# Nominal complementiser clauses, polysemy & lexical semantics

Peter R. Sutton
Universität Potsdam


peter-sutton.github.io/talks

# Informational Nouns

## Informational Nouns

- have at least one sense that denotes a piece, pieces, a body or bodies of information
  - which could be modelled as e.g., a proposition, some collection of propositions, ...

Examples of informational nouns in English are given in (1)

(1) *allegation*, *belief*, *book*, *fact*, *information*, *knowledge*, *newspaper*, *report*, *statement*

# Noun-related propositional complementiser clauses (NCCs)

*noun-related complementiser clause* NCC adopted from Müller 2023

Not all nouns can be used with NNCs:

(2) a. the allegation/belief/evidence/fact/information/report/statement that Bilbo found the ring
b. #the house that Bilbo met Gandalf
c. #the event/celebration that Gandalf set off fireworks

Rough consensus about noun-NCC constructions:

- relate the relevant proposition to what is denoted by the noun

But a naive hypothesis that NCCs are licensed by informational nouns is not supported (see data in Elliott 2020)

(3) #the book/brochure/document/newspaper that Bilbo found the ring

# Previous analyses: The semantics of NCCs

Kratzer 2006; Moulton 2015; Elliott 2020

Informational nouns express vanilla properties of individuals

(4) a. $[\![\text{claim}]\!] = \lambda w.\lambda x.CLAIM(w)(x)$
b. $[\![\text{fact}]\!] = \lambda w.\lambda x.FACT(w)(x)$

NCCs also express a property of individuals (5-b):

(5) a. $[\![\text{Alex lied}]\!] = \lambda w.LIE(w)(a)$
b. $[\![\text{that Alex lied}]\!] = \lambda w.\lambda x.[CONT(w)(x) = \lambda w.LIE(w)(a)]$

And via Predicate Modification (Heim and Kratzer, 1998):

(6) a. $[\![\text{claim that Alex lied}]\!]$
$= \lambda w.\lambda x.[CLAIM(w)(x) \wedge (CONT(w)(x) = \lambda w.LIE(w)(a))]$
b. $[\![\text{fact that Alex lied}]\!]$
$= \lambda w.\lambda x.[FACT(w)(x) \wedge (CONT(w)(x) = \lambda w.LIE(w)(a))]$

# *Contentful entities* and the puzzle

Kratzer 2006; Moulton 2015; Elliott 2020

(3) #the book/brochure/document/newspaper that Bilbo found the ring

NCCs are only defined for "contentful entities"

> *"The natural conclusion to draw from such contrasts is that the linguistic notion of a contentful entity is distinct from the intuitive notion." (Elliott 2020, p.63)*

> *"Rawlins assumes that media artifact nouns do in fact denote predicates over contentful entities, and it is not immediately obvious how to reconcile this claim with the distribution of that-clauses." (Elliott 2020, p.63, n.9)*

# Challenges for the contentful entities view

## Stipulative

- What are/are not *contentful entities* is stipulated

## Ungrounded

- The existence of contentful entities is based upon the very data they are supposed to explain

## Facts are not entities that have content

- NCCs are assumed to encode a *CONT* relation between contents and things that have contents
- But what entities are denoted by *fact* such that they HAVE content? From the philosophy literature:
  - facts are (true) propositions (Strawson, 1949) or
  - facts are situations that make propositions true (Austin, 1949) (but do not HAVE contents)

# Polysemy and Informational Nouns: Examples

## Examples: Nouns in different POLYSEMY CLASSES

- *book* is (at least) 2-ways polysemous
  - the physical book made of paper, and an informational sense, the body of information that is the contents of the book (see e.g., Pustejovsky 1995; Asher 2011)
- *statement* is (at least) 3-ways polysemous
  - a stating event or a physical artefact (e.g., a written document) and the informational contents of each of these (Sutton, 2022)
- *fact* is not polysemous in this way
  - has an informational sense
  - more on other senses of *fact* later

# Motivating an alternative hypothesis: polysemy

(7) a. the allegation/belief/evidence/fact/information/report/statement that Bilbo found the ring

b. #the book/brochure/document/newspaper that Bilbo found the ring

**The informational nouns in (7-a):**

| | |
|---|---|
| allegation, belief, evidence, report, statement | with an eventuality-denoting sense |
| fact, information | without a physical- or eventuality-denoting sense |

**The informational nouns in (7-b):**

| | |
|---|---|
| book, brochure, document, newspaper | with a physical entity-denoting sense, no eventuality-denoting sense |

# Identifying *book*-like nouns

(8) a. the allegation/belief/evidence/fact/information/report/ statement that Bilbo found the ring
b. #the book/brochure/document/newspaper that Bilbo found the ring
c. #the house that Bilbo met Gandalf
d. #the event/celebration that Gandalf set off fireworks

In order to distinguish nouns in (8-b) from those in (8-c,d) it will be helpful to find a construction that isolates nouns such as those in (8-b).

# The locative strategy for referring to informational contents

Some languages use locative PPs to target the informational contents of nouns like *book*:

(9) Context: Andrew is a character in the book:

a. #the book that Andrew lied

b. the book in which Andrew lied

(10) el libro en el que Andrés mintió
the.M book in REL.M that Andrés lied
"the book in which Andrés lied"

(11) das Buch in dem Andreas gelogen hat
the.N book in REL.N.DAT Andreas lied has
"the book in which Andreas lied"

# The locative strategy as an alternative to NCCs

## The information locative PP construction has a near complementary distribution with the NCC construction

(12) The information locative PP construction (I-LOC). Examples:

a. the information in this book/brochure/document/newspaper/report/statement

b.??the information in this allegation/belief/evidence/fact/information

### Only NCC nouns

- allegation, belief, evidence, fact, information

### Only I-LOC nouns

- book, brochure, document, newspaper

### Both NCC and I-LOC nouns

- report, statement

# Hypothesis

(H) The polysemy class of a noun determines whether that noun is felicitous in NCC, I-LOC, or both NCC and I-LOC constructions.

a. If a noun is, minimally INF-PHYS polysemous it can be used in I-LOC constructions. This includes:
   - *book*, *newspaper*, which are INF-PHYS polysemous;
   - *statement*, *report*, which are INF-PHYS-EV polysemous

b. If a noun has an INF sense (but not a PHYS sense), or is INF-EV polysemous, it can be used in NCC constructions. This includes:
   - informational nouns such as *fact*
   - *allegation*, *belief*, *statement* which are INF-EV polysemous

# Main questions

(Q1) To what extent do different types of nominal polysemy predict the licensing conditions for the NCC and I-LOC environments?

(Q2) What implications does the answer to Q1 have on the semantics of NCCs, and on how different informational nouns should be semantically classified?

# Still to come

- Corpus study
  1. Define a class of nouns that are felicitous in NCC and/or I-LOC
  2. Annotate these nouns for senses (e.g., physical entity, eventuality)
  3. Assess whether 2 predicts 1 (Conditional Inference Tree (CIT) analysis)
- Refine the CIT model with additional parameters
- Discussion of the implications for formal semantic analyses

# Corpus information

The UK Web Annotated Corpus (ukWaC)
(Ferraresi et al., 2008; Baroni et al., 2009)

- 1,547,594,305 tokens
- POS tagged, but not dependency parsed
- Accessed via SketchEngine (`https://www.sketchengine.eu`)
  - The UkWaC corpus is also available on request `https://wacky.sslmit.unibo.it`

# Methodology Overview

1. API requests[1] to pull (noisy) corpus data for NCC and I-LOC environments.
2. Dependency parsing using the *spaCy* Python package (Honnibal and Montani, 2017). Use a Python script to identify the nouns in these constructions. Manual checking and cleaning
3. Sense annotation of nouns using listed senses scraped from `www.wiktionary.org`
   - All nouns assumed to have an informational sense
   - Annotation only for ±P(hysical entity) and ±E(ventuality)
4. Conditional Inference Tree modelling
   - Probabilistic model that can be used to identify which nouns are correctly/incorrectly predicted to be felicitous in the NCC and I-LOC environments by the hypothesis

[1] Application Programming Interface request, a means of pulling large amounts of data automatically from a website or database.

# API requests & Dependency parsing

NCC

- Any sentence containing *N that DP*
  - Nb. Rel clauses filtered later after dependency parsing
- 1,078,043 corpus hits
- a random sample of 30,000 of these sentences was dependency parsed

I-LOC

- Any sentence containing *information in this/that* followed by an NP, with optional adjective and that could be a compound
- 1653 corpus hits (after duplicates removed)
- All sentences dependency parsed

# Cleaning: NCC

Not possible to differentiate the following structures from the dependency parse information

(13) a. Comment on the fact that they quit. (good case)
b. Stop in the event that they quit. (bad case)

- the script was adjusted to exclude all cases where the target noun is the object of a preposition

Other frequent false positives removed:

(14) a. so much/little N that ...
b. such an A N that ...

After duplicates were removed and the algorithm applied:

- 6935 sentences, 607 unique noun lemmas
- 75 most frequent carried forward for further analysis

# Demo

# Cleaning: I-LOC

## Output of python script

- 204 unique noun lemmas, the most frequent 75 were further analysed

## Some excluded lemmas:

- *section*, *chapter*, *column*, *paragragh*, *page*, *part*
  - Labour-saving measure: Not bad, per se, but all parts of larger informational entities (e.g., books)
- *area*, *category*, *class*, *connection*, *context*, *format*, *manner*, *matter*, *regard*, *respect*, *way*.
  - Classified types of information or relations between information (e.g. *the information in this connection was useful*

## The cleaning process left 50 unique noun lemmas and 732 sentences

# Re-running the corpus search

## Lemmas from each environment searched for in the other

- Since I only took a sample of most frequent lemmas, nouns that appear in both NCC and I-LOC could be missed
- Re-running the process on all identified lemmas ensured finding those that are felicitous in both environments
- Total of 119 unique lemmas

## An issue with *issue*

- *issue* was one of the nouns found in I-LOC and NCC constructions
- But clearly a case of lexical ambiguity
  - *issue*: problem/topic (cf. *Sachverhalt*, *Problem*, *Thema*)
  - *issue*: volume/tome (cf. *Ausgabe*, *Band*)
- *issue-1* (felicitous in NCC);
  *issue-2* (felicitous in I-LOC)

# Lemma lists

(15) *Only*-NCC nouns:
admission, advantage, agreement, allegation, argument, assertion, assumption, assurance, belief, chance, claim, complaint, concern, conclusion, confidence, confirmation, consensus, contention, conviction, danger, decision, declaration, doubt, evidence, expectation, fact, fear, feeling, finding, guarantee, hint, hope, hypothesis, idea, impression, indication, issue-1, knowledge, likelihood, news, notice, notion, opinion, order, perception, point, possibility, principle, probability, proof, proposal, question, realisation, reassurance, recognition, recommendation, reminder, requirement, risk, rumour, sense, sign, signal, speculation, suggestion, suspicion, theory, thought, view, warning

(16) *Only*-I-LOC nouns:
appendix, article, book, booklet, box, briefing, brochure, catalogue, database, directory, document, edition, factsheet, field, file, form, guide, handbook, issue-2, leaflet, list, manual, newsletter, pack, paper, prospectus, publication, series, site, table, text, volume, webpage, website

(17) *Both*-NCC&I-LOC nouns:
announcement, application, bulletin, case, email, (diary/log) entry, guidance, letter, message, note, (press) release, report, review, statement, story

# Sense Annotation

An independent means of annotating each noun for having either an eventuality-denoting sense, a physical entity-denoting sense, or both was needed

- I used a free version of the Lingua Robot API which provides access to JSON-formatted English wiktionary.org entities (`https://www.linguarobot.io`)
- Where the Wiktionary entry was not clear, or where there seemed to be at least one sense missing, the relevant entry in the Oxford English Dictionary (`https://www.oed.com`) was checked

# Sense Annotation: criteria

## Physical entity denoting senses

- referring to objects or animate individuals
- expressions such as *written* or *in writing*
- digital entities (*email*, *website*): treated as physical entities.

## Eventuality denoting senses

- reference to acts, actions, events, situations, circumstances, processes or (mental) states
- the use of relevant psych nouns and verbs were taken to indicate an eventuality qua mental state

# Results of sense annotation

## Assuming that all of the nouns have an informational sense:

(18) [+I, -E, -P] Inf nouns without eventuality or physical entity senses:
chance, danger, fact, hint, hypothesis, issue-1, likelihood, news, point, principle, probability, requirement, rumour, theory

(19) [+I, +E, -P] Inf nouns with an eventuality sense, but no physical entity sense:
admission, advantage, allegation, argument, assertion, assumption, assurance, belief, briefing, claim, complaint, concern, conclusion, confidence, consensus, contention, conviction, decision, doubt, expectation, fear, feeling, finding, hope, idea, impression, knowledge, notion, opinion, perception, possibility, question, realisation, reassurance, recognition, recommendation, risk, sense, speculation, suggestion, suspicion, thought, view

(20) [+I, -E, +P] Inf nouns with a physical entity sense, but no eventuality sense:
appendix, article, book, booklet, box, brochure, catalogue, database, directory, document, edition, email, entry, factsheet, field, file, form, guide, handbook, issue-2, leaflet, letter, list, manual, newsletter, note, pack, paper, press release, prospectus, publication, release, series, site, table, text, volume, webpage, website

(21) [+I, +E, +P] Inf nouns with both a physical entity sense and an eventuality sense:
agreement, announcement, application, bulletin, confirmation, declaration, evidence, guarantee, indication, message, notice, order, proof, proposal, reminder, report, review, sign, signal, statement, warning

# Stats package information

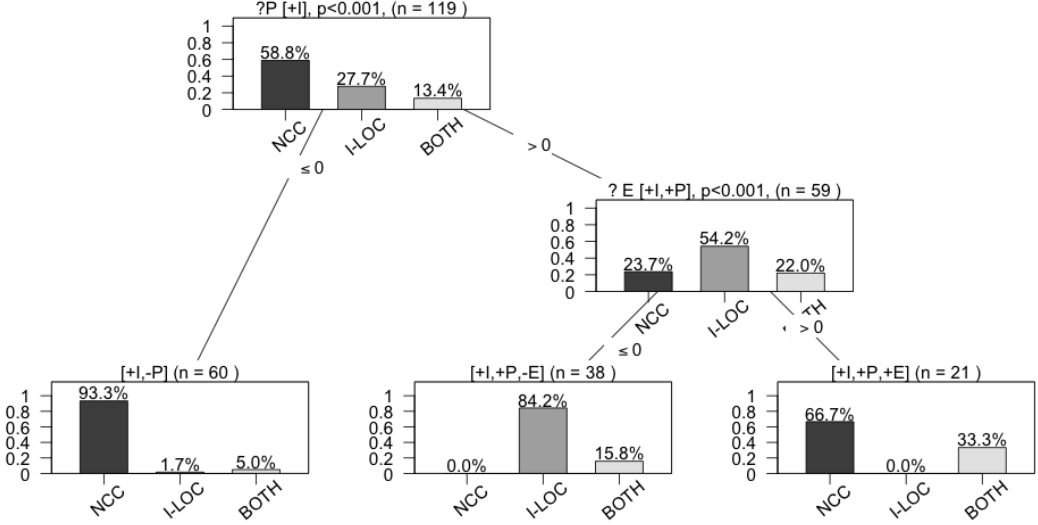
- Tree (CIT) analysis was applied using the PARTYKIT package in R (Hothorn and Zeileis, 2015)
- The model was run with the following inputs (all other parameters taken as defaults):

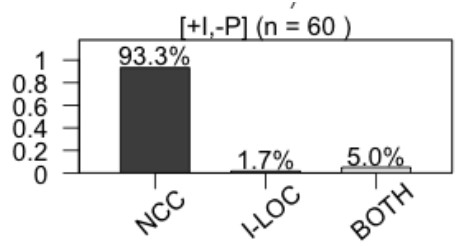
(22) `data.cit <- ctree(diagnostic ~ ev + phys, data = data)`

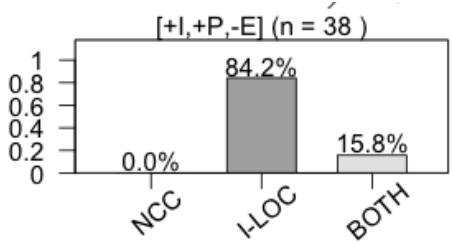
# Conditional Inference Tree Analysis

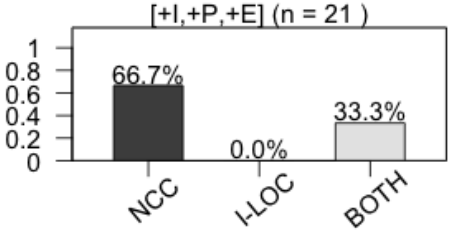
## CIT tree analyses are a form of recursive partitioning method

- The algorithm searches for ways of splitting the data (in this case the set of nouns), given the available predictors, in this case $\pm E$ and $\pm$P
- Splits are only chosen if they exceed a default significance value of $p = 0.05$ for rejecting the global null hypothesis (which is derived via randomly permuting the values of the response variables).
- Predictions are based on terminal node frequencies:
  - distributions over only-NCC, only-I-LOC or both-NCC&I-LOC
  - Predictions for the model are derived by taking the category with the highest frequency (first past the post).


?P [+I], p<0.001, (n = 119 )
58.8%
27.7%
13.4%
NCC
I-LOC
BOTH
≤ 0
> 0
? E [+I,+P], p<0.001, (n = 59 )
23.7%
54.2%
22.0%
NCC
I-LOC
≤ 0
> 0
[+I,-P] (n = 60 )
93.3%
1.7%
5.0%
[+I,+P,-E] (n = 38 )
0.0%
84.2%
15.8%
[+I,+P,+E] (n = 21 )
66.7%
0.0%
33.3%


[+I,-P] (n = 60 )
93.3%
1.7%
5.0%
1
0.8
0.6
0.4
0.2
0
NCC
I-LOC
BOTH


[+I,+P,-E] (n = 38 )
0.0%
84.2%
15.8%
1
0.8
0.6
0.4
0.2
0
NCC
I-LOC
BOTH


[+I,+P,+E] (n = 21 )
66.7%
0.0%
33.3%
1
0.8
0.6
0.4
0.2
0
NCC
I-LOC
BOTH

# Identifying bad predictions

Table: Predictions of the CIT model regarding use in the NCC and I-LOC environments, given their sense annotation. False predictions are highlighted in orange. This equates to a reliability of $1 - (17/119) \approx 0.86$.

| | | Actual category | | | |
|---|---|---|---|---|---|
| | | NCC | I-LOC | BOTH | Errors |
| Predicted category | NCC | **70** | 1 | 10 | 11 |
| | I-LOC | 0 | **32** | 6 | 7 |
| | BOTH | 0 | 0 | **0** | 0 |
| | Errors | 0 | 1 | 17 | 18 |

## Summary

- Mostly a result of over-extending the *only*-NCC prediction (10/18)
- Suggests that [+I, +P, +E] is not a sufficiently good predictor of this

# Next steps

- Deeper dive into where the model is going wrong
- Can we identify any other predictors?

# Where did the model go wrong?

(23) Nouns which are [+I,+P, +E] and felicitous in both the NCC- and I-LOC-environments, but falsely predicted to be only NCC.
*announcement*, *application*, *bulletin*, *message*, *report*, *review*, *statement*.

(24) Nouns which are [+I,+P, -E] and felicitous in both the NCC- and I-LOC-environments, but falsely predicted to be only I-LOC.
*email*, *entry*, *letter*, *note*, (*press*) *release*, *text*

## Nouns in (24):

- These nouns denote artefacts that are used as messages/missives
- That these artefacts are used to inform somebody of something (e.g., *message*) could be eventuality-related
- Some overlap with some of the nouns in (23): *announcement*, *bulletin*, *message*

# New predictor: Missive Nouns

Missive Nouns:

- (An) N we was sent out/written in order to inform $x$ of $y$.

(25) Missive nouns (nouns that are [+M]):
*announcement*, *bulletin*, *declaration*, *email*, (diary/log) *entry*, *letter*, *message*, *note*, *notice*, *order*, (*press*) *release*, *proposal*, *reminder*, *report*, *review*, *statement*, *text*, *warning*

# Another source of error

Original motivation for the hypothesis:

- Nouns such as *message* and *statement* can be informational, eventuality-denoting, or denote physical objects
- These nouns are felicitious in both NCC and I-LOC
- The following data was unexpected:

(26) [+I, +E, +P] nouns that are felicitous in the NCC environment, but are not felicitous in the I-LOC environment:
*agreement*, *confirmation*, *declaration*, *evidence*, *guarantee*, *indication*, *notice*, *order*, *proof*, *proposal*, *reminder*, *sign*, *signal*, *warning*

Possible pattern: Some of these nouns

- *confirmation*, *evidence*, *guarantee*, *indication*, *proof*, *reminder*, *sign*, *signal*, *warning*

Seem to be of a different nature that nouns such as *message* and *statement*

# Examples: *proof, sign*

- An object/eventuality can be proof/a sign that supports some proposition.
  - e.g., a fingerprint or lying can be proof/a sign that someone is guilty
- So-called *subject nominals* such as *proof*, *sign* have been identified as outliers in the NCC literature (see e.g., Safir 1985; Moulton 2009; Müller 2023)

(27) The best proof that John was not lying is that he was here last night. (Higgins 1973)

> *"It is clearly not plausible that both that-clauses refer to the same object or describe the content of the same object [...] They refer to two different things: one thing that is considered to be the proof, that is the thing by which something is proven, and another thing that is proven by that" (Müller 2023 p.93)*

Nouns such as *proof* and *sign* arguably relate to Saussurian signs which have a *signifier* and a *signified*

# Signifiers

To paraphrase Müller, that by which something is proven, and that which that thing proves

- This contrast can be seen in (28) and (29):

(28) a. The fingerprint is proof/a sign/an indication/a confirmation/a reminder of Alex's innocence.
b. [The] Oklahoma [bombing] was a warning that cancer was spreading in the body politic. [ukWaC, 939890]

(29) a. ?The handshake was an agreement between them.
b. ?Alex speaking in court was a statement that he is innocent.

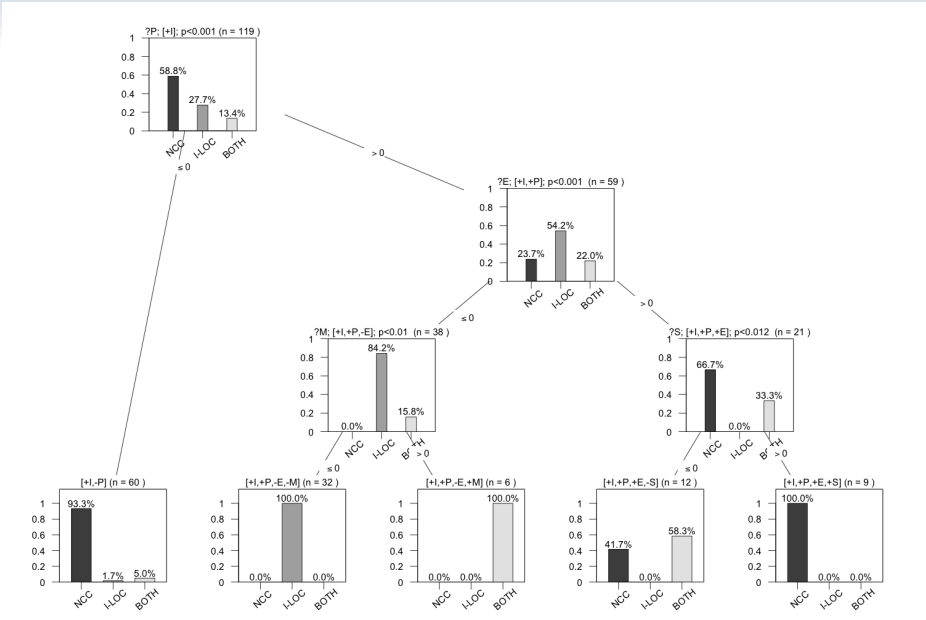
Annotating nouns as signifiers:

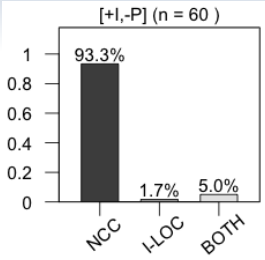
(30) Nouns that denote *signifiers* (nouns that are [+S]):
*confirmation*, *danger*, *evidence*, *guarantee*, *hint*, *indication*, *proof*, *reminder*, *sign*, *signal*, *warning*

# Re-running the CIT analysis with extra predictors

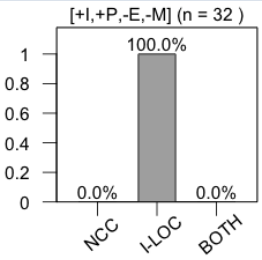
The CIT model was re-run taking ±S(ignifier) and ±M(issive) as extra predictors:

(31)
```
data.cit <- ctree(diagnostic ~ ev + phys + signifier
                          + missive, data = data)
```

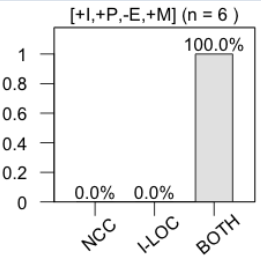

?P; [+I]; p<0.001 (n = 119 )
58.8%
27.7%
13.4%
NCC
I-LOC
BOTH
≤ 0
> 0
?E; [+I,+P]; p<0.001 (n = 59 )
23.7%
54.2%
22.0%
NCC
I-LOC
BOTH
≤ 0
> 0
?M; [+I,+P,-E]; p<0.01 (n = 38 )
0.0%
84.2%
15.8%
NCC
I-LOC
BOTH
≤ 0
> 0
?S; [+I,+P,+E]; p<0.012 (n = 21 )
66.7%
0.0%
33.3%
NCC
I-LOC
BOTH
≤ 0
> 0
[+I,-P] (n = 60 )
93.3%
1.7%
5.0%
NCC
I-LOC
BOTH
[+I,+P,-E,-M] (n = 32 )
100.0%
0.0%
0.0%
NCC
I-LOC
BOTH
[+I,+P,-E,+M] (n = 6 )
100.0%
0.0%
0.0%
NCC
I-LOC
BOTH
[+I,+P,+E,-S] (n = 12 )
41.7%
0.0%
58.3%
NCC
I-LOC
BOTH
[+I,+P,+E,+S] (n = 9 )
100.0%
0.0%
0.0%
NCC
I-LOC
BOTH



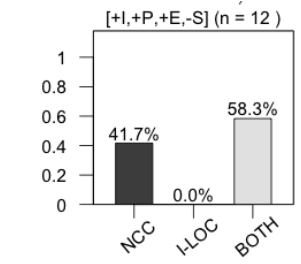

e.g., *fact*



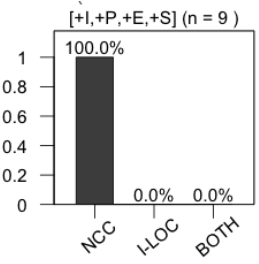

e.g., *book*




e.g., *letter*




e.g., *agreement, statement*




e.g., *sign*

# Identifying bad predictions

Table: First Model: Actual category vs. predicted category in the model

| | | Actual category | | | |
|---|---|---|---|---|---|
| | | NCC | I-LOC | BOTH | Errors |
| Predicted category | NCC | **70** | 1 | 10 | 11 |
| | I-LOC | 0 | **32** | 6 | 7 |
| | BOTH | 0 | 0 | **0** | 0 |
| | Errors | 0 | 1 | 17 | 18 |

Table: Second Model: Actual category vs. predicted category in the model

| | Actual category | | | |
|---|---|---|---|---|
| | NCC | I-LOC | BOTH | Errors |
| NCC | **65** | 1 | 3 | 4 |
| I-LOC | 0 | **32** | 0 | 0 |
| BOTH | 5 | 0 | **13** | 5 |
| Errors | 5 | 1 | 3 | 9 |

## Summary: New model

- $111/119 = 92.4\%$ accuracy
- The CIT model with extra predictors outperforms the first run of the model: 92.4% vs. 84.8%
- Most under-prediction of *both* NCC&I-LOC resolved
- Now some under-prediction of *only*-NCC

# Discussion

Table: Details of remaining errors in the second model

| noun | diagnostic | ev | phys | signifier | missive | cit prediction |
|---|---|---|---|---|---|---|
| briefing | loc | 1 | 0 | 0 | 0 | comp |
| guidance | both | 1 | 0 | 0 | 0 | comp |
| story | both | 1 | 0 | 0 | 0 | comp |
| case | both | 0 | 0 | 0 | 0 | comp |
| agreement | comp | 1 | 1 | 0 | 0 | both |
| declaration | comp | 1 | 1 | 0 | 1 | both |
| notice | comp | 1 | 1 | 0 | 1 | both |
| order | comp | 1 | 1 | 0 | 1 | both |
| proposal | comp | 1 | 1 | 0 | 1 | both |

## Annotation issues or data sparsity?

- Most nouns in the mid-section of the Table may in fact be used with a physical entity denoting sense, even if dictionaries do not list this
  - E.g., *story* can be used an an abbreviation of *story book*
- Nouns in the lower section of the Table may in fact be found in I-LOC constructions in a larger corpus

# Summary: Empirical Landscape

(Q1) To what extent do different types of nominal polysemy predict the licensing conditions for the NCC and I-LOC environments?

- Found evidence that polysemy class covers a large amount of the data ($\approx$ 85%)

With extra parameters, increased to $\approx$ 92%

- Missive nouns (*letter*, *message*)
- Signifier nouns (*evidence*, *signal*)

Next step: What does this mean for the semantics of informational nouns and NCCs?

# Implications for formal semantic analyses of NCCs

(Q2) What implications do [the corpus study findings] have on the semantics of NCCS, and on how different informational nouns should be semantically classified?

(Q2.1) Do NCCs hardcode the *contents* relation? $(\lambda p.\lambda x.[CONT_w(x) = p])$

- signifier nouns
- *fact*-like nouns

(Q2.2) What conditions are placed on the lexical semantics of nouns that are felicitous with NCCs in light of the corpus study findings?

- signifier nouns
- *fact*-like nouns
- missive nouns

## *signifier* nouns (*proof*, *evidence*)

(32) The boot print / That the DNA matches / Alex fleeing the scene } is evidence { that Alex is guilty. / of Alex's guilt.

The data in (32) indicate:

- *Alex is guilty*/*Alex's guilt* can be related to entities, propositions, eventualities
- This is a *support*, not a *contents* relation
  - ??the content of (that) the DNA matches is that Alex is guilty?

Implications for questions (Q2.1) and (Q2.2):

1. NCCs do not (always) encode a contents relation.
2. The similar meaning between the CP and PP constructions suggests that the *support* relation is encoded by the noun, not the NCC

# *fact*-like nouns

## Problem raised earlier

- facts either are (true) propositions or are parts of the world/situations that make such propositions true

## Sutton & Cooper, ms: *fact* is polysemous between these two readings

(33) a. The terrible condition of the bike lanes is a fact. (situation reading)
b. The fact that the bike lanes are terrible cannot be denied. (informational reading)
c. The terrible condition of the bike lanes is a fact which cannot be denied. (copredication)
d. ??The terrible condition of the bike lanes is a/the fact that they contain many potholes.

## (Q2.1) Do NCCs hardcode the contents relation?

- No: *fact* encodes a relation between proposition and witness/truth-maker, NOT between bearer and contents

# *fact*-like nouns – revising the hypothesis?

Are eventualities special cases of situations (Sutton & Cooper, ms)?

- Situations such as *the terrible condition of the bike lanes*
- Situations-qua-eventialities (including belief states and speech acts)

(Q2.2) What conditions are placed on the lexical semantics of nouns that are felicitous with NCCs in light of the corpus study findings?

- Plausibly, NCCs select for informational nouns which also have a situation-denoting sense
  - includes: *fact, belief, statement*
  - excludes: *book, newspaper*

# *missive* nouns (*letter, note*)

The problem for the original hypothesis:

- Missive nouns are $+$*physical-entity*, $-$*eventuality*
- Predicted to be infelicitous in NCC constructions
- But *I got a letter that I should pay a fine* is okay

Observation: variation in felicity depending on the content of the NCC:

(34) a. ?The broadcaster received a letter that a viewer complained about the programme.
b. The broadcaster received a letter that the presenter's comment was inappropriate.

# Routinised coercion

Reminiscent of cases of routinised coercion based on TELIC qualia (Pustejovsky 1995)

- E.g., eventuality-selecting verbs such as *start* can be felicitously combined with DPs such as *the book*
  - *began the book* in (35) is coerced to *began* READING *the book*

(35) Mary began the book. (Pustejovsky 1995)

Access to this eventuality is degraded in some constructions:

(36) ?War and Peace is a six-month book. (Chatzikyriakidis et al. 2025)

# Missive Ns in NCCs result from coercion

The TELIC-*quale* of a missive nouns (*letter*):

- a *communication eventuality e, such that e communicates that p.*
- *a letter* COMMUNICATING *that the presenter's comment was inappropriate*

For describing general topics, coercion is not straightforwardly available

- ?*The broadcaster received a letter* COMMUNICATING *that the/a viewer complained about the programme.*

And is completely out for *book*

- ??*This is the book* READING *that Bilbo found the ring.*

(Q2.2) What conditions are placed on the lexical semantics of nouns that are felicitous with NCCs in light of the corpus study findings?

- Depends on whether coercion phenomena are evidence of richer lexical structure vs. e.g. free pragmatic enrichment

# Summary: Implications for formal semantics

- *fact*-like and signifier nouns suggest that NCCs do not (always) encode a *contents* relation.
- Moreover the relevant relation is plausibly encoded by the polysemous noun.
- The ungainly original hypothesis can plausibly be made more streamlined if eventualities are analysed as special cases of situations.
- At the very least, the conceptual (even if not lexical) structure of missive nouns (*letter*) seems to differ from nouns such as *book*

# Conclusions

## Talking about informational content

- Complementiser strategy vs. locative strategy
  - Does this underlie a metaphysical/ontological distinction?
  - Or more a reflection of:
    - modifying verbal + CP complement strategies to fit the nominal domain
    - extending a container+contents metaphor to other nouns

## Future work (currently underway, with Jesse Humaljoki):

- Replication corpus study in Finnish
- Also looking towards Turkish (and maybe modern Greek)

# Thanks and Acknowledgements

Thanks for your attention!

**Acknowledgements:**

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number 552248395.

Thank you to Carla Bombi, Frank Grüneisen, and Alex Hogrebe for Spanish and German judgements.

# References I

Asher, N. (2011). *Lexical Meaning in Context: A Web of Words*. Cambridge University Press.

Austin, J. L. (1949). Truth. *Proceedings of the Aristotelian Society 24*, 111–128.

Baroni, M., S. Bernardini, A. Ferraresi, and E. Zanchetta (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation 43*(3), 209–226.

Benz, J. (2025). *Structure and interpretation across categories*. Ph. D. thesis, University of Pennsylvania.

Chatzikyriakidis, S., R. Cooper, E. Gregoromichelaki, and P. R. Sutton (2025). *Theories of Types and the Structure of Meaning*. Cambridge University Press. Part of the Cambridge Elements in Semantics series.

Elliott, P. D. (2020). *Elements of Clausal Embedding*. Ph. D. thesis, University College London.

Ferraresi, A., E. Zanchetta, M. Baroni, and S. Bernardini (2008). Introducing, evaluating ukWaC, a very large web-derived corpus of English. In S. Evert, A. Kilgarriff, and S. Sharoff (Eds.), *Proceedings of the 4th Web as Corpus Workshop*, Marrakech, Morocco, pp. 47–54. European Language Resources Association.

Grimshaw, J. (2015, 10). The light verbs say and say. In I. Toivonen, P. Csúri, and E. Van Der Zee (Eds.), *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. The MIT Press.

# References II

Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar*. Malden, MA: Blackwell.

Higgins, F. R. (1973). *The pseudocleft construction in English*. Ph. D. thesis, MIT.

Honnibal, M. and I. Montani (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Hothorn, T. and A. Zeileis (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research 16*, 3905–3909.

Kratzer, A. (2006). Decomposing attitude verbs. `https://semanticsarchive.net/Archive/DcwY2JkM/attitude-verbs2006.pdf`.

Major, T. (2021). *On the nature of "say" complementation*. Ph. D. thesis, University of California, Los Angenles.

Moulton, K. (2009). *Natural Selection and the Syntax of Clausal Complementation*. Ph. D. thesis, University of Massachusetts at Amherst.

Moulton, K. (2015). Cps: Copies and compositionality. *Linguistic Inquiry 46*(2), 305–342.

Müller, K. (2023). On noun-related complementizer clauses. In K. Axel-Tober, L. Gunkel, J. Hartmann, and A. Holler (Eds.), *On the nominal nature of propositional arguments*, pp. 93–120. Buske.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.

# References III

Safir, K. (1985). *Syntactic Chains*. Cambridge University Press.

Strawson, P. F. (1949). Truth. *Proceedings of the Aristotelian Society 24*, 129–156.

Sutton, P. R. (2022). Restrictions on copredication: a situation theoretic approach. *Semantics and Linguistic Theory (SALT) 32*, 335–355.

# CQL request: the NCC diagnostic

(37) `year|week|day|today|yesterday|tomorrow|century|time|thing|something|everything|anything|everyone|everything|anyone|someone|nothing|friend`

(38) `<s/> containing [tag="NN|NNS"&lemma!="(37)"] [word="that"][tag="PP.*|NP.*|DT"]`[1]

Any sentence containing *N that DP* (excluding the above mentioned cases).

# CQL request: The I-LOC diagnostic

(39) `<s/> containing`
`[lemma="information"][tag="IN"&lemma="in"]`
`[tag="DT"&lemma="this|that"][tag="JJ"]*`
`[tag="NN"]*[tag="NN|NNS"]`
Any sentence containing *information in this/that* followed by an NP, with optional adjective and that could be a compound.

# Comparison with Benz 2025

Benz (2025) defends an *allosemy* account:

- Different readings are associated with different structures. Unclear how this extends to coprediction constructions.
- Does not discuss negative data wrt informational nouns that are not felicitous in NCC constructions

# Signifier nouns in the Syntax literature

Recall: these 'subject nominals' can be used in double NCC constructions

(27) The best proof that John was not lying is that he was here last night. (Higgins, 1973)

Safir (1985), Moulton (2009):

- "subject nominals" such as *proof* and *evidence* are relational ($\langle st, \langle e, t \rangle \rangle$)
- *The proof (that Bob was guilty)* the type *st* argument is filled by the CP (or contextually), and the type *e* argument can be a situation, physical entity, or proposition
- Moulton (2009) generally analyses CPs as type $\langle e, t \rangle$
  - Assumes 'late merge' to reconcile these types

# Three-way polysemous nouns: *statement*

Three-way polysemous nouns, e.g., *statement* can be used in NCCs:

- The polysemy account makes predictions about which sense in NCCs should be activated

Predictions:

- Polysemy view ⇒ NCC constructions combined with predicates of physical entities (*in the filing cabinet*) should be degraded compared to predicates of events (*took but a few moments*)
- Contentful entities view ⇒ no contrast predicted without further assumptions

Polysemy view's prediction arguably borne out, although the judgment is subtle:

(40) a. Billie's statement that Alex lied took but a few moments (but its effects on their relationship were permanent).
b. ?Billie's statement that Alex lied is in the filing cabinet.
c. Billie's statement in which he claims that Alex lied is in the filing cabinet.

I also consulted two native speakers of German who detected a similar contrast

- Consultant's comment: the (b) variant, but not the (a) variant makes it clear that the statement is written, not spoken

(41) a. ?Berthas Stellungnahme, dass Alex gelogen hat, ist in dem Aktenschrank.
b. Berthas Stellungnahme, in der sie behauptet, dass Alex gelogen hat, ist in dem Aktenschrank.

# Takeaways

Not devastating for the contentful entities view

- Perhaps *statement* is polysemous between a contentful entities sense and a non-contentful entities sense

But then this begs the question:

- Are contentful entities in the denotation of e.g. *statement* not just eventualities?

Moreover:

- We have independent means of testing for whether a noun denotes eventualities
- We have no such for whether a noun denotes contentful entities

# *say* complementation

Grimshaw 2015; Major 2021

Nouns such as *book* are infelicitous in some say complementation:

(42) Grimshaw 2015, p.87

a. ??The sign/poster/book/article was saying that the park was closed.
b. ??The sign/poster/book/article said to the tourists that the park was closed.

But are not so bad in others:

(43) an historian called Paul Kennedy published a book saying that America was in decline as a super power [BNC]

*letter* has a similar pattern:

(44) a. The letter says that the park is closed.
b. The letter saying that the park is closed is on the sideboard.
c. ?The letter was saying that the park is closed.
d. ?The letter said to the recipient that the park is closed.