	
match 2 in X	My father	did not	give	me	his car	
match 3 in X			Give	peace	a chance	
match 4 in X	The mailman		gave		the guy	the finger

- In the discussion of frequency data so far, we have moved 'outwards' from a particular frequency
  - we started from A in P in X and successively
  - extended the range of frequencies from that to A in P(, Q, R, S, …) in X(, Y, Z, …)
  - now let's move 'inwards', by having a closer look at the *n* instances of A in P in X and their similarities
- Why? Because analogy/similarity between uses of A are relevant in a variety of contexts including
  - the formation of novel utterances in L1 acquisition is facilitated by their similarity to prior utterances
  - language change is driven by analogy/similarity
  - structural priming/persistence in 'targets' is affected by analogy/similarity to primes

Local similarity and its effect on structural priming/persistence

- Szmrecsanyi (2005, 2006) explores persistence
  - α-persistence: a structure *x* increases the probability of the same structure *x* at the next point of choice
  - β-persistence: a structure *x* increases the probability of a similar structure *y* at the next point of choice
    - 533 analytic vs. synthetic comparisons (BNC CG)
    - 35,558 tokens of futures (BNC DS)
    - 1048 tokens of particle placement (FRED)
- results
  - comparison *more* → * more analytic comparatives
  - (motion) *go* → * more *going-to* futures
  - particle placement: same phrasal verb → * more priming (same in Gries 2005)
- thus, lexical similarity in spite of syntactic differences results in re-activation of structures
- but, "the problem […] is in specifying the relevant features upon which similarity is measured" (Bybee 2010:62)

Global similarity and its effect on structural priming/persistence

- Snider (2009) explores
  - whether verb repetition increases priming (unsurpri-singly, it does; cf. Pickering & Branigan 1998, Szmrecsanyi 2005, 2006, Gries 2005)
  - whether prime-target similarity facilitates priming
  - crucially, he adopts an 'overall similarity' kind of perspective, using a multi-feature distance metric, Gower's metric, and a GLMEM
    - 1002 tokens of the dative alternation (switchboard)

$$d_{ij} = \frac{\sum_{k=1}^{p} w_k \delta(ij;k) d(ij;k)}{\sum_{k=1}^{p} w_k \delta(ij;k)}$$

- result: the more similar prime and target are, the more likely they also involve the same construction
- implication: lexical and structural priming may be more similar to each other than assumed so far: both respond to similar factors
- similarity operates on many many levels
- things we count may be affected (much) more by previous ones than by their properties per se!

Speakers keep track of all these kinds of distributional info … and fast!

- The previous discussion has – hopefully – shown that speakers keep track of vast amounts of multi-dimensional statistical/probabilistic co-occurrence information
- the amazing things are
  - how early this happens (recall the classic studies on infants recognizing transitional probabilities of syllables), …
  - … but also, maybe more 'dangerously', …
  - how fast this happens
    - in L1 acquisition
    - with adult native speakers
      - in L2 acquisition/learning
      - in language contact situations
    - …

And we find this in all sorts of domains

- In L1 acquisition (see above)
  - Casenhiser & Goldberg (2005) taught 5-7 yr. olds a new English construction meaning 'appearance' with 16 clips
- with adult native speakers (see above)
  - Boyd & Goldberg (2011) show how speakers learn not to use novel *a*-adjectives from only 3 exposures in a preemptive context
- in L2 acquisition/learning
  - Gries & Wulff (2009) found a (weak) accumulative priming effect in a sentence-completion task: over the course of the experiment, subjects primed themselves more to using construction *x* more completions of type *x*
- in language contact situations
  - Doğruöz & Gries (2012) find that, over the course of a small acceptability judgment experiment testing frequency and conventionality effects, subjects significantly relax their more critical judgments regarding unconventional utterances

Where does this lead us?

- It leads to a cline of co-occurrence complexity
  - 1: observed frequencies/percentages of *w*s in *c*
    - but this is very limited
  - 2: associations (AMs) of *w*s to *c* (against *w* & *c* overall)
    - all of approach 1 but more comprehensive and still limited
  - 3: full cross-tabulation of *w*s and *c*s
    - all of approaches 1-2 but way more comprehensive
    - why is this great?
      - we get type frequencies (cf. *G*)
      - we get token distributions and, thus, entropies
        - Goldberg, Casenhiser, & Sethuraman's (2004) study on learning a new *c*: 8-2-2-2-2 (*H*=2) with 4-4-4-2-2 (*H*=2.25)
        - Goldberg (2006) on the importance of low-variance samples
  - 4: adding dispersion to the mix:
    "Given a certain number of exposures to a stimulus, […] learning is always better when exposures or training trials are distributed over several sessions than when they are massed into one session." (Ambridge et al. 2006:175)

Where does this lead us?

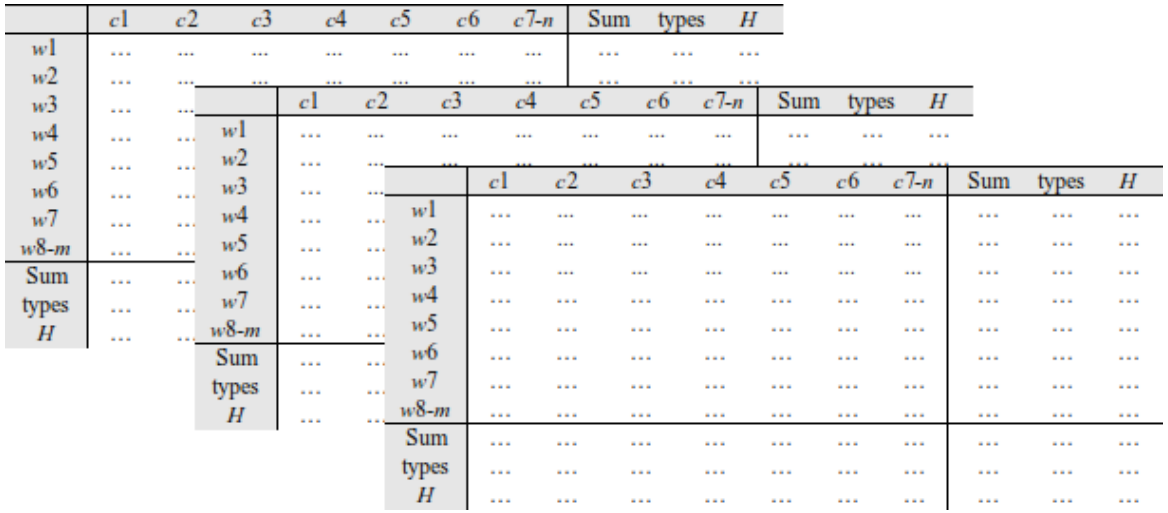
	c 1
w 1	**80**
w 2	**60**
w 3	**40**
…	…

	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

	*c*1	*c*2	*c*3	*c*4	*c*5	*c*6	*c*7-15	Sum	types	*H*
*w*1	**80**	*90*	*45*	*35*	*25*	*5*	*0*	**280**	6	2.26
*w*2	**60**	*0*	*310*	*0*	*0*	*0*	*0*	**370**	2	0.639
*w*3	**40**	*30*	*30*	*30*	*30*	*30*	*270*	**460**	15	3.902
*w*4	40	407	1	1	1	1	9	460	15	0.713
*w*5	40	420	0	0	0	0	0	460	2	0.426
*w*6	40	1	407	1	1	1	9	460	15	0.713
*w*7	40	0	420	0	0	0	0	460	2	0.426
*w*8-20	…	…	…	…	…	…	…	…	…	…
Sum	**1080**	948	1213	…	…	…	…	sum	15	…
types	…	…	…	…	…	…	…	…	…	…
H	…	…	…	…	…	…	…	…	…	…



- and wherever it says *w*1, *w*2, …, we would better distinguish senses $s1_{w1}$, $s2_{w1}$, …, sn_{w1}, $s1_{w2}$, …
(G&S 2004, C&B 2012)

Given that this is the required level of complexity, why does CA work at all?

- The values of the most frequently used AM in CA, p_{FYE}, are highly correlated with ΔP, the measure from associative learning (Ellis & Ferreira Junior 2009)
- CA is a good approximation of approach 3
  - because the ratios of observed percentages are 'weighted' by frequency of occurrence (cf. above)
  - because the token frequencies in cells *b* and *c* approximate the token distributions in cells
    - *w*1 x *c*2-15
    - *c*1 x *w*2-20
    - (and this is *d*)
- but how can this then be a good approximation?
  - the token distributions are all somewhat Zipfian
    - few types will be frequent
    - many types will be infrequent

Given that this is the required level of complexity, why does CA work at all?

	*c*1	*c*2	*c*3	*c*4	*c*5	*c*6	*c*7-15	Sum	types	*H*
*w*1	**80**	*90*	*45*	*35*	*25*	*5*	*0*	**280**	6	2.26
*w*2	**60**	*0*	*310*	*0*	*0*	*0*	*0*	**370**	2	0.639
*w*3	**40**	*30*	*30*	*30*	*30*	*30*	*270*	**460**	15	3.902
*w*4	40	407	1	1	1	1	9	460	15	0.713
*w*5	40	420	0	0	0	0	0	460	2	0.426
*w*6	40	1	407	1	1	1	9	460	15	0.713
*w*7	40	0	420	0	0	0	0	460	2	0.426
*w*8-20	…	…	…	…	…	…	…	…	…	…
Sum	**1080**	948	1213	…	…	…	…	sum	15	…
types	…	…	…	…	…	…	…	…	…	…
H	…	…	…	…	…	…	…	…	…	…

Where does *this* lead us?

- Goldberg (2006:89) says the better learnability of skewed input "may involve a type of cognitive anchoring"
- however, I think it's just as possible to say
  - Zipfian distributions involve less uncertainty than uniform or less Zipfian distributions: the more tokens fewer types account for, the lower *H*
  - one way of understanding the learning of categories and their productivity then could be Hebbian learning
    - where the frequent types strengthen the core of the category/slot (but on their own would 'become' the category)
    - where the many infrequent types
      - indicate the category is in fact an open category by keeping few core types from taking over the category
      - strengthen the core of the category via co-activation (cf. Zeldes 2011)

How might that reshape our view of constructions in general?

- Goldberg (1995:4): constructions = form-meaning pairing with at least one unpredictable property
- in Goldberg (2006:5)
  - something unpredictable is not longer a nec. condition
  - something can also be a construction by virtue of "sufficient frequency"
- given
  - an exemplar-based view in which
  - linguistic knowledge is a multidimensional space
    - with formal dimensions
    - with functional dimensions
- a construction is an uncertainty/*H*-reducing spike of a distribution in a part of multidimensional space with at least one dimension being 'functional'
- a frequency becomes "sufficient" when the frequency distribution involving such a confluence of 1+ formal and 1+ functional characteristics has become uncertainty-reducing / Zipfian enough

To conclude: if anything, we need more complex tools …

- We have seen a variety of guidelines to bear in mind
  - comparisons against H0 and various baselines can be useful, as can identifying directions of effects
  - exploring variability and dispersions of data is essential as is an awareness of subtle contextual effects, many different kinds of predictors, & various correlational structures
  - type frequencies, token distributions, and entropies are all relevant in general, as are sense distinctions …
- most of the above is relevant to, and to some extent at least, embodied in collostructional analysis
- however, CA is also a massive simplification
- exploration of the *whole range of dimensionality* is necessary, and can lead to interesting new perspectives – reversing that trend won't help
- the versatility/complexity of our quantitative tools *must* do justice to the versatility/complexity of our theories

Thank you!

http://tinyurl.com/stgries