



Referential and lexical-semantic factors in center-embedding

Emily Davis¹ & Robert Kluender²

1. Uniwersytet Mikołaj Kopernik w Toruniu

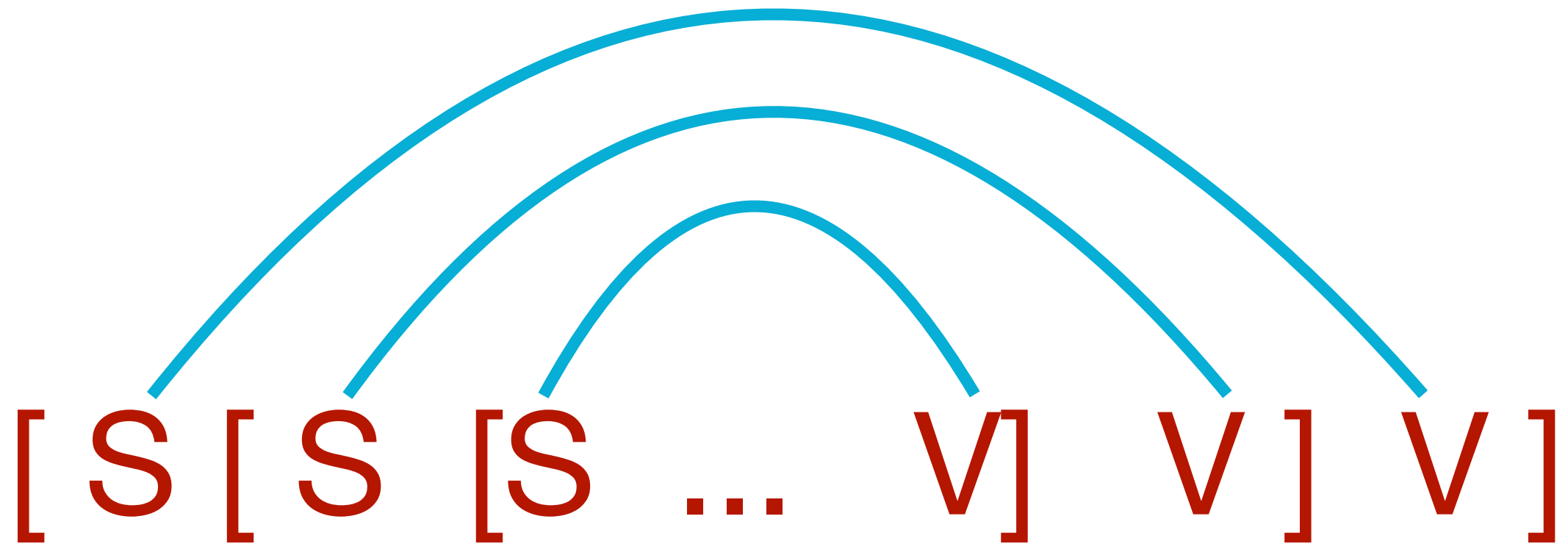
2. University of California, San Diego

General overview

- Center-embedding
 - Definition and significance
 - Experimental findings
- The role of referential and lexical factors
 - Subject referential form
 - Subject lexical semantics: concreteness/specificity

Center-embedding

General definition: nested dependencies



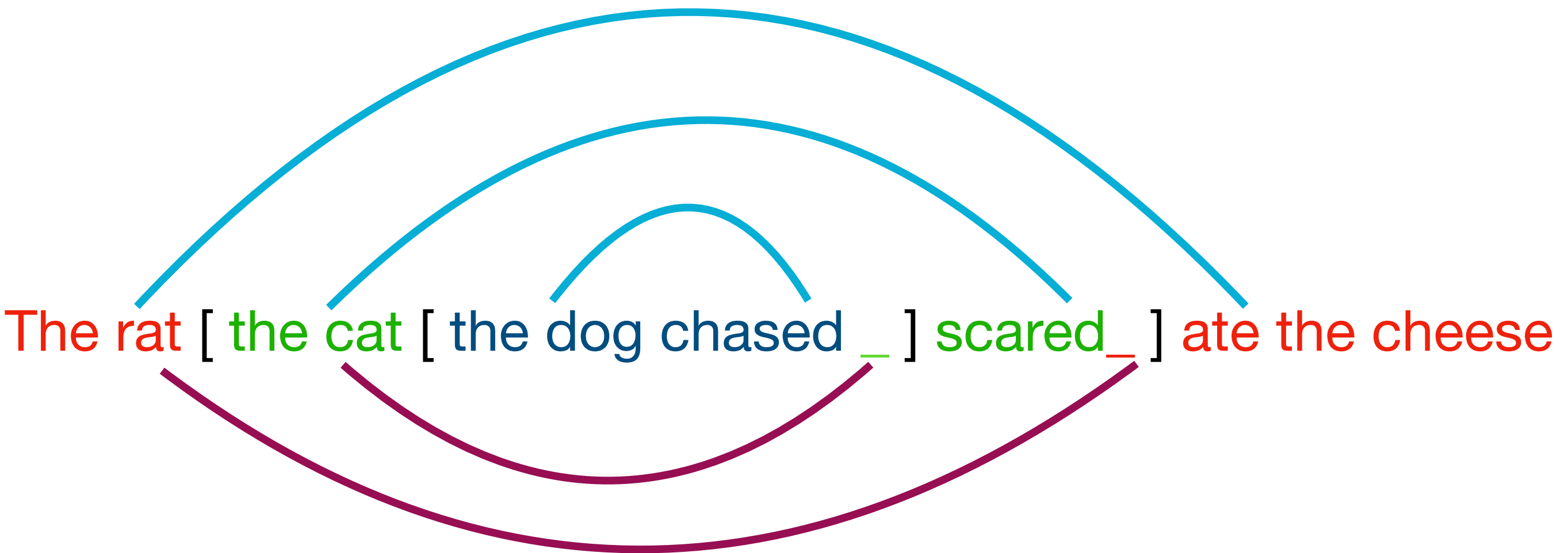
Center-embedding

Theoretical significance

- Formal grammar and recursion
 - context-free vs. finite-state grammar
- Processing
- Grammatical vs. acceptable
- Related syntactic issues
 - long-distance dependencies, islands

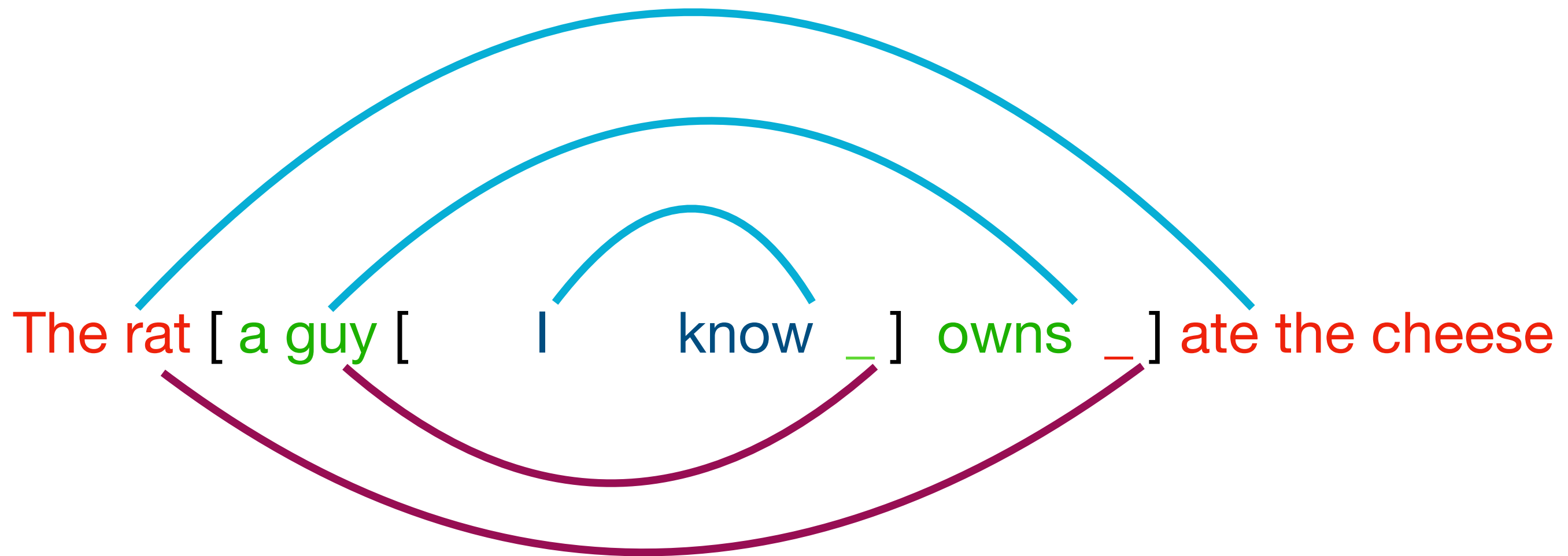
Center-embedding

Facilitating factors: subjects and verbs



Center-embedding

Facilitating factors: subjects and verbs



Center-embedding

Facilitating factors: reference form of subjects

Low Accessibility

- 
1. Full name + modifier
 2. Full "namey" name
 3. Long definite description
 4. Short definite description
 5. Last name
 6. First name
 7. Distal demonstrative + modifier
 8. Proximal demonstrative + modifier
 9. Distal demonstrative (+ NP)
 10. Proximal demonstrative (+ NP)
 11. Stressed pronoun + gesture
 12. Stressed pronoun
 13. Unstressed pronoun
 14. Clitic pronoun
 15. Gap or zero

High Accessibility

Center-embedding

Facilitating factors: reference form of subjects

C1 The student

C2 who the professor

C3 who {I/they/Jen/the scientist} collaborated with

C2 had advised

C1 copied the article.

(1st/2nd person pronoun >> 3rd person > name > def DP)

Center-embedding

Experimental evidence: reference form of subjects

The “deeper” the quantifier as subject,
the more acceptable the sentence

The journalist
[who the photographer
[who everyone met]
liked]
was at the party

>

The photographer
[who everyone
[who the journalist met]
liked]
was at the party

>

Everyone
[who the journalist
[who the photographer met]
liked]
was at the party

Center-embedding

Corpus findings: reference form of subjects

- Corpus studies (Karlsson 2007):
 - Multiple center-embedding is rare in English
 - (1/5000 sentences)
 - Pronouns preferred as subjects in embedded clauses
 - Especially the deepest embedded clause
- Extending these findings with a collection of attested examples...

Data collection

English multiple center-embedding “in the wild”



Data collection

English multiple center-embedding “in the wild”

C1 The chances

C2 that the persons

C3 who are now being frozen

C2 will return to life

C1 are practically indistinguishable from nil

Data collection

English multiple center-embedding “in the wild”

I regularly talk about how much I dislike kids, but

C1 the things

C2 people

C3 who profess utter devotion to their welfare

C2 do to them

C1 appall me

Data collection

English multiple center-embedding “in the wild”

C1 Something

C2 which every child

C3 you’ve met

C2 has taken for granted

C1 is seen as strange in other parts of the world

Center-embedding

Facilitating factors: reference form of subjects

Low Accessibility

- 
1. Full name + modifier
 2. Full "namey" name
 3. Long definite description
 4. Short definite description
 5. Last name
 6. First name
 7. Distal demonstrative + modifier
 8. Proximal demonstrative + modifier
 9. Distal demonstrative (+ NP)
 10. Proximal demonstrative (+ NP)
 11. Stressed pronoun + gesture
 12. Stressed pronoun
 13. Unstressed pronoun
 14. Clitic pronoun
 15. Gap or zero

High Accessibility

Subject referential form

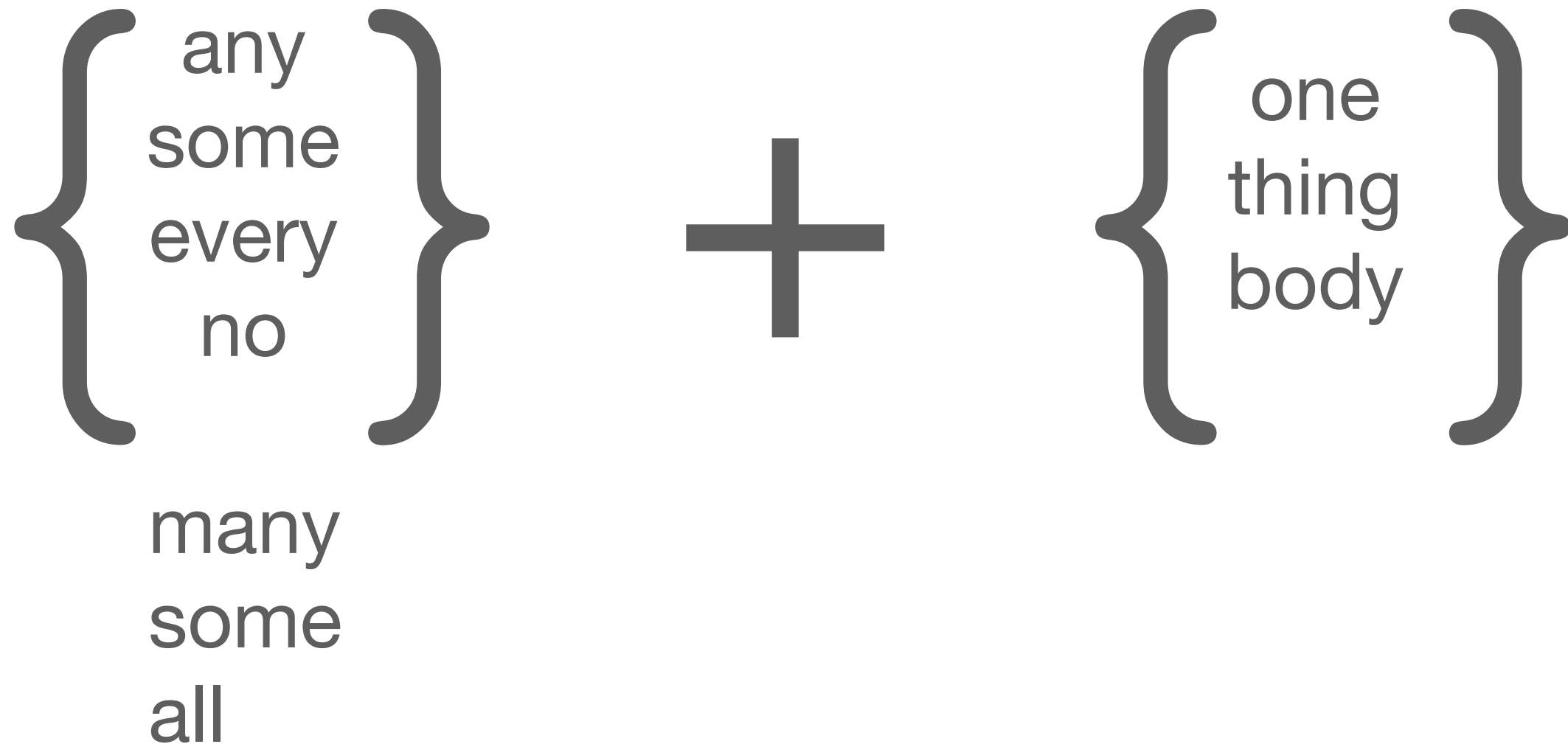
Classifying subjects: lower accessibility

- definites (specific)
 - *the* NP
- proper name
 - *Joe, Garfield, etc...*
- demonstrative
 - *this/that/these/those* NP(s)

Subject referential form

Classifying subjects: mid accessibility

- Quantifiers



Subject referential form

Classifying subjects: mid accessibility

- Quantified expressions
 - *some/all/many/few/no* NP
- Focused expressions
 - *the only/only the/only* NP

These involve an operator over a set:

- Quantifiers: individuals
- Only: alternatives

Subject referential form

Classifying subjects: mid accessibility “Generics”

- bare plurals
 - *people, things*
- mass nouns
 - *water, money*
- indefinite singulars
 - *an animal, a man*

Subject referential form

Classifying subjects: mid accessibility “Generics”

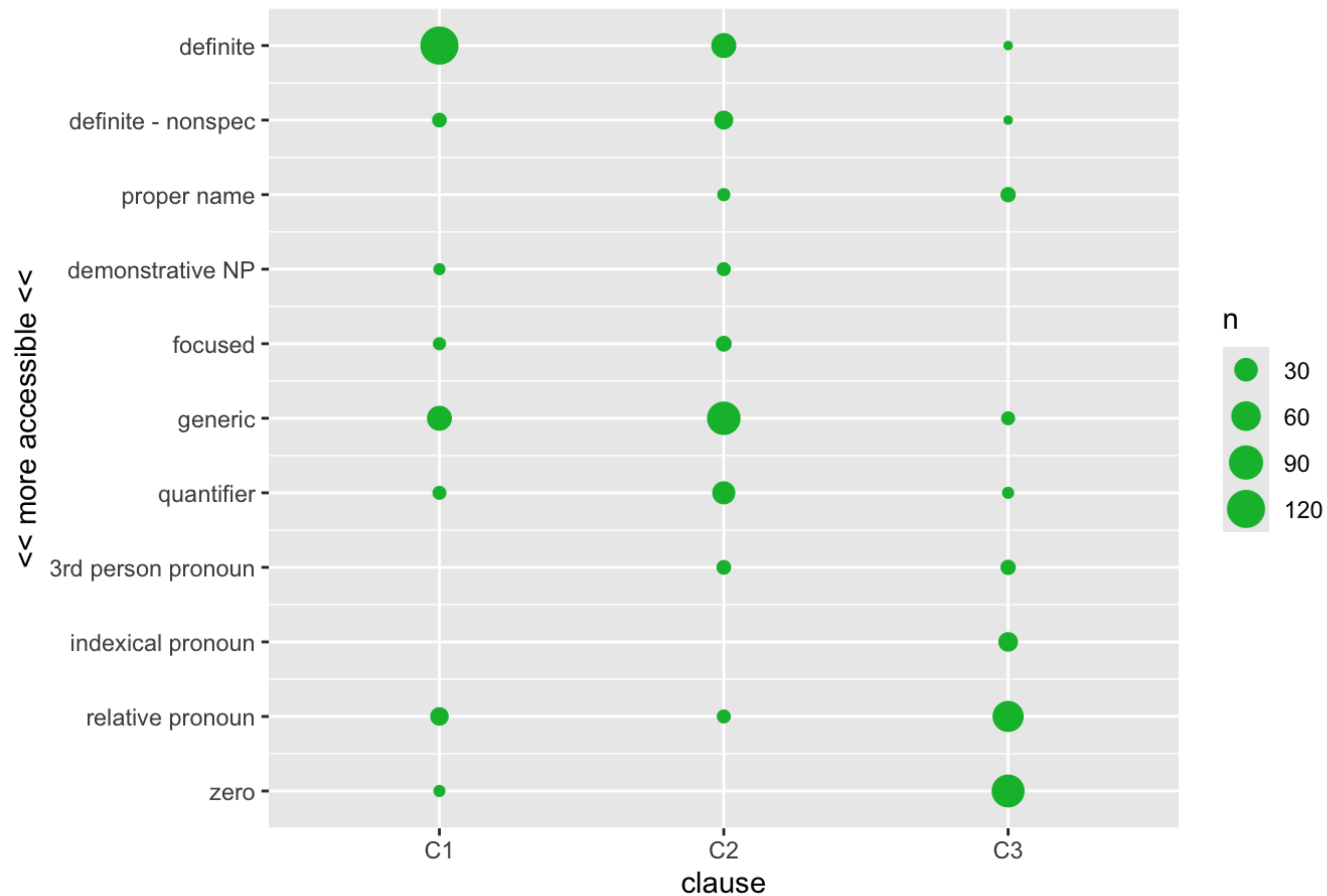
- Generic interpretation (not new referent)
 - bare plurals
 - ***Cats are mammals*** vs. *Cats were playing in the backyard*
 - mass nouns
 - ***Water is a liquid*** vs. *Water dripped down the walls*
- Indefinite singulars
 - ***A duck is a bird*** vs. *A duck landed on the roof*

Subject referential form

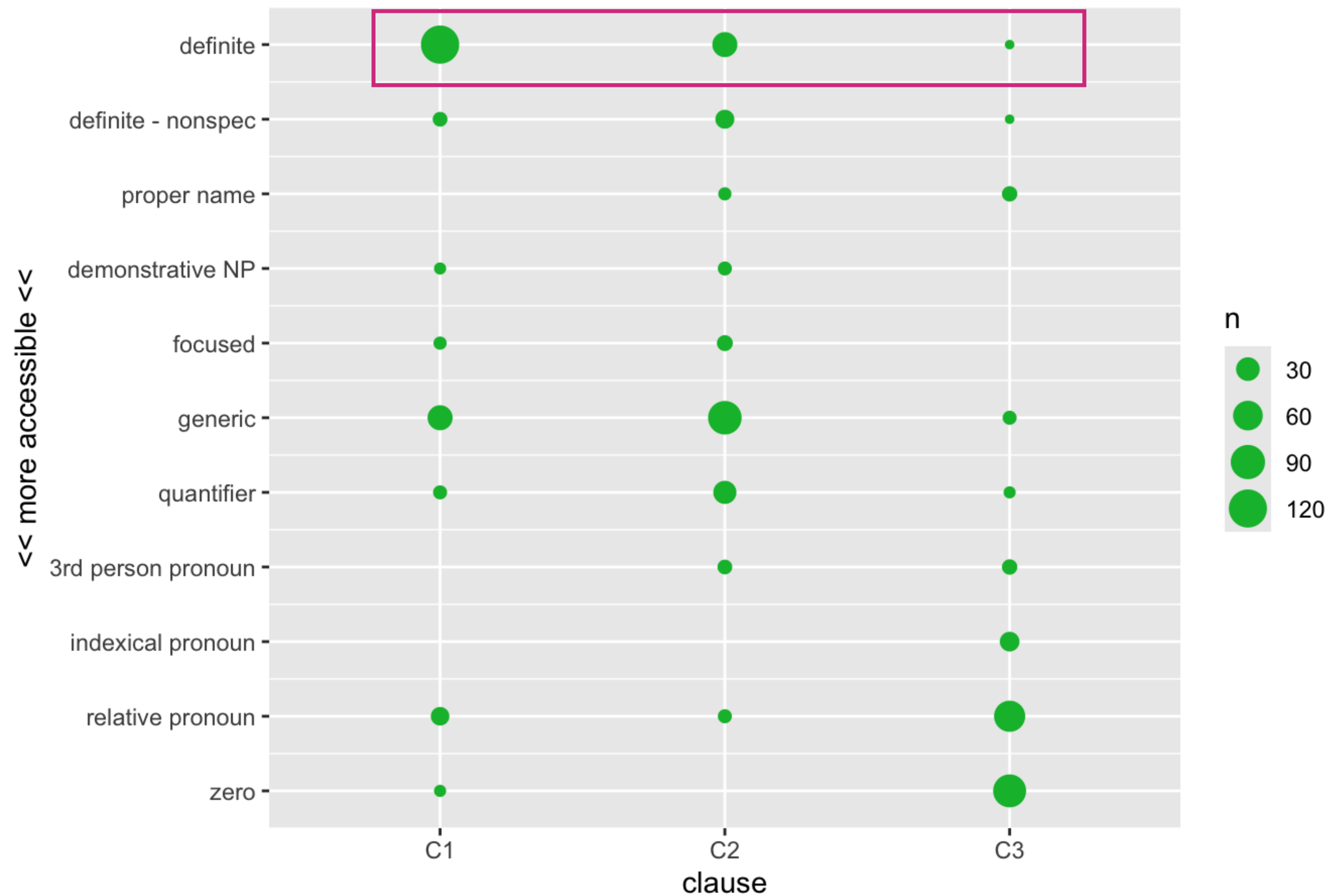
Classifying subjects: high accessibility

- Third person pronouns
 - *he, she...*
- Indexical pronouns
 - *I, you, we...*
- Relative pronoun in subject RC/free relatives
 - The man [**that** saw the cat]
 - **What** [you did] was wrong
- Zero forms
 - People [**Ø** using computers]

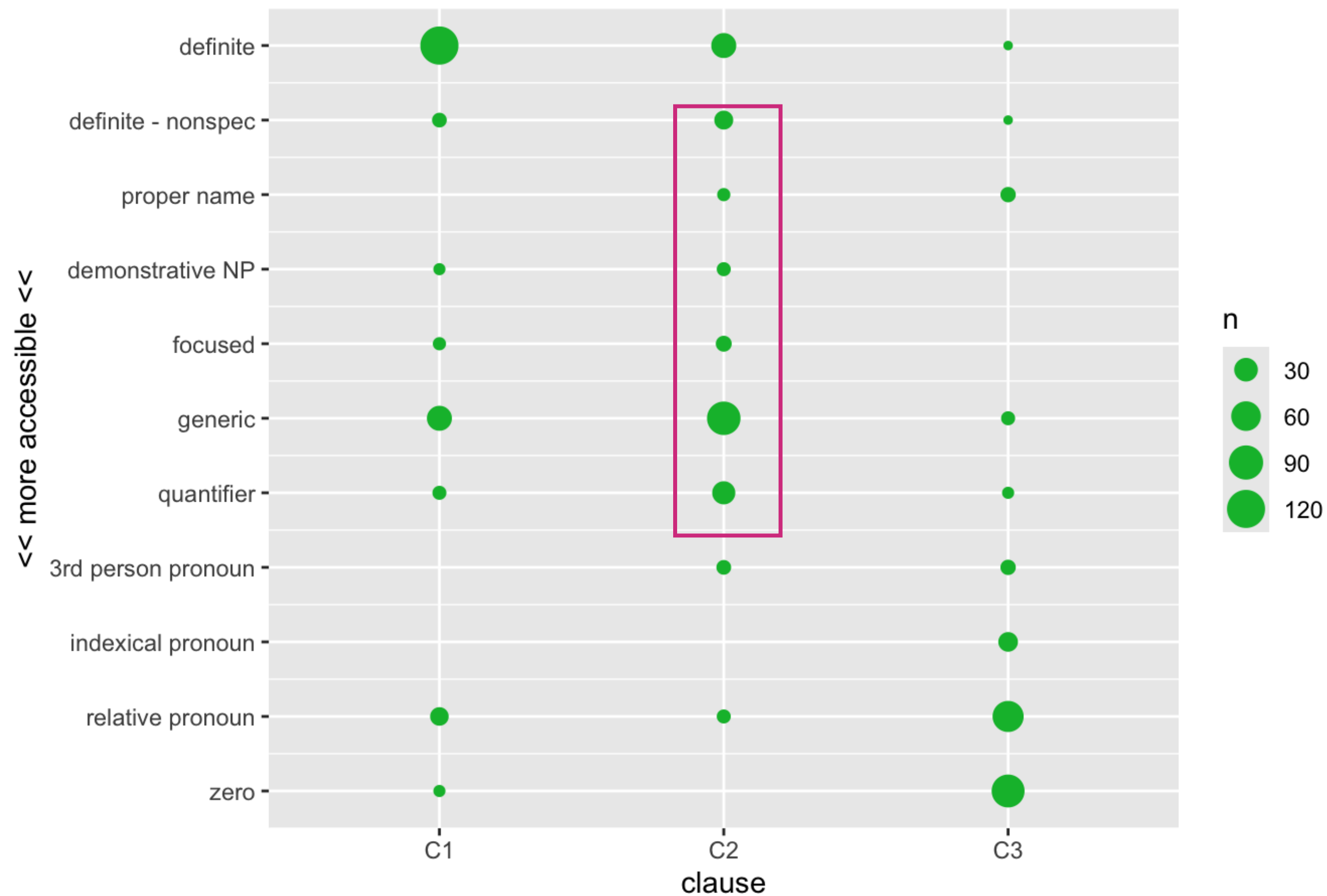
Type of subject with clause depth



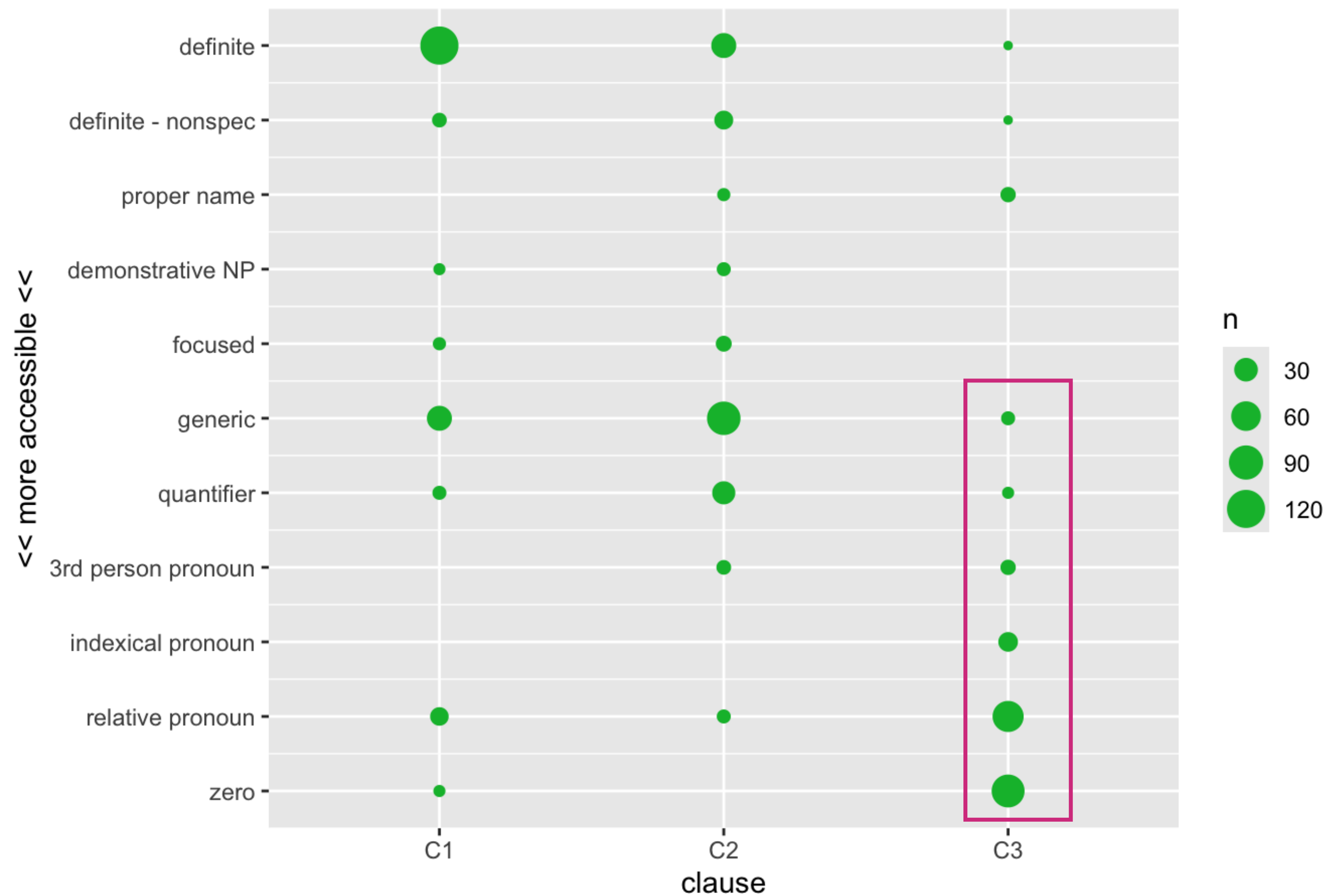
Type of subject with clause depth



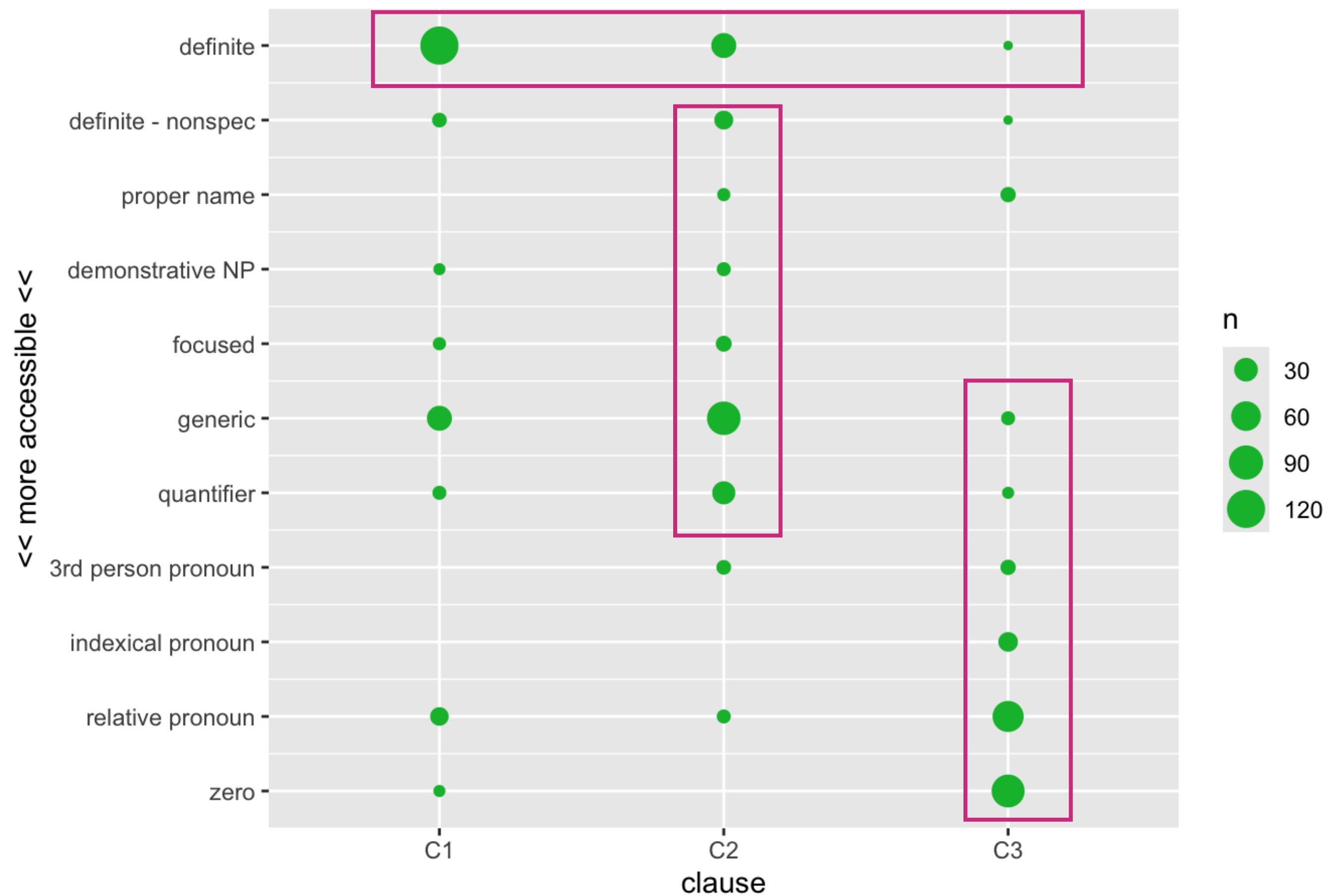
Type of subject with clause depth



Type of subject with clause depth



Type of subject with clause depth



Subject referential form

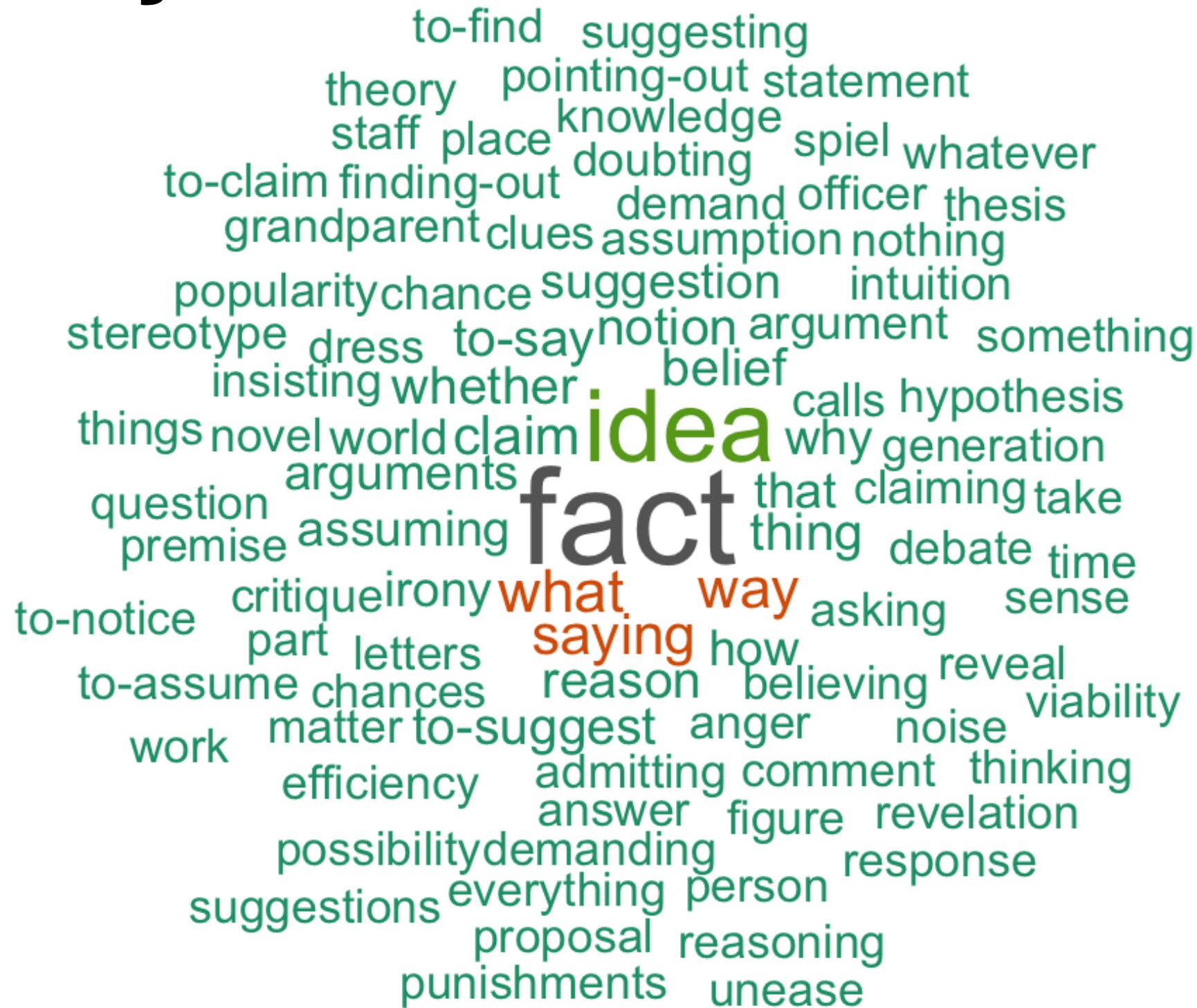
Overview

- Deeper embedded clause = more accessible subject
- C1: definite
- C2: generic/nonspecific/quantified
- C3: null or high-accessibility subject
- These subjects fit into a prototypical topic-comment discourse pattern
 - More on this later

Lexical semantics of clause subjects

- Within the subset of subjects that are common nouns (as opposed to quantifiers, pronouns, or proper names) what patterns appear?
- Certain words are very common in each clause

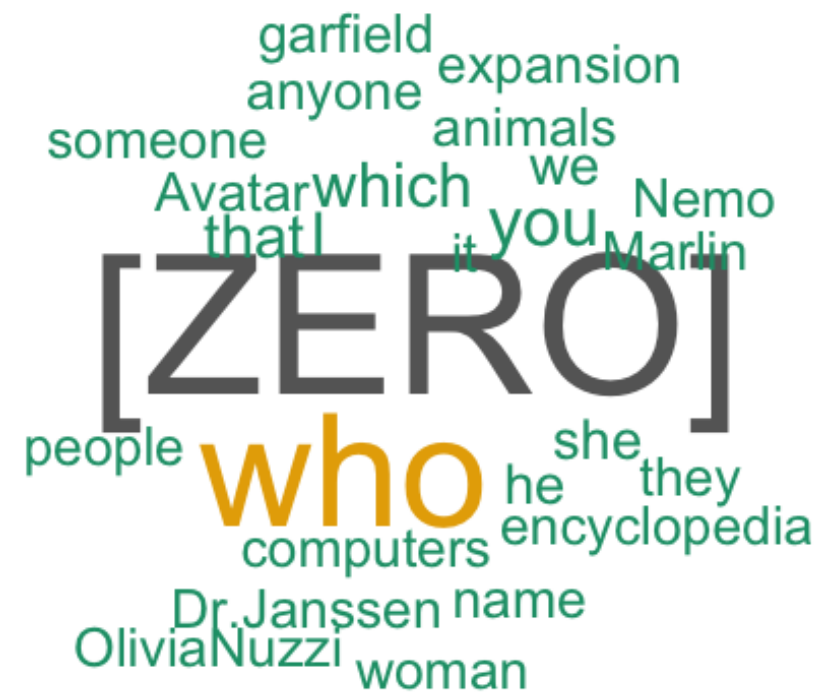
C1 subjects



C2 subjects



C3 subjects



Specificity and concreteness

Defining the concepts

- Specificity:

How broad or narrow a word's semantic extension is

animal < bird < duck < mallard

generic ← → *specific*



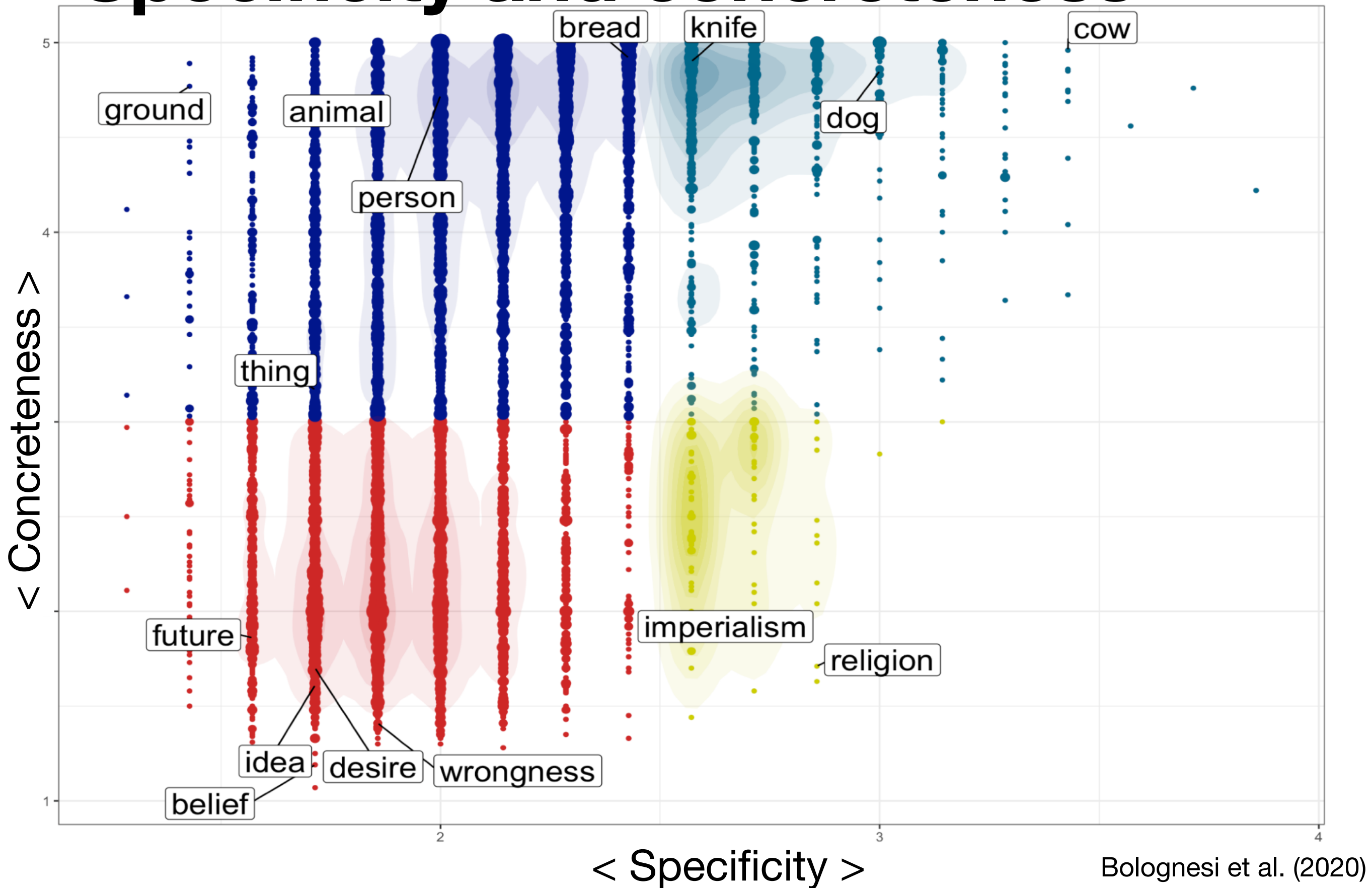
- Concreteness:

Whether a word refers to a tangible thing

belief, idea = abstract

tree, bread, animal = concrete

Specificity and concreteness



Specificity and concreteness

Processing effects

- Effect of specificity on relative clause processing (Hofmeister 2011)
 - More specific head nouns (e.g. *soldier* vs. *person*)
 - initially slow down processing
 - but facilitate retrieval of filler at gap
 - a. COMPLEX: The lieutenant could not remember **which soldier** the commander that was deeply respected ordered to scout the area ahead.
 - b. SIMPLE: The lieutenant could not remember **which person** the commander that was deeply respected ordered to scout the area ahead.

Specificity and concreteness

Processing effects

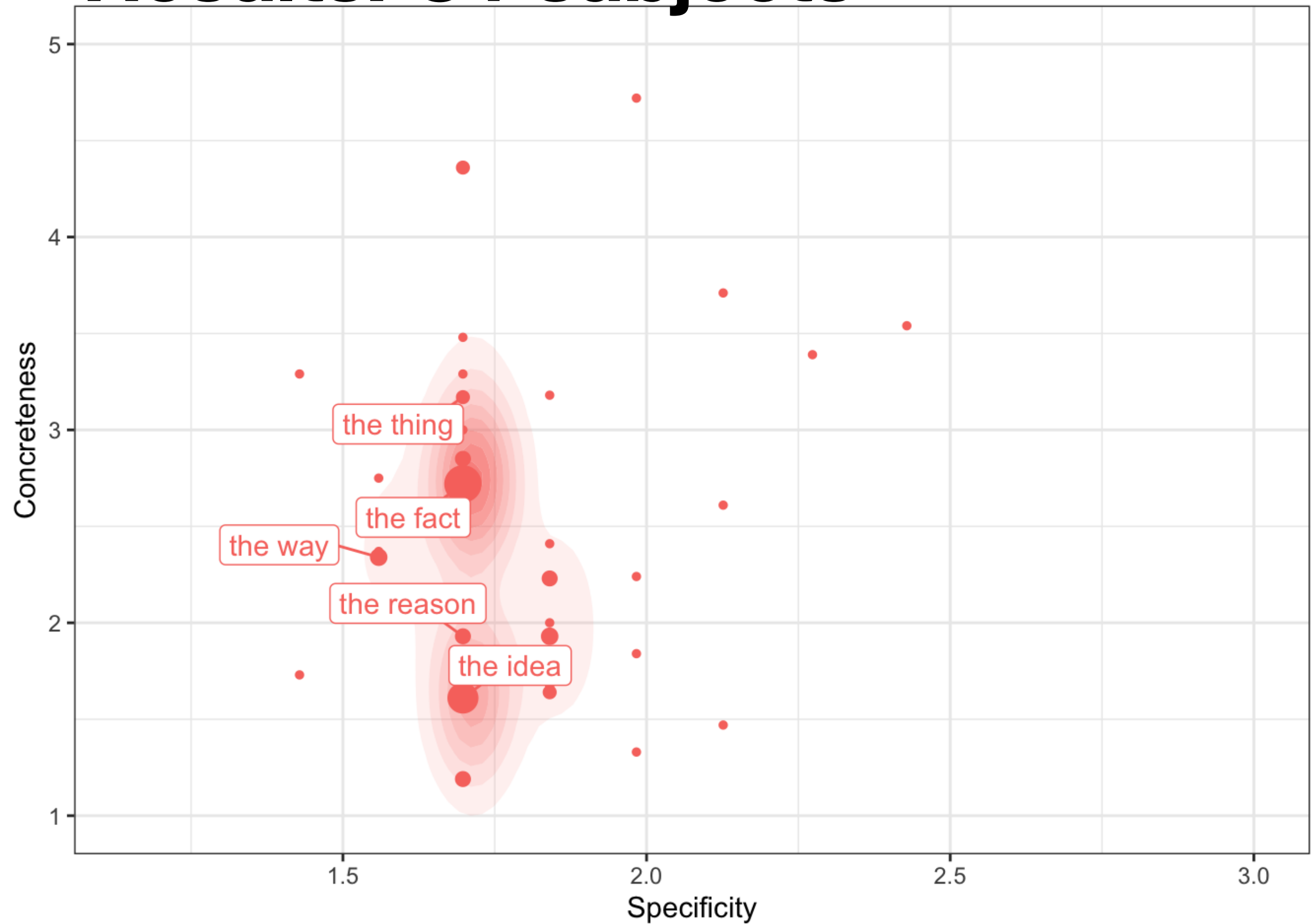
- Effects of concreteness and specificity on lexical decisions (Iliev & Axelrod 2017)
 - Higher specificity *prolongs* decision times
 - Higher concreteness *shortens* decision times

Specificity and concreteness

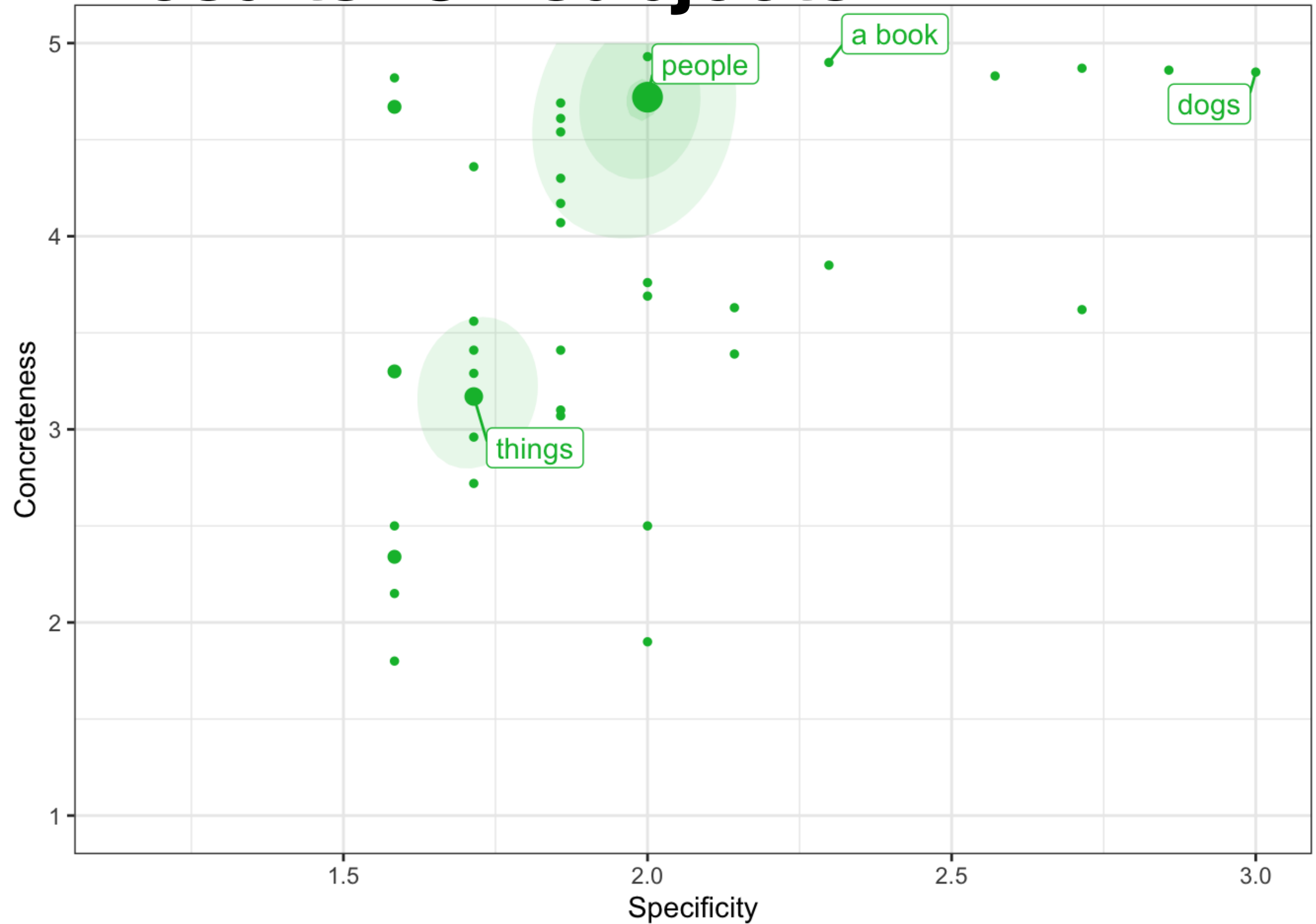
Implications for center-embedding

- Higher lexical specificity *impedes* processing
 - But may assist retrieval for relative clause heads
- Higher lexical concreteness *facilitates* processing
- More processable subjects are favored in embedded clauses
- ∴ We should expect to see low-specificity, high-concreteness subjects in embedded clauses
 - i.e. C2 and to a lesser degree C3
 - C3 usually avoids a new subject altogether

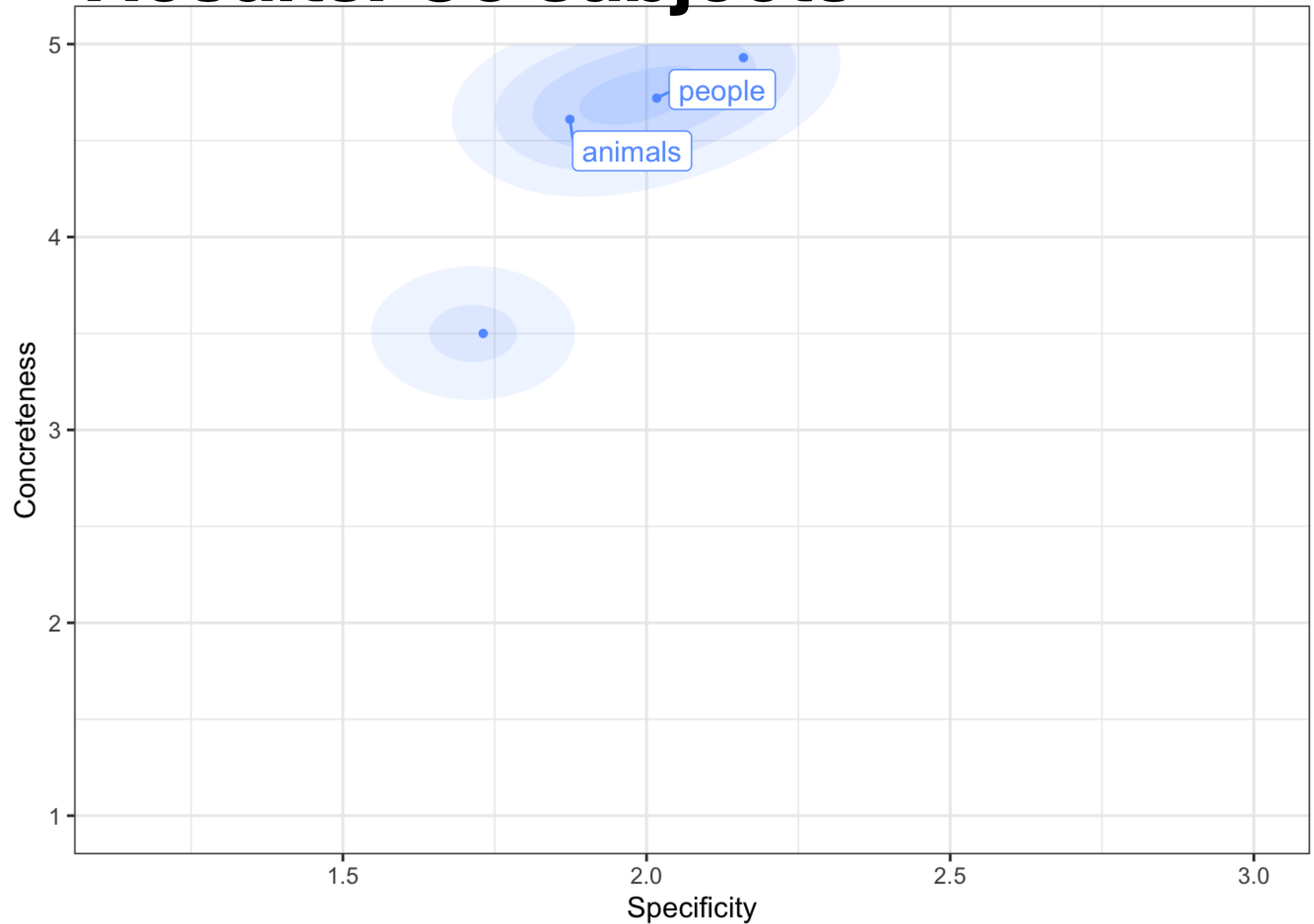
Results: C1 subjects



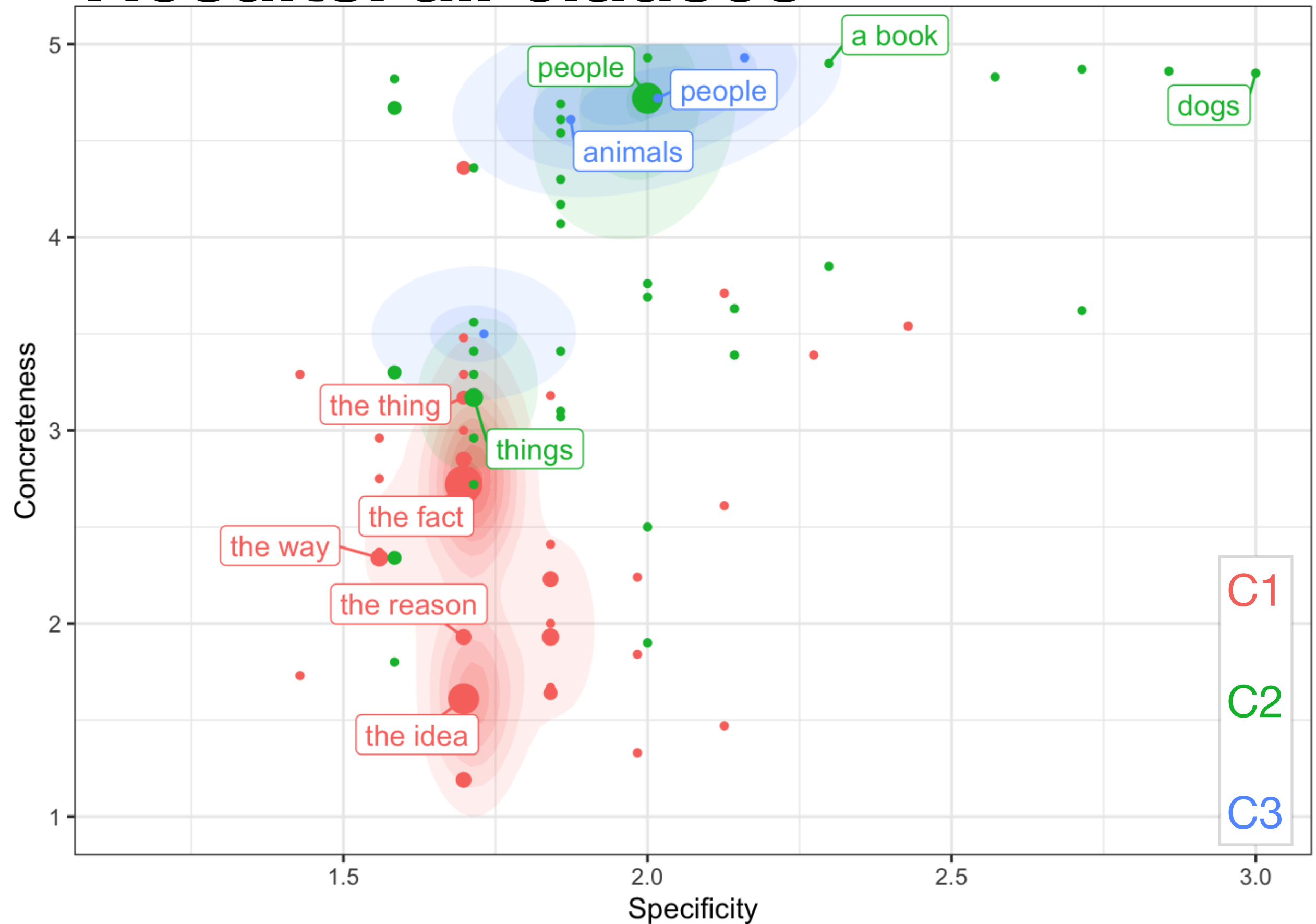
Results: C2 subjects



Results: C3 subjects

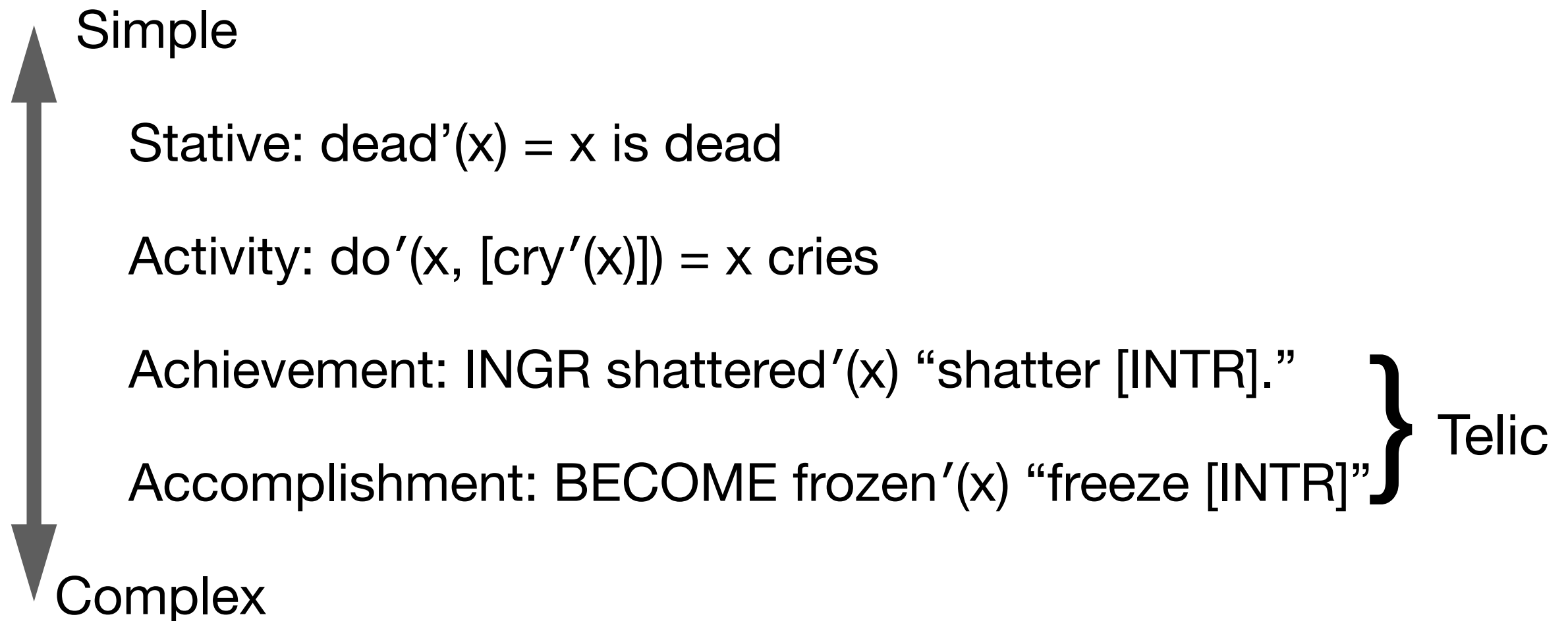


Results: all clauses

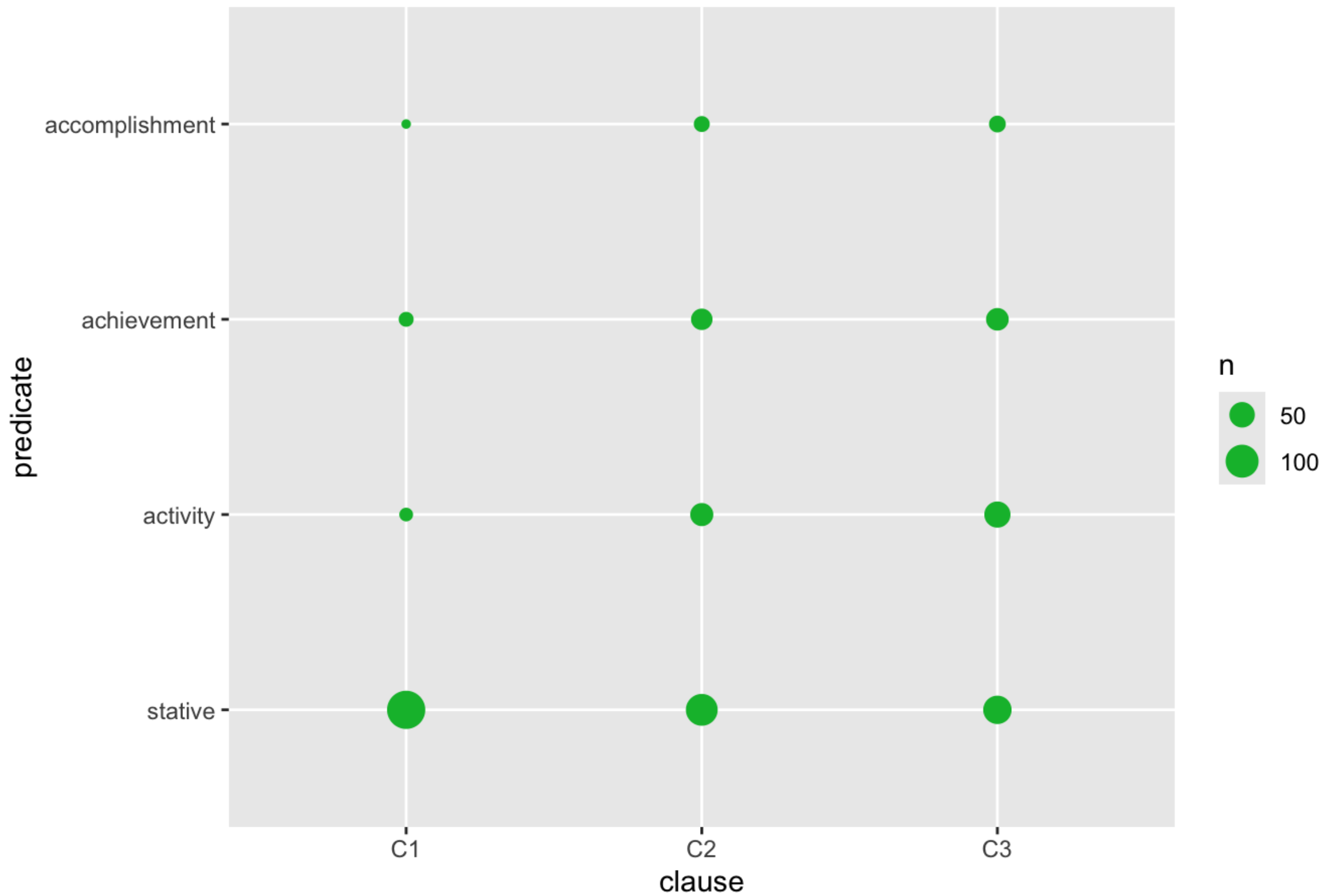


Predicates

Lexical aspect



Type of predicate with clause depth



Discourse structure

- Topic-comment discourse pattern
 - C1 is a comment on embedded clause C2, which is typically a proposition about a category/group
 - C2's subject (usually a generic) is further specified by C3
 - C3 therefore generally lacks a subject of its own

Discourse structure

C1 The chances abstract and definite subject as
 “anchor” for C2

C2 that the persons

C3 who are now being frozen

C2 will return to life

C1 are practically indistinguishable from nil

Discourse structure

C1 The chances

Complement clause

C2 that the persons

Concrete but generic subject

C3 who are now being frozen

C2 will return to life

C1 are practically indistinguishable from nil

Discourse structure

C1 The chances

C2 that the persons

C3 who are now being frozen

subject RC,

specifying subject of C2

Predicate is accomplishment

C2 will return to life

C1 are practically indistinguishable from nil

Discourse structure

C1 The chances

C2 that the persons

C3 who are now being frozen

C2 will return to life Predicate of C2 - achievement

C1 are practically indistinguishable from nil

Discourse structure

C1 The chances

C2 that the persons

C3 who are now being frozen

C2 will return to life

C1 are practically indistinguishable from nil

C1 predicate is stative,
comment on C2

Discussion

Related phenomena - Islands



- Relative clauses block WH-movement

*This is
[a paper]_i that the students criticized the professor [**who** understands __i].

*That's
[one trick]_i that the students mocked the teachers [**who**'ve been taken in by __i].

Discussion

Related phenomena - Islands



- Long dependencies become easier if the intervening material, especially any clause subjects, does not introduce new referents

This is
[a paper]_i that **we** really need to find **someone** [who understands _i].

That's
[one trick]_i that **I**'ve known a lot of **people** [who've been taken in by _i].

Discussion

Related phenomena - islands

Norwegian



(8) Pengar *avskyr* *jag* *folk* [som ger _]
money *hate* *I* *people* that give
'I hate people who give money [for Christmas]'
(Löwenadler 2015:43, ex. 15)



(9) Akvarell *beundrar* *jag* *alla* [som klarar av _]
watercolor *admire* *I* *all* that manage of
'I admire all those that are good at watercolor [painting]'
(Löwenadler 2015:43, ex. 14)

Discussion

Possibilities for further research

- Verb semantics
 - Classifying verbs by concreteness and specificity?
- Processing and acceptability experiments
 - Using adapted versions of attested examples, determine how the different factors interact

Discussion

Wrapping up

- Multiple center-embedding is not so unacceptable in English as commonly assumed
- Acceptable examples feature a range of referential and lexical-semantic factors that facilitate processing
 - More accessible subjects in embedded clauses
- Prototype discourse structure
- Parallels to factors that facilitate island exceptions
 - Lends support to a processing account of islands

Question time!

