Constructions and their semantics/behavior: collostructional analysis

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

Corpora for lexis and syntax …

- In the past two decades or so, corpus linguistics has turned out to be the fastest growing methodological tool
- as I mentioned before, it is probably fair to say, however, that
  - corpus studies have been particularly strong in the domain of lexis
    - KWIC approaches to individual words or lemmas
    - collocational approaches
      (much of that with an eye to help lexicographers)
  - syntactic matters have not been neglected, but …
  - … given that most corpora are either not annotated at all or 'only' POS-tagged, a focus on more easily retrievable lexical items is unsurprising
- in this talk, I want to focus more on things that are more grammatical/syntactic in nature

… but no divide between lexis and syntax

- However, as mentioned before$_1$, I do share the theoretical orientation
  - of frameworks such as
    - pattern grammar (in corpus linguistics)
    - cognitive linguistics / construction grammar (in theoretical linguistics)
  - that there is no clear divide between syntax and lexis
- thus, my discussion of grammatical issues here will also involve lexical aspects

Patterns and constructions: two very similar entities

- As mentioned before$_2$, patterns and constructions are very similar entities
  - Hunston and Francis (2000:37): "A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it."
  - Goldberg (2006:5): "Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency."

… plus a quantitative perspective

- In addition, the case studies to be discussed below will be quantitative in nature because, to me, corpus linguistics is a discipline concerned first and foremost with distributional patterns
- corpora do not contain meaning, function, concepts – they only contain
  - information on (relative) frequencies of occurrence: elements that occur x many times (with $x=0$ or $x>0$)
  - information on dispersion: elements that occur
    - $x \geq 0$ time in particular parts of corpora
    - at particular distances $d \geq 0$ from each other
  - information on (relative) frequencies of co-occurrence (collocation, colligation, etc.)
  - derivatives of the above (e.g., key words)

Frequency-based approaches to patterns and constructions: the standard

- Until fairly recently, much work in pattern grammar, construction grammar, and other linguistic approaches studying patterns/constructions
  - has so far relied on
    - raw frequencies of (co-)occurrence
    - eyeballing (sorted) frequency lists
  - in the study of
    - structure/POS-sensitive collocates of a node word
    - grammatically-defined frames: Adj+N, N+N, N+P+N, etc.
    - collocational frameworks: *a*+N+*of*
    - colligations / grammatical patterns / constructions
      - lexically partially specified: V+NP+*into*+V-*ing*, V+*from*+V-*ing*, V+POSS+*way*+PP, N+*waiting to happen*, etc.
      - lexically unspecified: V NP, V NP NP, etc.

Frequency-based approaches to patterns and constructions: the problems

- This approach can be problematic though
  - often, a mere eyeballing of frequencies can fail to un-cover important results (see below for much more detail)
  - for example, Hunston & Francis discuss the V+*from*+V-*ing* construction on the basis of corpus data, but
    - they only focus on one of the two slots of the construction
    - they attribute a notion of "some kind of forcefulness or even coercion" (p. 106) to the verbs that occur with this pattern, but both *force* and *coerce* are notably absent from the several fragmented lists of verbs they discuss
    - they verb whose occurrence in the first verb slot of this pattern is highest in their list is *talk* (317) - but *talk* is very frequent in general: it's no surprise that it shows up very often – what is more surprising is that *talk* occurs only 4.7 times more often in that pattern's first slot than *coax* although it is 96 times more frequent in general

Well, if there's no divide between syntax and lexis …

- This approach can be problematic though
  - there is a large body of work on how to best quantify collocational strength – now if lexis and syntax are not considered to be fundamentally different, then why not extend the same methodological sophistication to the study of syntax/grammar? …
  - … which raises the question whether the results reported so far can be improved on
- in this first part of the talk, I want to explain one such approach, exemplify it, and discuss a few of its advantages and areas of application

Collexeme analysis

- Collexeme analysis/strength quantifies the degree of association – attraction or repulsion – between
  - one word and
  - one slot in one pattern/construction

The *as*-predicative: frequencies

- The *as*-predicative
  - I never [$_{VP}$ saw [$_{NP\text{-}DO}$ myself] as [$_{XP}$ a designer]]
  - Politicians [$_{VP}$ regard [$_{NP\text{-}DO}$ themselves] as [$_{XP}$ being closer to actors]]
- to determine how this pattern is used and what the verbs occurring in it reveal about the pattern's meaning, a typical corpus-linguistic approach would be to look at (the most frequent) verbs in this pattern
- 687 tokens of this construction (107 V types) were retrieved from the ICE-GB (cf. Gries, Hampe, & Schönefeld 2005, 2010)

Verb	Frequency in C
see	111
describe	88
regard	80
know	79
use	42
treat	21
take	18
define	18
view	12
recognis/ze	12
…	…

The *as*-predicative: frequencies and why they don't help that much 1

Verb	Frequency in C	Frequency gen
see	111	1988
describe	88	259
regard	80	99
know	79	2120
use	42	1228
treat	21	92
take	18	1653
define	18	83
view	12	41
recognis/ze	12	114
…	…	…

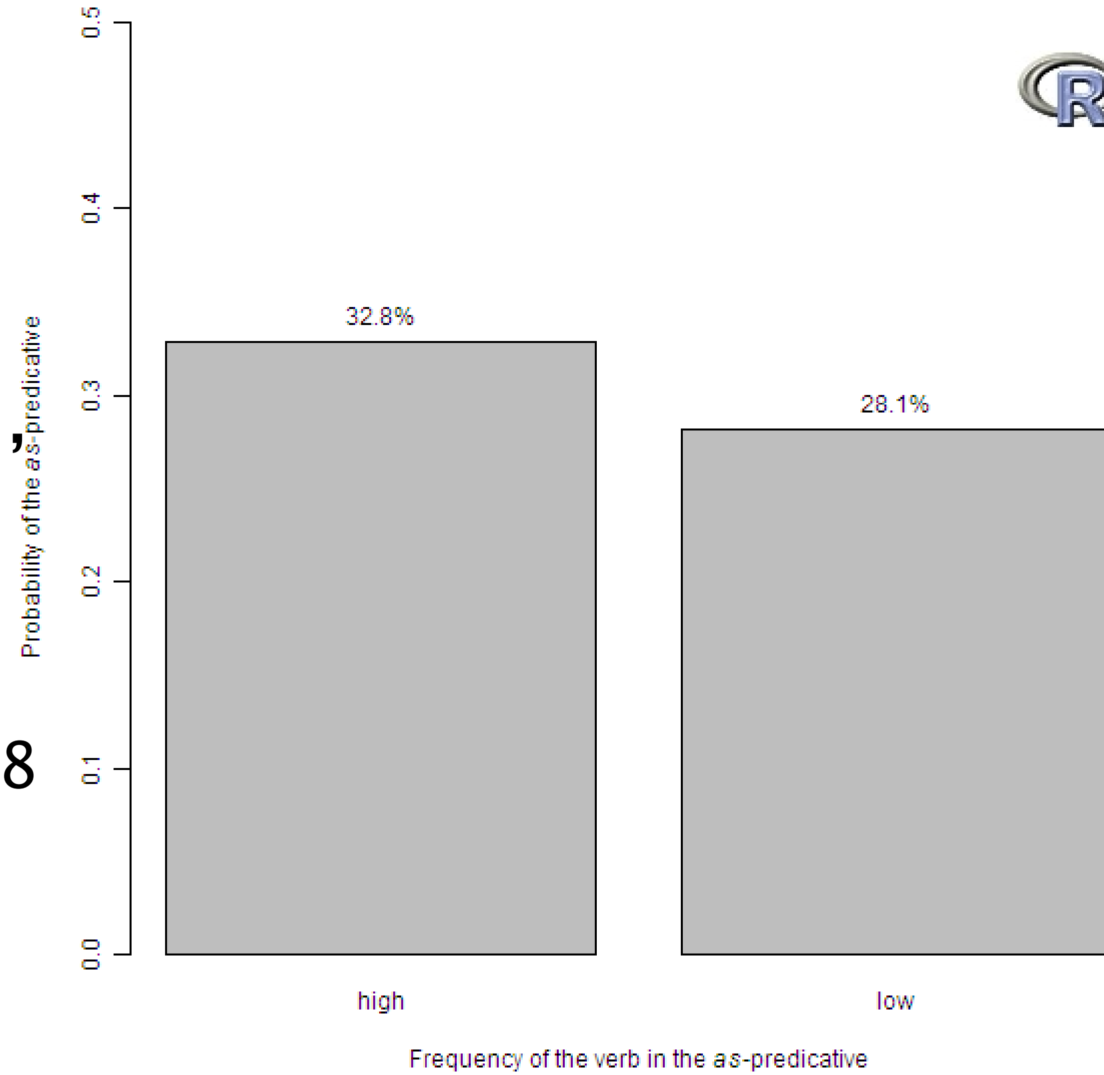
- It should not take long to see a problem with that approach: the frequencies in the pattern have not been normalized regarding the verbs' overall frequencies
- of course, *know*, *see*, and *take* occur in this pattern often – they are frequent verbs in general
- but maybe such frequent verbs are, *qua* their high frequency in the construction, still connected to this construction in speakers' minds …
- to test this, we ran an experiment

The *as*-predicative: the experimental design

- We chose one set of verbs that occurred in the *as*-predicative frequently, and one that occurred there infrequently
- we then created for each verb one active and one passive sentence fragment (because the *as*-predicative is strongly attracted to the passive)
  - *The biographer depicted the young philosopher* ________
  - *The young philosopher was depicted* ________
- subjects (64) were asked to complete the fragments
- the usual experimental controls: filler items with many other different verbs, pseudo-randomization of order of presentation, each subject saw each verb in only one fragment, saw only only one of each experimental condition, etc.

The *as*-predicative: the experimental results 1

- 493 responses that were unambiguously codable as *as*-predicatives or other constructions
- when the verb was from the high frequency group, the proportion of *as*-predicatives was indeed higher …
- … but not statistically significantly so: p=0.098 ns
- frequency of occurrence is not a good predictor



The *as*-predicative: frequencies and why they don't help that much 2

- Frequency of occurrence is not a good predictor, but there are other shortcomings, too
  - small frequencies quickly inflate percentages

	as-predicative	other construct.	totals
V: *re-elect*	1	0	1

  - one does not know the direction of the observed effect: is 0.03 more frequent than expected or less?

	as-predicative	other construct.	totals
V: *think of*	6	200	206

  - in fact, it is more frequent than expected because 0.03>0.005

with 6870, it would be less, though

	as-predicative	other construct.	totals
V: *think of*	6	200	206
	687	137977	138664

  - but how do we know the expected frequency, and what about the statistical significance of that difference?

Collexeme analysis: the table and how to compute expected frequencies

- The direction of your effect: is 7 (i.e., 1.518%) more frequent than expected or less?

	ditransitive	other construct.	totals
V: *bring*	7	454	461

- expected freq. of ditransitive *bring* = $\frac{1035}{138664} \times 461 \approx 3.44$
- since obs>exp, the ditransitive attracts *bring*
- is that difference – 7 observed vs. 3.44 expected – significant?
- no: $p_{\text{Fisher-Yates exact test}} \approx 0.06$

Collexeme analysis: the table and the significance of the distribution

- One out of many possible statistical tests than can be applied to such a table is the Fisher-Yates exact test

	as-predicative	other construct.	totals
V: *think of*	6	200	206
other verbs	681	137777	138458
totals	687	137977	138664

- $p_{\textit{Fisher-Yates exact test}}$=0.00062
- $-\log_{10}\ p_{\textit{Fisher-Yates exact test}}$=3.209
- if verbs are sorted according to this measure – which was called collexeme strength – does that make a difference?

R Console

File Edit Misc Packages Help

```
Press <Enter> to continue ...

Which kind of analysis do you want to perform?

1: collocational/ collostructional strength, i.e. collexeme analysis (cf. <1*.txt> for an example)
2: (multiple) distinctive collocates or distinctive collexeme analysis (cf. <2*.txt> for an example)
3: co-varying collexeme analysis (cf. <3*.txt> for an example)

Selection: 1

C o l l o c a t i o n a l / c o l l e x e m e   a n a l y s i s  . . .

This kind of analysis computes the degree of attraction and repulsion between
one word or construction and many other words using a user-defined statistic;
all these statistics are based on 2-by-2 tables, and attraction and repulsion
are indicated in a separate column in the output.

What is the word W / the name of the construction C you investigate (without spaces)?
1: as-predicative

Enter the size of the corpus (in constructions or words) without digit grouping symbols!
1: 138664

Enter the frequency of as-predicative in the corpus you investigate (without digit grouping symbols)
1: 687

Which index of association strength do you want to compute?

1: -log10 (Fisher-Yates exact, one-tailed) (= default)
2: log-likelihood
3: Mutual Information
4: Chi-square
5: log10 of odds ratio (adds 0.5 to each cell)

Selection: 1

How do you want to sort the output?

1: alphabetically
2: co-occurrence frequency
3: faith
4: collostruction strength

Selection: 4

Enter the number of decimals you'd like to see in the results (and '99', when you want the default output)!
1: 4

Where do you want the output ('text file' will append to already existing file with the same name)?

1: text file (= default)
2: terminal

Selection: 1

To compute the collocational strength of one word W to many other words <A, B, ..., ?>,
you need a text file with the following kind of table (with column names!):

Word    Freq_A-?_in_Corpus      Freq_A-?_&_W
A       ...                     ...
B       ...                     ...
...     ...                     ...

To compute the collostructional strength of one construction C to the words <A, B, ..., ?>,
you need a text file with the following kind of table (with column names!):

Word    Freq_A-?_in_Corpus      Freq_A-?_in_C
A       ...                     ...
B       ...                     ...
...     ...                     ...

Your table must not have decimal points/separators and ideally has no spaces (for the latter, use '_' instead)!
Also, don't forget that R's treatment of alphanumeric characters is case-sensitive!

Choose this text file with the raw data!        Press <Enter> to continue ...
```

R Console

File Edit Misc Packages Help

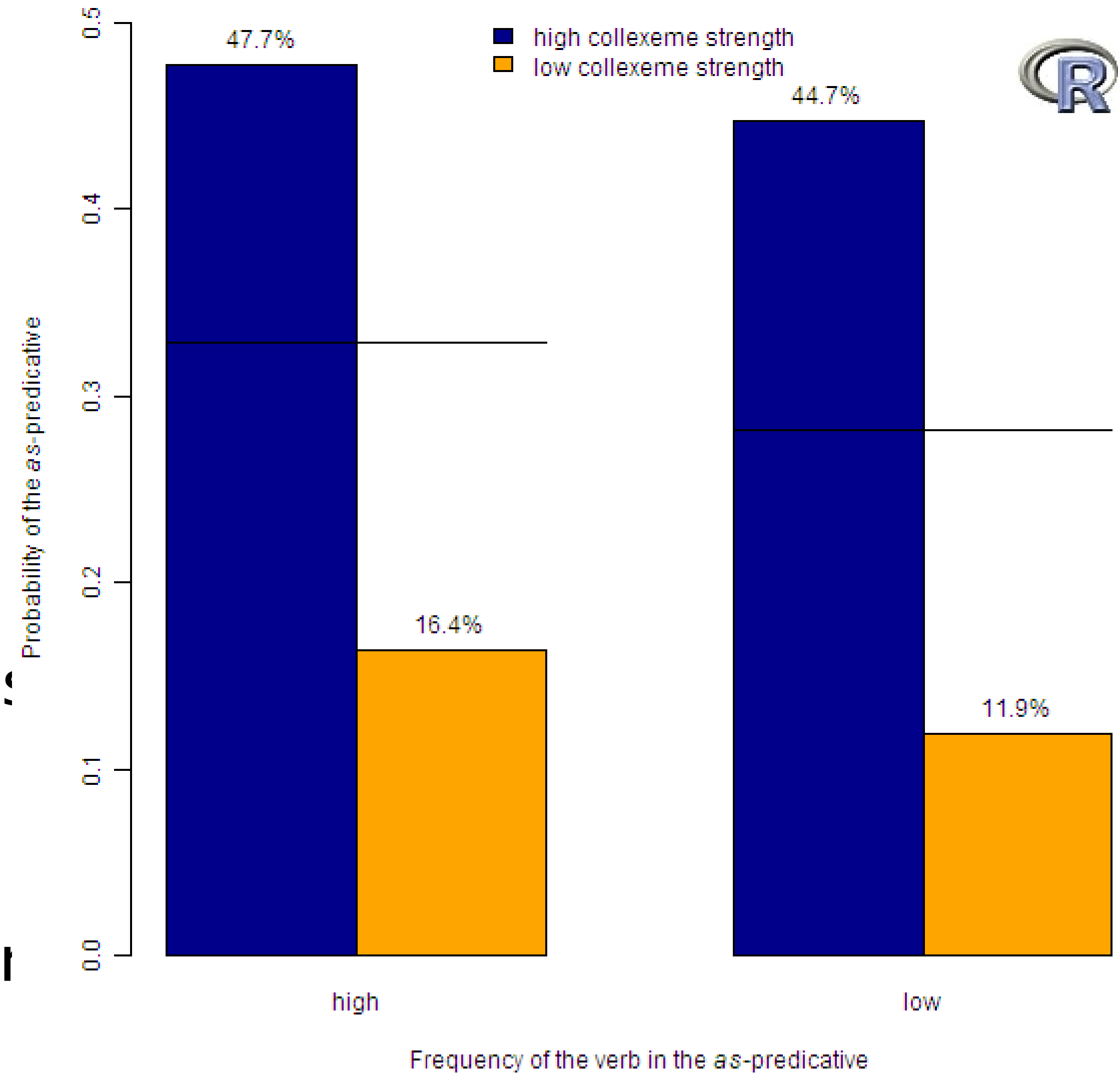
```
|--------------------------------------------------------------------|
| This output is provided without any warranty on an as-is basis by  |
| Stefan Th. Gries <http://www.linguistics.ucsb.edu/faculty/stgries/> |
| Please cite the program as mentioned in <readme.txt>. Thanks a lot! |
|--------------------------------------------------------------------|


word.freq: frequency of the word in the corpus
obs.freq: observed frequency of the word with/in  as-predicative
exp.freq: expected frequency of the word with/in as-predicative
faith: percentage of how many instances of the word occur with/in as-predicative
relation: relation of the word to as-predicative
coll.strength: index of collocational/collostructional strength: -log10 (Fisher-Yates exact, one-tailed), the higher, the stronger

          words word.freq obs.freq exp.freq   relation  faith delta.p.constr.to.word delta.p.word.to.constr coll.strength
1        regard        99       80   0.4905 attraction 0.8081                 0.1163                 0.8037      166.4757
2      describe       259       88   1.2832 attraction 0.3398                 0.1269                 0.3354      134.8705
3           see      1988      111   9.8494 attraction 0.0558                 0.1480                 0.0516       78.7897
4          know      2120       79  10.5034 attraction 0.0373                 0.1002                 0.0328       42.7964
5         treat        92       21   0.4558 attraction 0.2283                 0.0301                 0.2235       28.2235
6        define        83       18   0.4112 attraction 0.2169                 0.0257                 0.2120       23.8431
7           use      1228       42   6.0840 attraction 0.0342                 0.0525                 0.0295       21.4248
8          view        41       12   0.2031 attraction 0.2927                 0.0173                 0.2878       17.8610
9           map        23        8   0.1140 attraction 0.3478                 0.0115                 0.3429       12.7957
10    recognis|ze      114      12   0.5648 attraction 0.1053                 0.0167                 0.1004       12.1590
11   categoris|ze       10       6   0.0495 attraction 0.6000                 0.0087                 0.5951       11.5246
12     perceive        28        6   0.1387 attraction 0.2143                 0.0086                 0.2094        8.3037
13         hail         4        3   0.0198 attraction 0.7500                 0.0044                 0.7451        6.3165
14      appoint        35        5   0.1734 attraction 0.1429                 0.0071                 0.1379        6.0734
15    interpret        35        5   0.1734 attraction 0.1429                 0.0071                 0.1379        6.0734
16        class         5        3   0.0248 attraction 0.6000                 0.0044                 0.5951        5.9201
17     denounce         7        3   0.0347 attraction 0.4286                 0.0043                 0.4236        5.3793
18      dismiss        25        4   0.1239 attraction 0.1600                 0.0057                 0.1551        5.1577
19     consider       264        9   1.3080 attraction 0.0341                 0.0113                 0.0292        5.0789
20       accept       178        7   0.8819 attraction 0.0393                 0.0089                 0.0344        4.4673
21         name        41        4   0.2031 attraction 0.0976                 0.0056                 0.0926        4.2816
22      portray        19        3   0.0941 attraction 0.1579                 0.0043                 0.1530        3.9563
36    advert_to         4        2   0.0198 attraction 0.5000                 0.0029                 0.4951        3.8354
23     diagnose         6        2   0.0297 attraction 0.3333                 0.0029                 0.3284        3.4403
24     think_of       206        6   1.0206 attraction 0.0291                 0.0073                 0.0242        3.2088
25       depict         8        2   0.0396 attraction 0.2500                 0.0029                 0.2451        3.1721
```

The *as*-predicative: the experimental results 2

- The difference in the ranking seems small, but what happens in the experiment?
- the verbs used in the experiment did not just come from two frequency groups (hi/lo), but also from two collexeme-strength groups (hi/lo)
- the effect of collexeme strength is highly significant and more than 3.8 times as strong as that of frequency



The *as*-predicative: the experimental results 3

Variable	Level	Fixed Eff.	*se*	*z*	*p*	
(Intercept)		0.22	0.67	0.33	0.743	
Voice	passive	-0.78	0.3	-2.62	0.009	**
Collexeme strength	lo	-7.06	1.87	-3.77	<0.001	***
Frequency	lo	-1.84	1.11	-1.66	0.098	.
Faith		0.26	1.9	0.14	0.892	
Voice X Collexeme strength	passive/lo	1.66	0.57	2.91	0.004	**
Collexeme strength X Frequency	lo/lo	6.58	2.41	2.73	0.006	**
Collexeme strength X Faith	lo	217.49	81.64	2.66	0.008	**
Collexeme strength X Frequency X Faith	hi/lo	5	2.94	1.7	0.088	.
Collexeme strength X Frequency X Faith	lo/lo	-289.52	108.31	-2.67	0.008	**

Variable	*var*	*sd*
Verb-specific effects	1.4	1.18

The *as*-predicative: results from an additional self-paced reading task

Variable	F	effect size	$p_{two\text{-}tailed}$	
TokenFrequency	0.257	0.001	0.612	ns
Voice	0.180	0.001	0.672	ns
Collexeme strength	3.438	0.014	0.065	ms
Frequency	1.111	0.005	0.293	ns
Voice X Collexeme strength	0.021	0.000	0.886	ns
Voice X Frequency	0.053	0.000	0.819	ns
Collexeme strength X Frequency	0.609	0.002	0.436	ns
Voice X Collexeme strength X Frequency	0.622	0.003	0.431	ns

Advantages of this approach

- This approach has many general advantages
  - it is in line with the assumption of no strict dichotomy between syntax and lexis (in using a statistical approach from lexical collocations)
  - it is descriptively more adequate (in downtoning the effect of overall frequent words and providing the direction of an effect)
  - it provides an impression of the robustness of the statistics
  - it has a larger degree of predictive power (at least in these experiments, for more see next talk)

Distinctive collexeme analysis

- Collexeme analysis/strength quantifies the degree of association – attraction or repulsion – between
  - one word and
  - one slot in one pattern/construction
- distinctive collexeme analysis quantifies the degree of association
  - of one word to
  - one slot in 2+ patterns/constructions
    - 2 patterns: *will*-future vs. *going to*-future, active vs. *be*-passive
    - 3 patterns: *will*-future vs. *going to*-future vs. *shall*, active vs. *be*-passive vs. *get*-passive

Distinctive collexeme analysis: expected frequencies and significance test 1

- It is based on a similar kind of co-occurrence table as the simple collexeme analysis

	ditransitive	*to*-dative	totals
V: *award*	7	3	10
other verbs	1028	1916	2944
totals	1035	1919	2954

- the main changes
  - the two columns are now the two 'competing' patterns
  - the overall *n* is now the sum of both 'competing' patterns (and not some corpus size anymore)
- $p_{\textit{Fisher-Yates exact test}}$=0.026
- $-\log_{10}\ p_{\textit{Fisher-Yates exact test}}$=1.585
- then verbs are sorted according to this measure, which was called distinctive collexeme strength

Distinctive collexeme analysis: expected frequencies and significance test 2

- The frequencies of *bring* in the ditransitive and the *to*-dative

	ditransitive	*to*-dative	totals
V: *bring*	7	82	89
other verbs	951	732	1683
totals	958	814	1772

- expected freq. of ditransitive *bring* = $\frac{958 \times 89}{1772} \approx 48.12$
- since obs<exp, *bring* repels the ditransitive and attracts the *to*-dative
- is that difference – 7 observed vs. 48.12 expected – significant?
- yes: $p_{\text{Fisher-Yates exact test}} \approx 2.672962\text{e-}21$ (which is 0.000000000000000000002672962) and its log=20.57

Distinctive collexeme analysis: an application to the dative alternation

- When this method is applied to the dative alternation, it returns all the verbs that encode characteristic senses of the patterns
  - ditransitive
    - change of possession, communication, perceiving as receiving, satisfaction condition, cause not to receive, etc.
    - many with typically close contact between AGT & REC
  - *to*-dative
    - caused accompanied motion
    - many with transfer over a distance

Ditransitive		***to*-dative**	
VERB	$\log_{10} p$	VERB	$\log_{10} p$
give	119.74	*bring*	8.83
tell	57.06	*play*	5.84
show	11.08	*take*	3.74
offer	9	*pass*	3.65
allow	8.24	*make*	2.17
cost	8.01	*sell*	1.86
teach	5.83	*do*	1.82
buy	4.11	*supply*	1.54
wish	3.27	*read*	1.22
earn	3.19	*feed*	1.07

Distinctive collexeme analysis: many more applications

- This way of looking at grammatical patterns has many applications
  - synchronic complementation patterns
    - Gilquin (2006): periphrastic causatives in English
    - Wulff (2006): *go and* V vs. *go* V
    - Hommerberg & Tottie (2007): *try to* vs. *try and*
    - Wulff (2008): *go/come/try and* V vs. *go/come/try* V
  - modification of hedges: Gries and David (2007)
  - diachronic language change
    - Hilpert (2006): verbal complementation of *shall*
    - Hilpert (2008): Germanic future constructions
    - Margerie (2008): *fairly*
  - cultural differences
    - Wulff, Stefanowitsch, & Gries (2007): British vs. American *into*-causative using (physical) force and pressure vs. communication verbs

Distinctive collexeme analysis: many more applications

- This way of looking at grammatical patterns has many applications
  - second/foreign language acquisition
    - alternation preferences of second/foreign language learners: Gries & Wulff (2005) on the dative alternation and Gries & Wulff (2009) on *to* vs. *ing* complementation
  - psycholinguistic processing
    - language comprehension: collexeme strengths and eye-tracking regarding verb subcategorization preferences in online sentence comprehension: Wiechmann (2008)
    - language production: lexical effects on syntactic persistence: Gries (2005), Szmrecsanyi (2005, 2006)
- (more on this later)

Covarying collexeme analysis

- Collexeme analysis/strength quantifies the degree of association – attraction or repulsion – between
  - one word and
  - one slot in one pattern/construction
- distinctive collexeme analysis quantifies the degree of association
  - of one word to
  - one slot in 2+ patterns/constructions
    - 2 patterns: *will*-future vs. *going to*-future, active vs. *be*-passive
    - 3 patterns: *will*-future vs. *going to*-future vs. *shall*, active vs. *be*-passive vs. *get*-passive
- covarying collexeme analysis quantifies the degree of association
  - of one word in one slot of a construction to
  - another word in another slot of the same construction

Quantitative corpus linguistics: covarying collexeme analysis ('methods')

- The frequencies of *blackmail* (or others) into *accepting* (or others) in the *into*-causative

	V: *accepting*	other verbs	totals
V: *blackmail*	14	37	51
other verbs	8	141	149
totals	22	178	200

- exp. freq. of *blackmail into accepting* = $\frac{22 \times 51}{200} \approx 5.61$
- since obs<exp, *blackmail* attracts *accepting*
- is that difference – 14 observed vs. 5.61 expected – significant?
- yes: $p_{\text{Fisher-Yates exact test}} \approx 6.60712e\text{-}21$ and its log=20.18

Results of a co-varying collexeme analysis for the *into*-causative

```
Co-varying collexeme analysis for: into-causative

words1: words in the 1st slot of into-causative
words2: words in the 2nd slot of into-causative
freq.w1: frequency of word1 in into-causative
freq.w2: frequency of word2 in into-causative
obs.w1_2.in_c: observed frequency of both words in both slots in into-causative
exp.w1_2.in_c: expected frequency of both words in both slots in into-causative
relation: relation between observed and expected frequency
coll.strength: index of co-varying collexeme strength: -log(Fisher exact, 10) , the higher, the stronger
```

	words1	words2	freq.w1	freq.w2	obs.w1_2.in_c	exp.w1_2.in_c	relation	coll.strength
1	talk	letting	670	127	50	8.72	attraction	25.2505606
2	talk	surrendering	670	65	31	4.46	attraction	18.7824311
3	talk	staying	670	90	36	6.18	attraction	18.5614734
4	torture	confessing	27	57	11	0.16	attraction	17.9263322
5	bounce	accepting	112	402	29	4.62	attraction	15.2071148
6	force	making	607	513	74	31.92	attraction	11.3808449
7	talk	going	670	146	36	10.03	attraction	11.1015213
8	dupe	buying	235	441	37	10.62	attraction	10.6870160
9	draw	commenting	107	8	6	0.09	attraction	10.3806688
10	coerce	having	315	135	23	4.36	attraction	10.3369707
11	terroris|ze	fleeing	48	15	6	0.07	attraction	10.3032562
12	shock	understanding	79	20	7	0.16	attraction	9.9058960
13	plunge	mourning	6	3	3	0.00	attraction	9.8882334
14	stimulate	producing	16	59	6	0.10	attraction	9.5400911
15	fool	seeing	127	48	10	0.62	attraction	9.3751673
16	pressure	pleading	615	29	14	1.83	attraction	9.3612485
17	shame	cleaning	156	15	7	0.24	attraction	8.8689809
18	talk	coming	670	73	22	5.01	attraction	8.7764782
19	mislead	buying	74	441	18	3.35	attraction	8.5112565
20	parlay	landing	7	7	3	0.01	attraction	8.1016619
21	delude	supposing	8	10	3	0.01	attraction	7.3630637
22	entice	buying	97	441	19	4.39	attraction	7.2925849
23	frighten	fleeing	164	15	6	0.25	attraction	7.0407731
24	dupe	carrying	235	36	9	0.87	attraction	6.9019070
25	lock	using	50	85	7	0.44	attraction	6.6592127
26	bounce	announcing	112	24	6	0.28	attraction	6.6423643
27	pressure	having	615	135	26	8.51	attraction	6.6049524
28	hoodwink	leaving	49	139	8	0.70	attraction	6.4167949
29	lull	expecting	52	14	4	0.07	attraction	6.1604223
30	con	paying	168	245	17	4.22	attraction	5.9938877

Results of a co-varying collexeme analysis for the *into*-causative

- General semantics of the *into*-causative
  - the agent forces/tricks the patient into doing an activity the patient would not normally want to do
- general characteristics of the pairs of
  - the forcing/tricking verbs in cause slots
  - the forced/tricked-activity verbs in result/effect slots
  - they instantiate culturally-specific frames of entrenched cause-effect relationships
    - commercial-transaction frame, confession frame, …
- culture-specific differences – let' test AmE vs. BrE
  - causes
    - AmE: communication, physical force; patient is restricted
    - BrE: stimulation, negative emotion, threaten, physical force; patient is set into motion
  - effects
    - AmE: light verbs
    - BrE: communication

Summary

- Collostructional analysis
  - extends the notion of association measures to the domain of associations in(volving) constructions
    - collexeme analysis: words and cxs
    - distinctive collexeme analysis: words to 1 of 2 cxs
    - multiple distinctive collexeme analysis: words to 1 of 3 cxs
    - co-varying collexeme analysis: words to words in 1 cxs
  - involves observed co-occurrence frequencies
  - rewards high co-occurrence frequencies
  - but normalizes observed co-occurrence frequencies against expected co-occurrence frequencies using any association measure (usually p_{FYE})
  - has garnered experimental support
    - sentence-completion task
    - self-paced reading task
- how can one not like this?

Thank you!

http://tinyurl.com/stgries