Quantifying Relational Nouns in Corpora

Lelia Glass
School of Modern Languages
Georgia Institute of Technology

May 2021

Work in progress!
Sortal vs. relational nouns

- *Tree*: sortal, one-place, characterizes individuals
  - $\lambda x[\text{tree}(x)]$
  - Like intransitive verbs

- *Cousin*: relational, two-place, relates 2 individuals
  - $\lambda x\lambda y[\text{cousin}(x, y)]$
  - Like transitive verbs

- Consequences for modification, typology, acquisition, similarities between nouns and verbs, possession ...

Sortal vs. relational nouns

- *Tree*: sortal, one-place, characterizes individuals
  - *My tree* — possessor-head relation is flexible, depends on context
  - “Alienable” — temporary, incidental
    - (in some languages, morphologically marked)

- *Cousin*: relational, two-place, relates 2 individuals
  - *My cousin* — possessor-head relation is (by default) provided lexically
  - “Inalienable” — inherent
    - (in some languages, morphologically less marked)

---

Which nouns are relational?

- The ??tree of Jane
- I found out your ??tree
- A man walked in with his ??tree
- Can identify a tree without reference to other entities
- The cousin of Jane
- I found out your cousin
- A man walked in with his cousin
- Cannot identify a cousin without reference to other entities

Which nouns are relational?

• The tree of Jane
• I found out your tree
• A man walked in with his tree
• Can identify a tree without reference to other entities

• Debated, contradictory, inconclusive, too binary?
  • Is phone relational?
  • Some theories also allow type-shifting... [stay tuned]

Barker 1992, Vikner & Jensen 2002, Partee & Borschev 1998...
Relational nouns
Approximating relationality

- Lübner 2011: Different nouns/referents “match” different types of determiners
  - Unique nouns prefer *the* — *the sun*
  - Relational nouns prefer possessives — *my cousin*
- Nouns are more frequent, less marked, require less contextual support with “matching” determiners
- Relational nouns are more likely to appear in possessives than other nouns; most possessives involve relational nouns
- Using the researcher’s own labeling of “relational nouns”

Nissim 2004, Lübner 2011, Kolkmann 2016...
%Possessive — Definition

• Instead, make no assumption about which nouns are relational — use %Possessive as a proxy for relationality

• Of all tokens of a given noun type, what % are possessive?

• Data: all 2-word noun phrases in 5 million words of comments from AskReddit, January 2018

• *My car, their idea, Mike’s sister* — possessive

• *The problem, an office, some shoes* — non-possessive
%Possessive — Validation

- *Tree*: <6% possessive
  - we had a huge monsoon [...] Blew my tree nearly over
  - yeah, when you hit a tree at 60 mph it doesn't really matter if you're in a new car

- *Cousin*: 88% possessive
  - I have fond memories [...] playing in the sand with my cousins
  - In the UK it’s legal to marry your cousin
  - i would feel more conflicted if i had grown up knowing her as a cousin
%Possessive — Validation

- Computed %Possessive for all noun lemmas in AskReddit
- Labeled each for its ontological class in WordNet

  - kinship (cousin)
  - body-part (foot)
  - human (snob)
  - abstract (content)
  - artifact (phone)
  - occupation (doctor)
  - location (area)
  - natural kind (tree)
  - excluded “unknown/other” (DNC)

Miller et al 1995 (WordNet)
Possessive — Validation

- Kinship, body-parts are most possessive — cousin, foot
- The most prototypical relational nouns
- Natural kinds are least possessive — tree
- The most prototypical sortal nouns

Possessive and non-possessive token counts in AskReddit for nouns in each ontological class

R Core Team 2012
%Possessive — Validation

% of possessive tokens in AskReddit for nouns in each ontological class

- natural_kind
  - tree: N=6233
- location
  - area: N=9031
- occupation
  - doctor: N=4974
- artifact
  - phone: N=32021
- abstract
  - content: N=112312
- human
  - snob: N=26976
- bodypart
  - foot: N=9665
- kinship
  - cousin: N=17692

R Core Team 2012
%Possessive — Validation

- %Possessive correlates highly significantly (in linear regressions) with hand-labeled datasets of relational nouns from NomBank and Williams 2018.
%Possessive — Advantages!

- “Good-enough” approximation (I argue) for relationality
  - Continuous
  - Objective
  - Easily computed across the lexicon
- Allows us to investigate:
  - *Which nouns are more or less relational, and why?*
  - Socio-cultural dimension of relationality/possession?
  - Synchronic, diachronic variation?
Which nouns are relational?

• The ontological class of the noun(’s referent) matters!
  • Kinship, body parts *(cousin, foot)* — describe relations?
  • Abstractions *(willingness, content)* — retain argument structure of underlying verb/adjective?
  • Artifacts *(book, phone)* — related to a creator/user?
  • Natural kinds *(tree, cloud)* — hard to infer a possessive relation

• People interact with each ontological class in different ways

Which nouns are relational?

- It matters how people/societies (conventionally, culturally) interact with an entity!
  - Culturally immanent artifacts are inalienably possessed
    - *My arrow*
  - *My car* — “more relational” than *my bus* because more common to own a car?
  - *My toothbrush* — easily interpreted as relational because everyone uses their own
  - *My cat* — “more relational” because they’re common pets?

Conventional interaction

• Claim:
  • A noun is *more relational* ...
    • (as measured by %Possessive)
  • ... when human interaction with its referent is *more conventional*
    • But how to measure conventional interaction??
      • Like relationality, not easy to quantify!
Approximating convention

- A noun is *more relational* when human interaction with its referent is *more conventional*

- But how to measure conventional interaction??

  - Two metrics:
    - Per-million-word count
    - Definiteness ratio

  - Two strategies:
    - Compare across nouns
    - Compare across communities
Approximating convention

• A noun is *more relational* when human interaction with its referent is *more conventional*

• But how to measure conventional interaction??

  • Per-million-word count

    • The more conventionally people interact with something,...

    • the more they might talk about it
Approximating convention

- A noun is *more relational* when human interaction with its referent is *more conventional*

- But how to measure conventional interaction??

  - Definiteness ratio
    - The more conventionally people interact with something, ...
    - the more they might treat it as (easily inferred to be) discourse-familiar and thus definite

  - At this barn, *the horses* see *the vet* once a year

Clark 1975, Spenader 2001, Roberts 2003
Across nouns, across communities

- A noun is *more relational* when human interaction with its referent is *more conventional*

- But how to measure conventional interaction??

- Compare across nouns within AskReddit
  - *Phone vs. kite*

- Compare the same noun to itself across communities (subreddits)
  - *Knife* for cooks vs. laypersons

- Different communities use different conventions!
A noun is *more relational* when human interaction with its referent is *more conventional*

Predict:
- More-%Possessive nouns...
  - *Proxy for: More relational*
  - ... should be more frequent
    - *Proxy for: More conventional interaction with their referent*

TRUE across nouns, TRUE across communities!
Freq. across nouns

- Excluded abstract nouns (*willingness*) — morphologically complex, unclear how people interact with them.

- For 1702 unique nouns in 5 million words of AskReddit, collected:
  - %Possessive, per-million-word count, ontological class.
Freq. across nouns

- \( m \leftarrow \text{lm}(\text{percent\_poss} \sim \text{ont\_class} + \text{pmw\_count}, \text{data}=d) \)
  - Ont class: kinship, bodyparts significantly more possessive, natural kinds significantly less so
  - Significant positive effect of pmw\_count (\( \beta = 0.06, t=20, p < 0.001 \)) — as predicted!
  - \( R^2 = 37\% \)
Freq. across nouns
Freq. across communities

- Used Fisher Exact Test to identify 136 nouns that are significantly more often possessive in a particular specialty subreddit vs. AskReddit

<table>
<thead>
<tr>
<th></th>
<th>Possessive</th>
<th>Not Possessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>AskReddit</td>
<td>N=4 Go ahead and bring your knife to a gun fight</td>
<td>N=47 yet another celeb who has gone under the knife to alter their appearance</td>
</tr>
<tr>
<td>Cooking</td>
<td>N=10 Press the parsley stalks with the side of your knife</td>
<td>N=15 I never had [a peeler] before and usually did it with a knife</td>
</tr>
</tbody>
</table>
Freq. across communities

• Examples of nouns (136 total, 82 unique) used significantly more often as possessive in 32 specialist subreddits
  
  • r/Horses: blanket, horse, owner, vet....
  
  • r/Cooking: counter, kitchen, knife....
  
  • r/BabyBumps: baby, child, doctor, hospital, shower....
Freq. across communities

• Wilcoxon Test for paired samples, comparing pmw_count in AskReddit vs. in the speciality subreddit in which they’re more often possessive
  
• $V = 1531, p = < 0.001$

• AskReddit pmw median = 49, specialty pmw median = 170

• As predicted, the same noun is more frequent in the specialist subreddit in which it’s more often possessive
Freq. across communities

log per-million-word counts of each noun in AskReddit vs in the speciality subreddit where it's more often possessive

ont_type
- artifact
- human
- location
- occupation
- bodypart
- kinship
- natural_kind

log_per-million-word count

log_askreddit_pmw  log_specialty_pmw
Definiteness ratio

• A noun is *more relational* when human interaction with its referent is *more conventional*

• Predict:
  • More-%Possessive nouns...
    • *Proxy for: More relational*
  • ... should be more definite
    • *Proxy for: More conventional interaction with their referent*
      • *A box, the box, my box* = 50% definite

• FALSE across nouns; but TRUE across communities!
Def. ratio across nouns

• m <- lm(percent_poss ~ pmw_count + ont_class * percent_def, data=d)

• Significant effects of:
  • ont_class
  • pmw_count
  • interaction between ont_class and percent_def

• But no main effect of percent_def — contrary to prediction
Def. ratio across nouns
Def. ratio across communities

- Same 136 (82 unique) non-abstract nouns found in a Fisher Exact Test to be more often possessive in a specialty subreddit vs. AskReddit
  - r/Horses: blanket, horse, owner, vet....
  - r/Cooking: counter, kitchen, knife....
  - r/BabyBumps: baby, child, doctor, hospital, shower...
Def. ratio across communities

- Wilcoxon Test for paired samples, comparing percent_def in AskReddit vs. in the specialty subreddit in which they’re more often possessive
  - $V = 2806$, $p = < 0.001$
  - AskReddit median = 71%, specialty median = 79%
  - As predicted, the same noun is more often definite (vs. indefinite) in the specialty subreddit in which it’s more often possessive
Def. ratio across communities

percent definite (vs. indefinite) uses of each noun in AskReddit vs in the speciality subreddit where it's more often possessive

ont_type
- artifact
- human
- location
- occupation
- bodypart
- kinship
- natural_kind
Mostly as predicted!

• Predict:
  • More-%Possessive nouns...
    • *Proxy for:* More relational
  • ... should be more frequent, more often definite
    • *Proxy for:* More conventional interaction with their referent
  • Across nouns: More-possessive nouns are more frequent, but not more definite [perhaps confounded by other factors...]
  • Across communities: The same noun is more frequent, more often definite where it’s more possessive
Examples (across nouns)

- *Phone* vs. *kite* in AskReddit
  - Per million: 168 for *phone*, 2 for *kite*
  - %Possessive: 67% for *phone*, 0% for *kite*
  - %Definite: 75% for *phone*, 0% for *kite*

  - As a 16 yo, I shouldn't need restrictions on how long I’m using my *phone*, right?
  - He got fired for repeatedly showing up to work high as a *kite*

- Arguably: *phone* is “more relational” than *kite* because people interact more conventionally with phones than kites

  - (both artifacts)
Examples (across nouns)

- *Dog* vs. *horse* in AskReddit
  - %Possessive: 46% for *dog*, 13% for *horse*
  - Per million: 191 for *dog*, 34 for *horse*
  - %Definite: 57% for *dog*, 31% for *horse*
    - Currently watching Netflix with *my dog* on my lap
    - I live in Texas and I’ve never ridden a *horse* here
- Arguably: *dog* is “more relational” than *horse* because people interact more conventionally with dogs than horses
  - (both natural kinds)
Examples (across communities)

- *Knife* in r/Cooking vs. r/AskReddit
  - 40% possessive in Cooking, 8% in AskReddit
  - Per million: 320 in Cooking, 10 in AskReddit
  - 33% definite in Cooking, 30% in AskReddit
  - Flailing *the knife* on the stone [is] inefficient and unsafe for a beginner (Cooking)
  - I think *the knife* was just a coincidence, she’s not a murderer (AskReddit)
  - Arguably: *knife* is “more relational” for cooks because cooks interact more conventionally with knives than other people
Examples (across communities)

- *Horse* in r/Horses vs. r/AskReddit
  - 38% possessive in Horses, 13% in AskReddit
  - Per million: 4895 in Horses, 34 in AskReddit
  - 43% definite in Horses, 31% in AskReddit
  - *My horse* is still barefoot and never needed shoes before, during, and after having white line disease (Horses)
  - I live in Texas and I’ve never ridden a horse here (AskReddit)
  - Arguably: *horse* is “more relational” for equestrians because they interact more conventionally with horses
Review!

- Relational nouns (*cousin*) — binary, theory-dependent idea
- Here instead: continuous, objective proxy for relationality
  - %Possessive
  - Which nouns are more or less relational, and why?
- Claim: *A noun is more relational when human interaction with its referent is more conventional.*
  - Corpus evidence (approximating convention via frequency, definiteness ratio; across nouns, across communities)
Significance — Formal semantics

- Depending on your theory, maybe...
  - Some nouns (*cousin*) are inherently relational, others (*tree*) are inherently sortal [Barker 1992, Vikner & Jensen 2002]
    - Sortal nouns are made possessive via different means from relational nouns?
    - Sortal nouns type-shifted to relational in possessives?

Significance — Formal semantics

- Depending on your theory, maybe...
  - All nouns are inherently sortal [Payne et al 2013]
    - Possessive construction requires “free R” relation between possessor, head — saturated by context
  - Easier for some nouns (cousin) than others (tree)

Significance — Formal semantics

- None of these theories explains which nouns are relational or not and why
- So they all need to be complemented by a theory that tries to answer that question — like the one offered here!

Significance — Lexical semantics

- Study lexical semantics at the scale of the lexicon
- Challenge & promise of approximating abstract ideas via corpus metrics
  - Can question validity of metrics for relationality & convention
  - But at least they allow for hypothesis-testing!

- Down to its structure, language is social!
  - The semantic type of a noun is shaped by the conventions of the people who use that noun
Thank you!

Thanks for being the first-ever audience to hear about this topic!

I am grateful for your feedback & time!

lelia.glass@modlangs.gatech.edu
Selected references

- Clark, Herbert. 1975. Bridging.
Selected references

- Karvovskaya, Lena. 2018. The typology and formal semantics of adnominal possession.
- Levin, Beth, Glass, Lelia, & Jurafsky, Dan. 2019. Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds.
Selected references

- R Core Team. 2012. R: A Language and Environment for Statistical Computing.
- Williams, Adina. 2018. Representing Relationality,