Detecting Selectional Behavior of Complex Types in Text

Anna Rumshisky
Victor Grinberg
James Pustejovsky

Brandeis University

Brief Outline

(1) Some aspects of selectional behavior of complex types

(2) Automatic methodology for detecting some of that behavior

Complex Types

- Complex types in GL are a mechanism for dealing with selectional behavior of nouns
- Contexts in which complex types occur may select for
  - any of the component types
  - the complex type itself

newspaper-n ((PHYS * INFO) * COMPANY)

He folded the newspaper carefully (PHYS)

He doesn't believe the newspapers (INFO)

He reads the newspaper every day (PHYS * INFO)

He tried to sue the newspaper for calling him a liar (COMPANY)

Selector Contexts

- Usually, something in the context tells you what is is

lunch-n (EVENT * FOOD)

> I have my *lunch* in the backpack (FOOD)
>
> Your *lunch* today was longer than usual (EVENT)

- Typically, predicates or modifiers
- Selector contexts may be quite similar, and yet select for different types

newspaper-n ((PHYS * INFO) * COMPANY)

> I finished this *newspaper* two hours ago (PHYS * INFO)
>
> I started this *newspaper* two years ago (COMPANY)

Argument Position Asymmetries

- A certain reading may be preferred in a certain argument position
  - E.g. for ANIMAL · FOOD nouns the subject position tends to disprefer the FOOD sense
  - Consistent with the systematic relation between senses (where each sense corresponds to one component type)

chicken-n

subject

a. ANIMAL: peck, look, wander, come, cross, follow, die

object

a. ANIMAL: count, chase, kill, shoot, slaughter, skin, pluck, sacrifice

b. FOOD: eat, serve, prefer, turn, dip, stuff, carve, baste, roast, simmer

c. ANIMAL · FOOD: poach, cook

Multiple Selection

- Complex types also allow for <u>multiple selection</u>

  We had a *delicious* (FOOD), *leisurely* (EVENT) lunch

  He *finished* (PHYS * INFO) the newspaper and *folded* (PHYS) it carefully

- Different selector contexts for the same dot object select for different component types of the dot

- Selectors may or may not be syntactically similar

  That *three-course* (FOOD) lunch sure *took forever* (EVENT).

Clustering Task

- The usual notion of *word sense disambiguation* (single sense per occurrence) may be difficult to apply to dot objects
- But it is often clear which type a particular selector prefers:

lunch-n

object

a. FOOD: eat, cook, enjoy, prepare, take, bring, etc.

b. EVENT: skip, finish, attend, miss, host, cancel, etc.

adjectival modifier

a. FOOD: light, delicious, three-course, excellent, liquid, home-cooked, half-eaten, heavy, substantial, etc.

b. EVENT: leisurely, early, annual, celebratory, official, private, weekly, etc.

- We would like to do this automatically (!)

Distributional Similarity

- Typically, such tasks are addressed using the notion of distributional similarity
  - Get all the contexts in which the word occurs
  - Compare contexts for different words
- Context gets represented as a feature vector

  $<(\text{feature}_i, \text{value}_i)> = <(\text{feature}_1, \text{value}_1), (\text{feature}_2, \text{value}_2), \ldots>$

- Each feature corresponds to some element or parameter of the context
  - bag of words; populated grammatical relations
- Measure how close two words (e.g. eat-v, cook-v) are distributionally
  - e.g. cosine between vectors; other measures of how often words occur in similar contexts

Distributional Similarity

- Can we use it?
  - In our task, selector contexts do not need to be distributionally similar
  - They only need to be <u>similar in context</u>

    (= activate the same component type)

  construction-n (EVENT * PHYS)

  object

    a. EVENT: finance, oversee, complete, supervise, halt

    b. PHYS: examine, build, photograph

- *Generic distributional similarity* may be low

  sim(finance-v, complete-v); *sim*(build-v, fotograph-v)

- *Contextualized similarity* must be high

  c_sim(finance-v, complete-v, (construction-n, object_of))

Reformulating Our Task

- Must group together selectors that occur <u>in the same grammatical relation</u> with the target according to the type they select
  - i.e. <u>activate the same component type of the target noun</u>
- These selectors must be similar with respect to the target noun *w* and that grammatical relation *R*
  - i.e. similar with respect to the target context (*w, R*)
  - e.g. (*w, R*) = (lunch, object_of)
  - *c_sim*(organize-v, miss-v, (lunch, object_of))

Reformulating Our Task

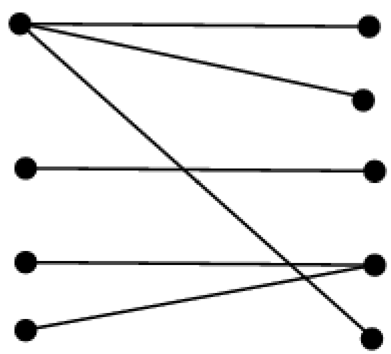
- We want to be able to tell that organize-v and miss-v select the same component type for lunch-n
- **How do we do that?**
  - We're going to look at all the contexts in which the target occurs, and find other words <u>like the target</u> that occur with the same selectors
    - all things that can be eaten, cooked, organized, and missed
  - Then cluster them according to which “sense” of lunch they behave closer to

Clustering with Inverse Image

- lunch-n (EVENT / FOOD):

conference-n	attend-v
meeting-n	hold-v
seminar-n	miss-v
parade-n	organize-v
rehearsal-n	cancel-v
wedding-n	host-v
...	...
sandwich-n	eat-v
stew-n	cook-v
pudding-n	serve-v
meat-n	prepare-v

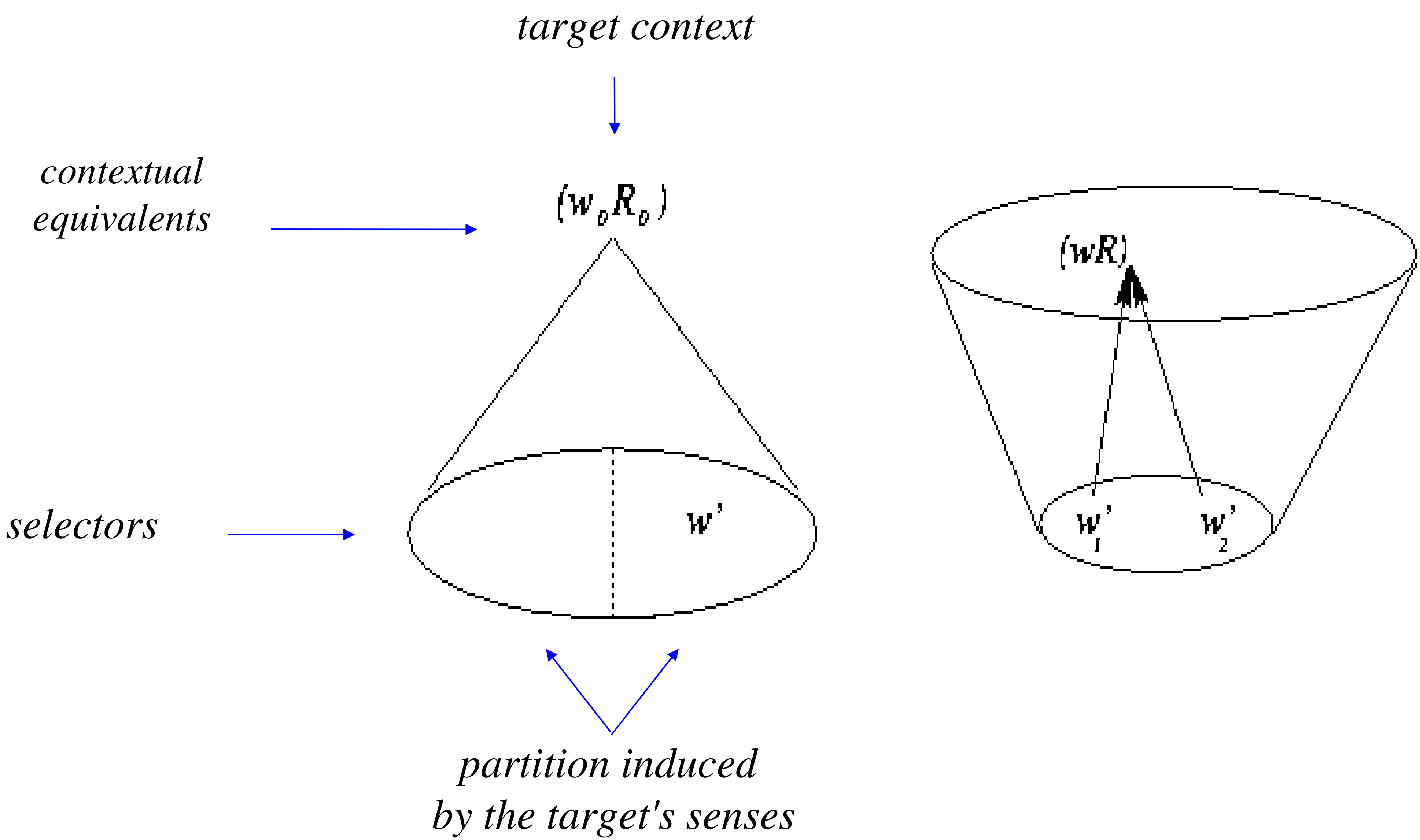
Think about it as a bipartite graph:



What are we doing?

- When computing semantic similarity based on distributional behavior, some contexts are “more equal than others”.
- Effectively, we are grouping together *licensing contexts*
  - they license similar use of selectors with respect to the target
  - simply put, it is the fact that you can cancel meetings, seminars, and weddings makes cancel select the EVENT aspect of lunch
- Thus, our goal is to obtain clusters of *contextual equivalents* for each component type (= “sense”) of the target word.

Clustering with Inverse Image



Clustering Algorithm-I

(1) Identify the set of selector contexts in which the target word was found in corpus.

(2) Take the inverse image of the above set under grammatical R^{-1}. A word is considered a potential contextual equivalent if it occurs in R^{-1} with the specified number of the target's selectors. (thres = 5)

(3) Obtain a set of "good" selectors for each potential contextual equivalent

   i. Take all the selector contexts in which both the target and the contextual equivalent are found. Compute two conditional probability scores for each selector: P(s | R w) and P(s | R t), where *s* is the selector context, *w* is the potential contextual equivalent, and *t* is the target word.

Clustering Algorithm-II

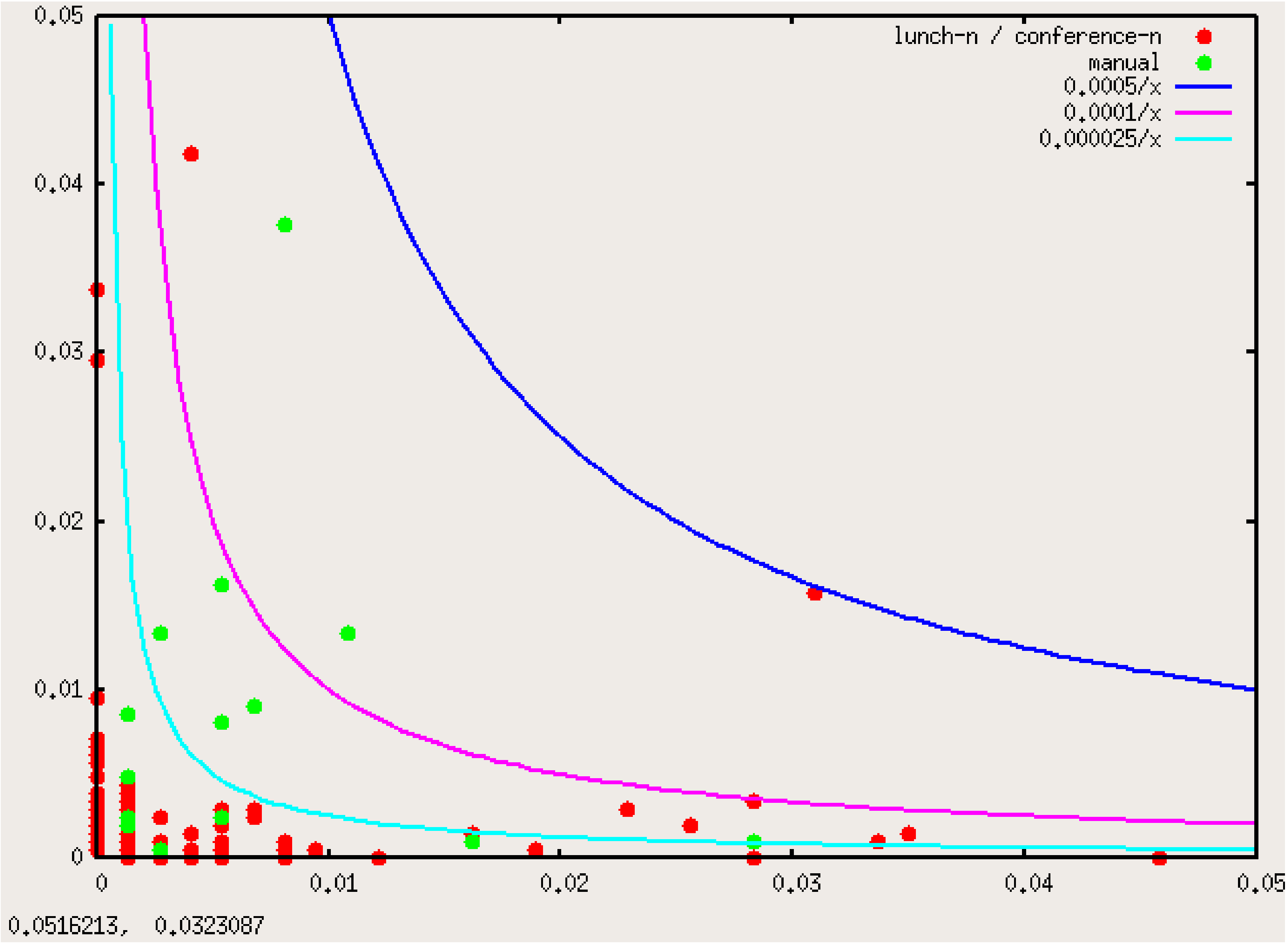
ii. Identify selectors that select the same interpretation both for the target noun and for potential contextual equivalent.

E.g. Given the target (lunch-n, object_of): for sandwich-n, we would need to select eat-v, cook-v, serve-v, prepare-v, etc; for conference-n, we would need to select attend-v, organize-v, miss-v, cancel-v, etc.

- "Good" selectors will must occur often enough with both words (modeled as having both *conditional probabilities* relatively high).
- NB Conditional probability will depend on how frequent the appropriate sense is for each of the two words.
- We pick the top-K selectors that maximize the *geometric mean* of two conditional probability values. K = 20

Choosing selectors for lunch-n/conference-n

(with R = object_of)



Choosing Good Selectors

- We now have a contextualized list of selectors for each potential candidate for contextual equivalency, with the appropriate association scores

Choosing Good Selectors

	lunch-n		sandwich-n	
	count	P(s|Rw)	count	P(s|Rw)
'eat-v'	93	.1253	93	.2035
'take-v'	48	.0647	30	.0656
'get-v'	40	.0539	25	.0547
'make-v'	17	.0229	56	.1225
'want-v'	19	.0256	17	.0372
'bring-v'	21	.0283	13	.0284
'finish-v'	21	.0283	8	.0175
'buy-v'	14	.0189	12	.0263
'prepare-v'	21	.0283	7	.0153
'serve-v'	42	.0566	3	.0066

Choosing Good Selectors

	lunch-n		conference-n	
	count	P(s|Rw)	count	P(s|Rw)
'attend-v'	15	.0202	263	.1251
'hold-v'	10	.0135	379	.1803
'give-v'	23	.0310	33	.0157
'tell-v'	2	.0027	285	.1356
'organize-v'	6	.0081	79	.0376
'take-v'	48	.0647	6	.0029
'call-v'	3	.0040	88	.0419
'arrange-v'	8	.0108	28	.0133
'get-v'	40	.0539	4	.0019
'bring-v'	21	.0283	7	.0033

Clustering Algorithm-III

(4) Compute the similarity matrix for the potential contexual equivalents.

- We compute the similarity measure as the sum of minima, which is effectively equivalent to set-theoretic overlap used in Jaccard and Dice measures.

$$c\_sim(w_1, w_2, (t, R)) = \Sigma_{s \in S' \cup S''} \min(P(s \mid R\ w_1), P(s \mid R\ w_2))$$

where $S' = Selectors^K(w_1)$ and $S'' = Selectors^K(w_2)$ are the contextualized selector lists chosen in the previous step.

	sandwich-n	conference-n	rehearsal-n
	P(s|Rw)	P(s|Rw)	P(s|Rw)
'attend-v'	.0000	.0129	.0145
'cancel-v'	.0000	.0108	.0017
'hold-v'	.0066	.0108	.0017
'eat-v'	.0005	.0132	.0009
'serve-v'	.0000	.0528	.0660

Clustering Algorithm-IV

- Unlike the standard numerical extensions of Jaccard and Dice, we do not normalize the sum of minima either by the size of the union $|S' \cup S''|$, or by the average size of each selector set $(|S'| + |S''|)/ 2$
- This allows us to avoid having high similarity scores for high-frequency words among candidates for contextual equivalentcy
  - E.g. man-n and thing-n (or variety-n and range-n) are frequent and omnivorous, and so any of the “good” selectors for the target lunch-n would also be “good” for both of them
  - If they are the closest pair amongst all candidates for contextual equivalency – they would merge first in clustering and immediately contaminate the clusters

Clustering Algorithm-V

- These are effectively promiscuous words that occur frequently with all selectors, including the “good” (i.e. reliable) selectors for each of target's senses.
- But conditional probabilities for their selectors are low due to their high frequencies. The sum of their minuma will also be low.
- Thus, normalizing the sum of minima by any value also reflecting this high frequency will remove the advantage that less promiscuous words have over such generics.

Clustering Algorithm-VI

(5) Perform agglomerative hierarchical clustering of potential contextual equivalents of the target's senses, using this similarity metric obtained as described above.

(6) Compute Average Pairwise Similarity (APS) between the elements of each cluster. As we proceed from the bottom of the dendrogram up, APS for the clusters decreases.

- We compute the percent decrease in APS (APS derivative) for every cluster merge point

(7) Select several seed elements from the high-scoring selectors (e.g. MI-based) of the target and trace their merges in the dendrogram

- Cut the dendrogram at the points that have high percent decrease, selecting the clusters obtained prior to the APS-decreasing merge.
- If selected clusters coincide for several seeds, select those clusters

Target lunch-n / Seed conference-n

num	*id*	*inter-c APS*	*intra-c APS*	*APS jump*	*resulting cluster*
1	3985	0.445	0.445	0.00	[conference-n] [seminar-n]
2	4012	0.430	0.435	0.02	[meeting-n] [conference-n seminar-n]
3	4097	0.397	0.416	0.04	[rally-n] [meeting-n conference-n seminar-n]
4	4263	0.342	0.387	0.07	[reunion-n] [rally-n meeting-n conference-n seminar-n]
5	4394	0.314	0.363	0.06	[ceremony-n] [reunion-n rally-n meeting-n conference-n seminar-n]
6	4493	0.295	0.332	0.09	[inquest-n fair-n] [ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]
7	4656	0.267	0.318	0.04	[congress-n] [inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]
8	4674	0.264	0.307	0.03	[disco-n] [congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]
9	4784	0.246	0.280	0.09	[talk-n ballot-n election-n referendum-n] [disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]
10	5001	0.223	0.272	0.03	[summit-n] [talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]
11	5072	0.216	0.265	0.03	[hearing-n] [summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n]

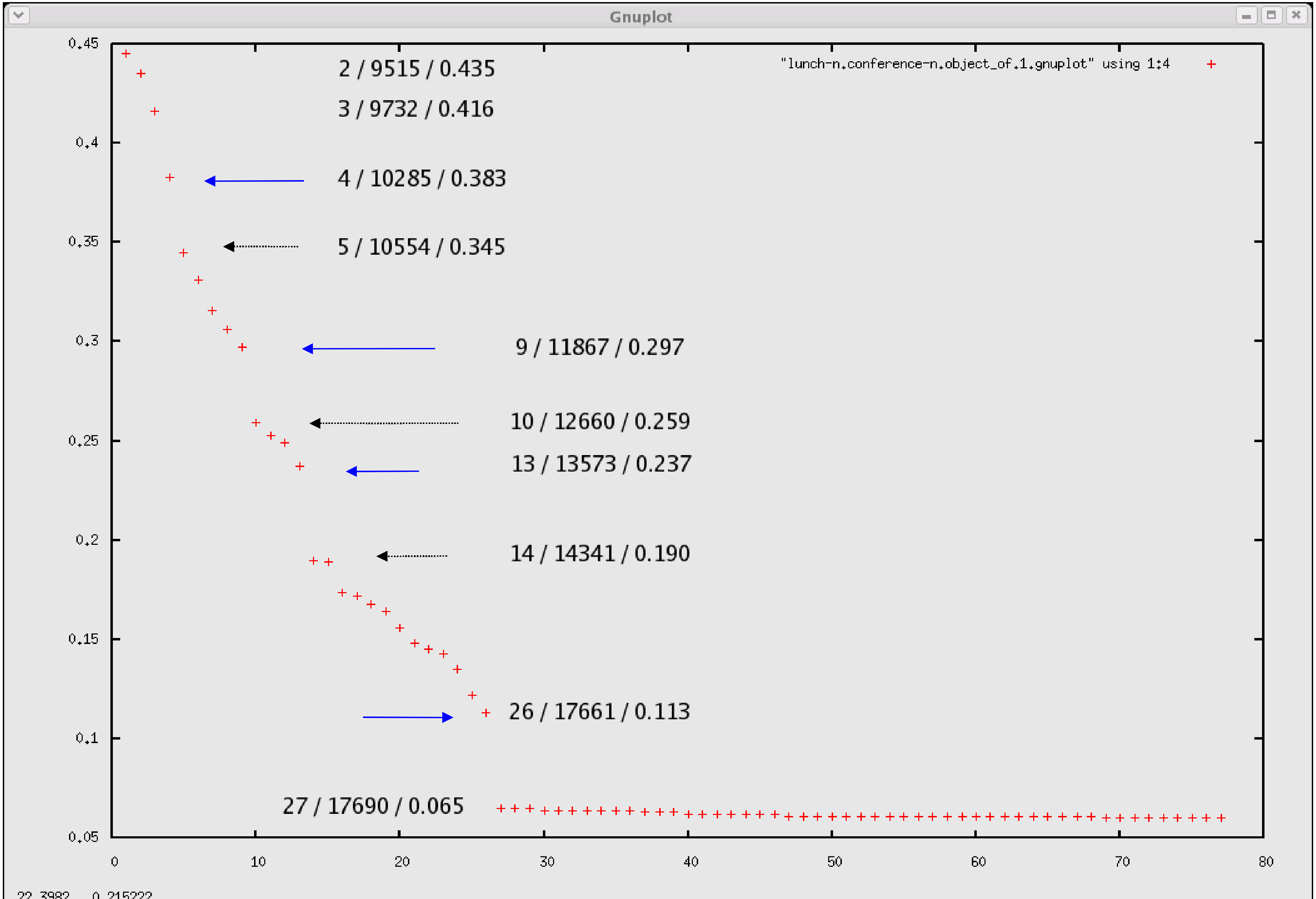
Target lunch-n / Seed conference-n

num	*id*	*inter-c APS*	*intra-c APS*	*APS jump*	*resulting cluster*
12	5294	0.197	0.224	0.15	[hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n] [parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n]
13	5316	0.196	0.219	0.02	[outing-n barbecue-n exhibition-n festival-n] [hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n]
14	5631	0.174	0.214	0.02	[outing-n barbecue-n exhibition-n festival-n hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n] [dance-n gathering-n]

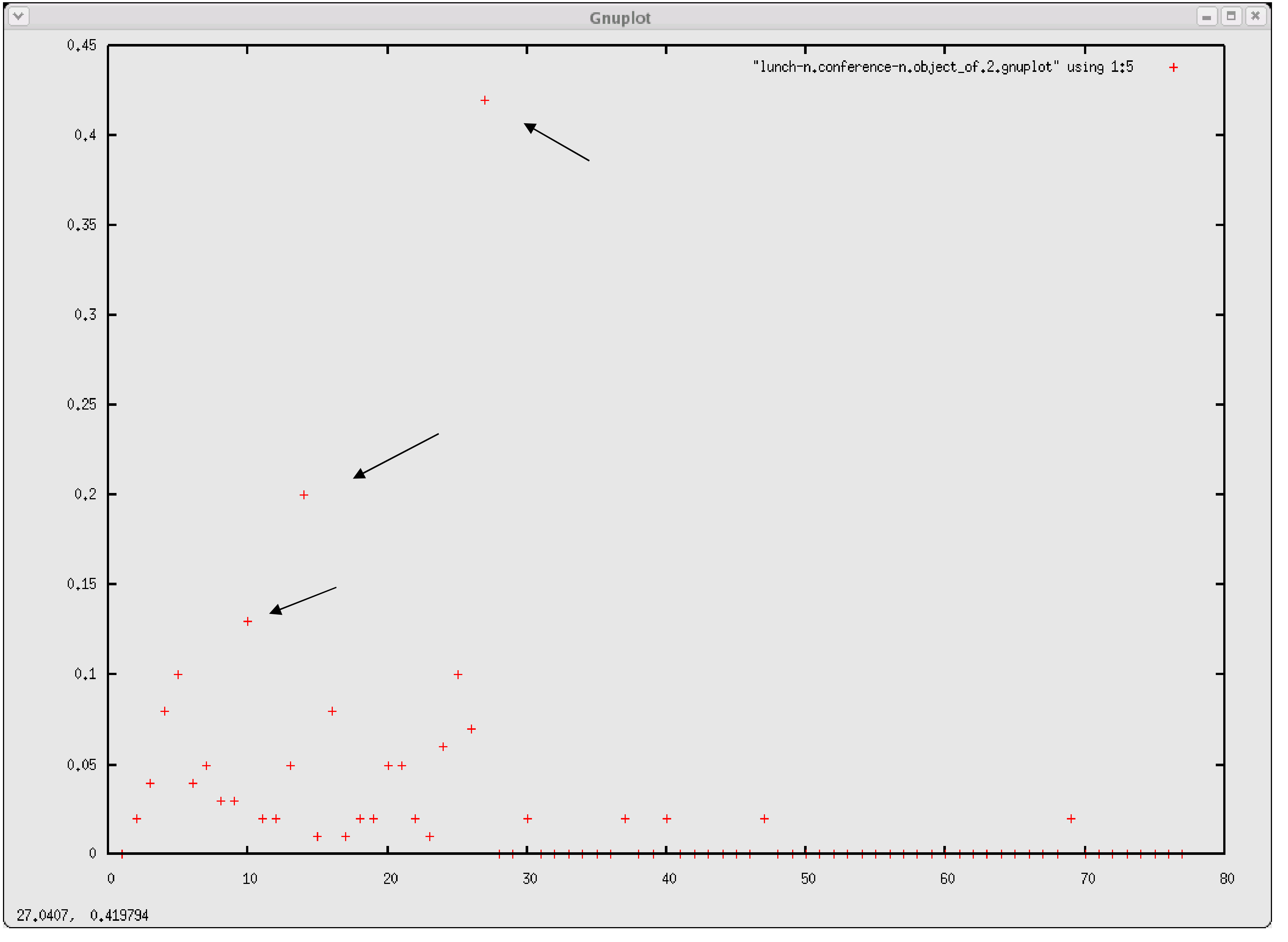
Target lunch-n / Seed conference-n

num	*id*	*inter-c APS*	*intra-c APS*	*APS jump*	*resulting cluster*
15	6299	0.138	0.206	0.04	[outing-n barbecue-n exhibition-n festival-n hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n dance-n gathering-n] [event-n procession-n]
16	6347	0.136	0.200	0.03	[tournament-n contest-n] [outing-n barbecue-n exhibition-n festival-n hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n dance-n gathering-n event-n procession-n]
17	6401	0.134	0.185	0.08	[candle-n captive-n portfolio-n office-n post-n presidency-n] [tournament-n contest-n outing-n barbecue-n exhibition-n festival-n hearing-n summit-n talk-n ballot-n election-n referendum-n disco-n congress-n inquest-n fair-n ceremony-n reunion-n rally-n meeting-n conference-n seminar-n parade-n rehearsal-n wedding-n funeral-n clinic-n feast-n celebration-n session-n workshop-n demonstration-n concert-n briefing-n lecture-n reception-n banquet-n luncheon-n dance-n gathering-n event-n procession-n]

Intra-cluster APS for lunch-n / conference-n



Intra-cluster APS % increase (derivative) for lunch-n / conference-n



Lunch-n

- Food

  ['juice-n','cocktail-n','alcohol-n','wine-n','ale-n','brandy-n','vodka-n','champagne-n', 'beer-n','pint-n','whiskey-n','gin-n','sherry-n','straw-n','corn-n','liver-n','cereal-n', 'goose-n','vegetable-n','rice-n','pasta-n','stuffing-n','dish-n','tomato-n','pea-n', 'bean-n','ham-n','turkey-n','mushroom-n','potato-n','chicken-n','carrot-n','bacon-n', 'cabbage-n','nut-n','apple-n','orange-n','lettuce-n','dessert-n','chip-n','food-n', 'snack-n','buffet-n','steak-n','salad-n','sandwich-n','dinner-n','meal-n','lunch-n', 'breakfast-n','supper-n','beef-n','sweet-n','crisp-n','chop-n','sausage-n','pizza-n', 'meat-n','chocolate-n','banana-n','spaghetti-n','yogurt-n','ice-cream-n','donut-n','mint-n','honey-n','jam-n','soup-n','toast-n','tea-n','coffee-n','bread-n','cheese-n','cake-n','curry-n','bun-n','biscuit-n','pudding-n','marmalade-n','jelly-n','pie-n','porridge-n','tart-n','pastry-n','stew-n','sauce-n','hay-n','butter-n','roll-n','cream-n']

- Event

  ['tournament-n','contest-n','outing-n','barbecue-n','exhibition-n','festival-n','hearing-n','summit-n','talk-n','ballot-n','election-n','referendum-n','disco-n','congress-n','inquest-n','fair-n','ceremony-n','reunion-n','rally-n','meeting-n','conference-n','seminar-n','parade-n','rehearsal-n','wedding-n','funeral-n','clinic-n','feast-n','celebration-n','session-n','workshop-n','demonstration-n','concert-n','briefing-n','lecture-n','reception-n','banquet-n','luncheon-n','dance-n','gathering-n','event-n','procession-n']

food seed: sandwich-n

event seed: conference-n

Clustering Algorithm-VII

(8) For each of the target's selectors *s* in grammatical relation *R*, compute the following score for each of the chosen cluster *C*:

- $\Sigma_{e \in C} P(s \mid R\ c)$
- this score indicates how likely selector *s* is to pick the sense of the target associated with *C*

Selector Assignment for lunch-n, object_of

hard assignment: $\Sigma_{\text{equiv} \in C}$ P(selector | equiv)

eat-v	FOOD	host-v	EVENT	supply-v	FOOD
cook-v	FOOD	cancel-v	EVENT	make-v	FOOD
serve-v	FOOD	organize-v	EVENT	organize-v	EVENT
skip-v	FOOD	include-v	FOOD	set-v	EVENT
finish-v	FOOD	order-v	FOOD	throw-v	FOOD
enjoy-v	EVENT	grab-v	FOOD	need-v	FOOD
prepare-v	FOOD	give-v	EVENT	blow-v	FOOD
attend-v	EVENT	spoil-v	FOOD	carry-v	FOOD
miss-v	EVENT	share-v	FOOD	estimate-v	FOOD
take-v	FOOD	hold-v	EVENT	follow-v	EVENT
provide-v	EVENT	pack-v	FOOD	lay-v	FOOD
get-v	FOOD	appreciate-v	FOOD	deliver-v	FOOD
bring-v	FOOD	like-v	FOOD	forget-v	FOOD
buy-v	FOOD	offer-v	FOOD	manage-v	EVENT
arrange-v	EVENT	plan-v	EVENT	leave-v	FOOD
want-v	FOOD	interrupt-v	FOOD	pass-v	FOOD

Soft Selector Assignment for lunch-n, object_of

Selector	Selected Type	A-score	Type	A-score	Confidence
eat-v	FOOD	.089	EVENT	.002	.087
cook-v	FOOD	.024	EVENT	.003	.021
serve-v	FOOD	.024	EVENT	.002	.022
skip-v	FOOD	.002	EVENT	.000	.002
finish-v	FOOD	.009	EVENT	.002	.007
enjoy-v	EVENT	.016	FOOD	.006	.010
prepare-v	FOOD	.009	EVENT	.003	.006
attend-v	EVENT	.099	FOOD	.001	.098
miss-v	EVENT	.002	FOOD	.001	.001
host-v	EVENT	.010	FOOD	.000	.010
cancel-v	EVENT	.003	FOOD	.000	.003
organize-v	EVENT	.034	FOOD	.000	.034
include-v	FOOD	.013	EVENT	.011	.002
order-v	FOOD	.008	EVENT	.001	.007
give-v	EVENT	.045	FOOD	.010	.035
share-v	FOOD	.004	EVENT	.002	.002
hold-v	EVENT	.004	FOOD	.157	.153

Lunch-n

- All selectors that occur with lunch-n in the object position
- Sorted on MI(selector, (target, rel))
- Performance much higher than the majority baseline
  - but need a more robust evaluation
- Polysemous selectors linked to the correct sense:

  hold-v lunch-n EVENT

  - all occurrences (10) in the BNC are in the EVENT sense

  leave-v lunch-n FOOD

  - 3 out of 4 occurrences in the BNC are in the FOOD sense

Noun vs. Verb Targets (side remark)

- The same procedure for identifying contextual equivalents may be applied equally to other types of targets besides the dot nominals.
- For example, for verbs!

  Given the target (lunch-n, object_of):

  for sandwich-n, we would need to select verbs such as eat-v, cook-v, make-v, etc.

  for conference-n, we would need to select verbs such as eat-v, cook-v, make-v, etc.

  Given the target (deny-v, object):

  for confirm-v, we would need to select report-n, existence-n, allegation-n,

  for grant-v, we would need to select access-n, right-n, approval-n, permission-n, etc.

- For verbs, selectors are nouns; for nouns, selectors are verbs

Conclusions

- Method for deriving automatically sets of *selectors that select for a specific component type*
- We avoid the common computational pitfalls in distributional similarity-based clustering by computing clusters of *short contextualized vectors*
- Association scores obtained for each selector with respect to the resulting clusters give a *measure of certainty*
- Can be *extended to several relations* by combining obtained selector assignments
- Can do this without the human labor (!)

Thank you!