

# The Phenomenal Corpus: A Resource for Evaluating Grammatical Phenomenon Discovery

Ned Letcher



THE UNIVERSITY OF  
MELBOURNE

# Grammatical Phenomena in Precision Grammars

## Desideratum

Indexing grammars by their contained grammatical phenomena, making grammar components more discoverable.

### Grammar engineering applications

- Enhanced documentation of grammars
- eg can leverage existing implementations for inspiration

### Language documentation applications

- Navigate exemplars in descriptive grammars by phenomena
- Retrieve additional examples from treebanks
  - ▶ Bender et al 2012
  - ▶ Bouma et al 2014

# Grammatical Phenomena in Precision Grammars

## Challenges

- How to define *grammatical phenomenon*?
- Analyses touch multiple types often across different files
- How to classify the “dark matter” of the grammar?
  - ▶ Constraints for removing spurious ambiguity
  - ▶ Constraints for excluding chart edges etc

## Approaches to the problem

1. Phenomenon-centric grammar engineering
  - ▶ eg CLIMB and other metagrammar strategies
2. Documentation tools and infrastructure for existing grammars
  - ▶ eg Lextype DB/Lingtype DB

# An Emerging Paradigm

Developing synergistic pairings of

1. Signature discovery tools
2. Example retrieval tools

Examples

- Gretel + Paqu (Taalportaal)
- Discovery procedure + semantic graph querying (ESD Project)
- Typediff + Fangorn

Flexibility

- Defers many of the challenging decisions to grammar writers
- Can be used off-line for documentation
- Can be used on-line by end users

# Evaluation

Given a phenomenon, how readily can users discover the signatures of phenomena and compose queries to retrieve examples from a treebank?

Properties of required resource

- Corpus data (ie not manually curated)
- Exhaustively annotated
- Token-level annotation
- Framework independent

# The Phenomenal Corpus

Focus: largely syntactically characterised phenomena

Chosen text

- 200 lines of Sherlock Holmes, *The Speckled Band*
- 477 lines of PTB (Section 8)

Chosen Constructions

- Relative clauses
- Imperative clauses
- Passive clauses

# Characterising Phenomena

- Assume we have a prose description of phenomena.
- eg from descriptive grammar

---

Passive clause	A valence-modifying construction, which yields the patient role occupying the subject position and the agent being found in an optional prepositional phrase.
----------------	---

---

Relative clause	A subordinate clause which serves to restrict or elaborate on a nominal referent in the higher clause.
-----------------	--

---

Complement clause	Subordinate clause which functions as the argument to a verbal predicate and itself possesses the constituent structure of a clause.
-------------------	--

---

# Methodology

1. Annotation guidelines prepared
2. Guidelines refined over *The Speckled Band*
3. WSJ test divided into two sections, each doubly-annotated
4. “Gold Standard” produced through disagreement resolution process



# Methodology (brat)

1 complement clause  
He also is a consensus manager, insiders say.

2 Associates say complement clause  
Mr. Hahn picked up that careful approach to management as president of Virginia Polytechnic Institute.  
3 But he also showed a willingness to take a strong stand.

4 In 1970, Mr. Hahn called in state police to arrest student protesters relative clause  
who were occupying a university building.  
5 The son of a physicist, Mr. Hahn skipped first grade because his reading ability was so far above his classmates.

6 Mr. Hahn agrees that he has a complement clause "retentive" memory, but friends say complement clause  
that's an understatement.

7 complement clause  
"The record of companies relative clause  
that have diversified isn't all that impressive," he says.

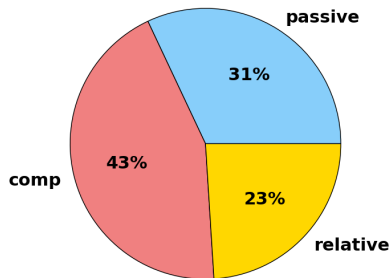
8 complement clause  
It would be a good match, Mr. Hahn and many analysts say,  
complement clause  
of two healthy companies with high-quality assets and strong cash flows.

9 To accommodate the additional cash assistance,  
passive complement clause  
the House Appropriations Committee last week was required to reallocate an estimated \$140 million from the Pentagon.

## Annotations Collected

	<b>SB</b>	<b>WSJ-08</b>
Lines	250	477
Passive	25	102
Relative	62	141
Complement	54	283
Total	162	526

<b>Error type</b>	<b>Total</b>	<b>%</b>
Missing instance	156	56
Spurious instance	26	10
Incorrect label	3	1
Span mismatch	72	26
Conjunction split	3	1
Missing punctuation	18	6
All errors	278	100



# Inter-Annotator Agreement (IAA)

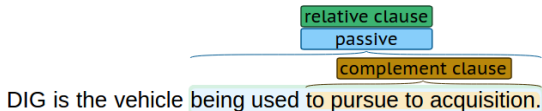
Typical approach for annotation tasks

- Chance-corrected measures: eg Fleiss' kappa

$$\kappa = \frac{\text{observed agreement above chance}}{\text{attainable agreement above chance}}$$

Challenges Presented by this task:

1. Overlapping spans
2. Annotator's delimiting *and* labelling units



# Solving Problem of Overlapping Spans

1. Perform IAA on a per-class basis
2. Use line-copying process for nested phenomena of same class

"We asked police to investigate why they are allowed to distribute the flag in this way.

The diagram illustrates the overlapping spans for the sentence "We asked police to investigate why they are allowed to distribute the flag in this way." The "complement clause" spans from "to investigate" to "in this way." The "passive" spans from "are allowed" to "in this way." The overlapping region is highlighted in yellow.

"We asked police to investigate why they are allowed to distribute the flag in this way.

The diagram illustrates the overlapping spans for the sentence "We asked police to investigate why they are allowed to distribute the flag in this way." The "complement clause" spans from "to investigate" to "in this way." The "passive" spans from "are allowed" to "in this way." The overlapping region is highlighted in yellow.

"We asked police to investigate why they are allowed to distribute the flag in this way.

The diagram illustrates the overlapping spans for the sentence "We asked police to investigate why they are allowed to distribute the flag in this way." The "complement clause" spans from "to investigate" to "in this way." The "passive" spans from "are allowed" to "in this way." The overlapping region is highlighted in yellow.

"We asked police to investigate why they are allowed to distribute the flag in this way.

The diagram illustrates the overlapping spans for the sentence "We asked police to investigate why they are allowed to distribute the flag in this way." The "complement clause" spans from "to investigate" to "in this way." The "passive" spans from "are allowed" to "in this way." The overlapping region is highlighted in yellow.

# Solving Problem of Delimiting & Labelling

## Approach 1: Fleiss' kappa over digitised character tokens

Text: We have the money to buy.  
Tokens: 00000000000000000001111111

### Problem

- Blind to span boundaries
- Measures aggregate agreement (ie span length is irrelevant)

Annotator *A*: 00001111111110000000

Annotator *B*: 00001111111110000100

Annotator *A*: 0000111101110000100

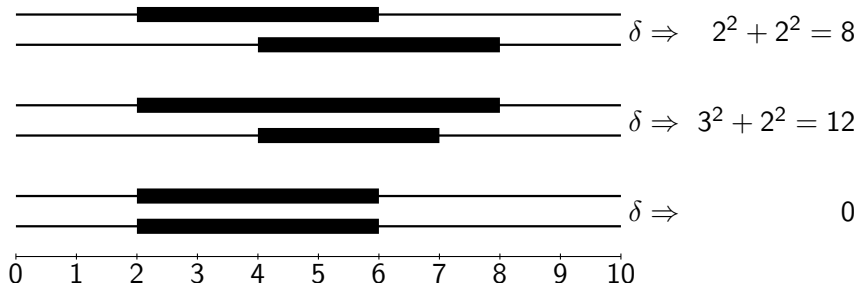
Annotator *B*: 00001111111110000100

## Solving Problem of Delimiting & Labelling

### Approach 2: Krippendorf's $\alpha_U$

- Divides text up into unit/gap sections
- Agreement is determined through pairwise comparison of sections
- Uses a squared difference function for comparison
  - ▶ peripheral disagreement weighted lower than core disagreement

The squared difference function:



# IAA Results

	WSJ-08 1		WSJ-08 2	
Raters	1 & 2		1 & 3	
Coefficient	$\kappa$	$\alpha_U$	$\kappa$	$\alpha_U$
Passive	0.828	0.675	0.676	0.268
Relative	0.783	0.568	0.766	0.410
Complement	0.806	0.742	0.802	0.763

Interpreting kappa-like scores:

$\kappa$  < 0 worse than chance  
 $\kappa$  = 0 as good as chance  
 $\kappa$  > 0 better than chance

$\kappa > 0.8$  "good reliability"  
 $0.67 < \kappa < 0.8$  "some reliability"

(unclear if this applies to  $\alpha_U$ )

## Problems with $\alpha_U$

1. Unclear impact difference function has on interpretation
2. Missing annotations are penalised by the square of their length
3. Less frequent categories penalised more heavily
  - ▶ Property shared by all chance-corrected scores
  - ▶ However possibly amplified by 2)
4. Situations where removal of overlapping spans *increased* agreement:

### Annotator 1:

This was in addition to a more parochial \$4.5 million authorization for a health center in South Carolina upheld by a 273-121 vote in the House last night.

passive

### Annotator 3:

This was in addition to a more parochial \$4.5 million authorization for a health center in South Carolina upheld by a 273-121 vote in the House last night.

passive





# IOBE Results

	Passive			Relative			Complement		
	P	R	F1	P	R	F1	P	R	F1
I	0.999	0.729	0.843	0.999	0.918	0.957	0.954	0.922	0.938
B	1.000	0.707	0.829	0.983	0.894	0.937	0.906	0.906	0.906
E	1.000	0.744	0.853	0.947	0.885	0.915	0.887	0.887	0.887
O	0.977	1.000	0.988	0.990	1.000	0.995	0.976	0.986	0.981

Table: Section 2, Annotator 1

	Passive			Relative			Complement		
	P	R	F1	P	R	F1	P	R	F1
I	0.820	0.826	0.823	0.866	0.828	0.847	0.989	0.874	0.928
B	0.444	0.390	0.416	0.902	0.833	0.866	0.940	0.745	0.832
E	0.829	0.744	0.784	0.864	0.836	0.850	0.940	0.804	0.867
O	0.985	0.984	0.985	0.978	0.984	0.981	0.962	0.998	0.980

Table: Section 2, Annotator 3

# Conclusion

## Observations

- Low recall annotation task; may need double annotation
- Krippendorff's  $\alpha_U$  not so helpful in practice
- Chasing a single measure of agreement of questionable value
- IOBE tags useful for fine-grained disagreement/error analysis
  - ▶ Using precision recall for each annotator if gold standard
  - ▶ Using kappa between annotators if not

## Next steps for the Phenomenal Corpus

- Gauge efficacy of user-created queries using Typediff