

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 ...

Löbner 1985, de Bruin & Scha 1988, Barker 1992, Asmuth & Gentner 2005, Löbner 2011, Barker 2011....

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)

de Bruin & Scha 1988, Barker 1992, Heine 1997, Asmuth & Gentner 2005, Barker 2011, Karvovskaya 2018....

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*

Barker 1992, Genter & Asmuth 2005, Barker 2011, Barker 2016....

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*
- 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

- Löbner 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, Löbner 2011, Kolkman 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

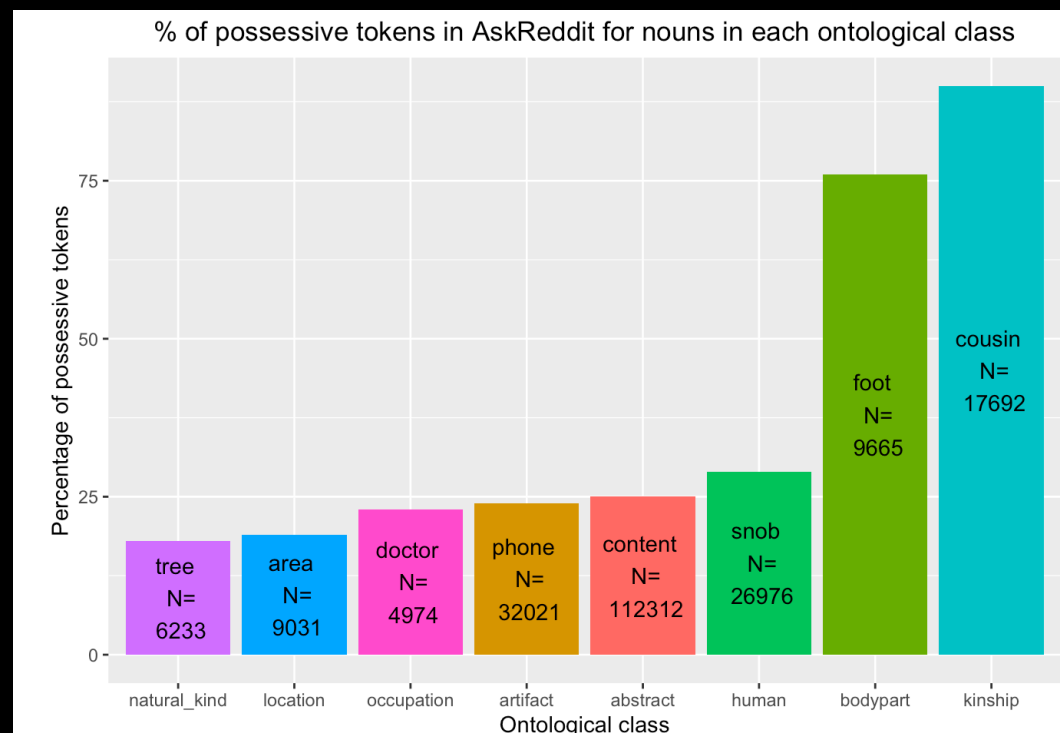
- *Tree*: <6% possessive
 - we had a huge monsoon [...] Blew my tree nearly over
 - yeah, when you hit a tree at 60 mph it doesn't really matter if you're in a new car
- *Cousin*: 88% possessive
 - I have fond memories [...] playing in the sand with my cousins
 - In the UK it's legal to marry your cousin
 - 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*)

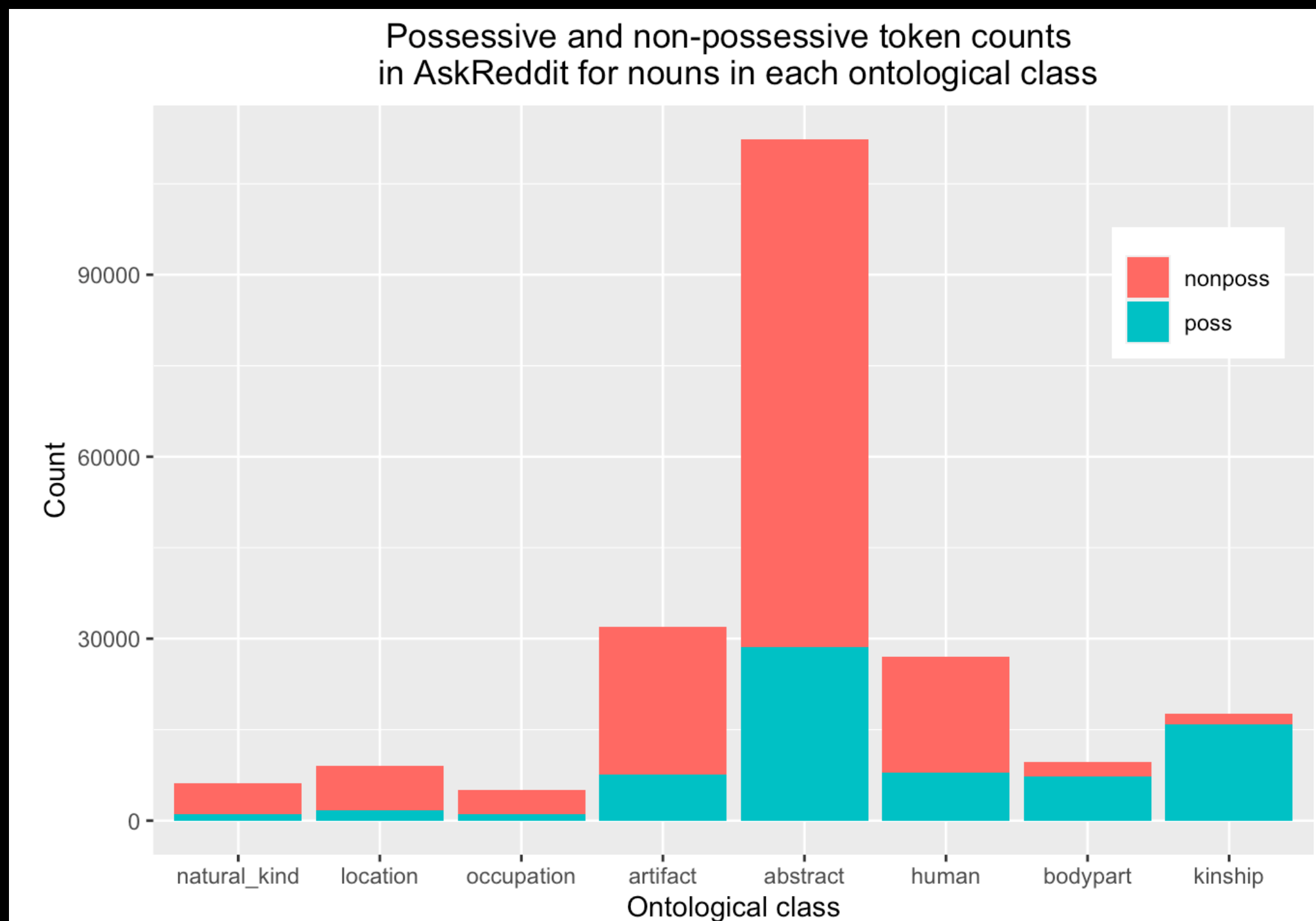
%Possessive — Validation

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

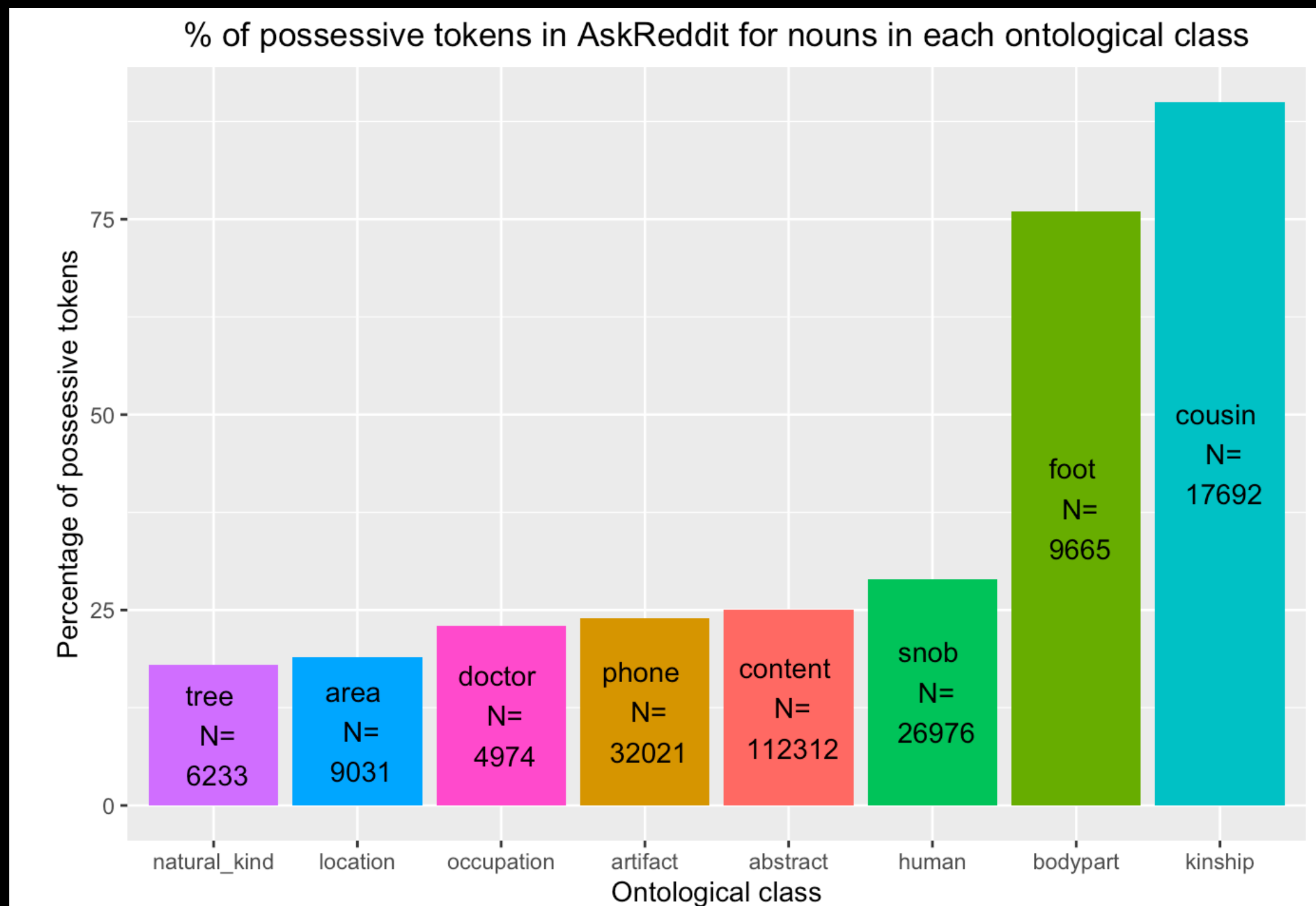


Nichols 1988, Heine 1997, Aikhenvald 2012

%Possessive — Validation

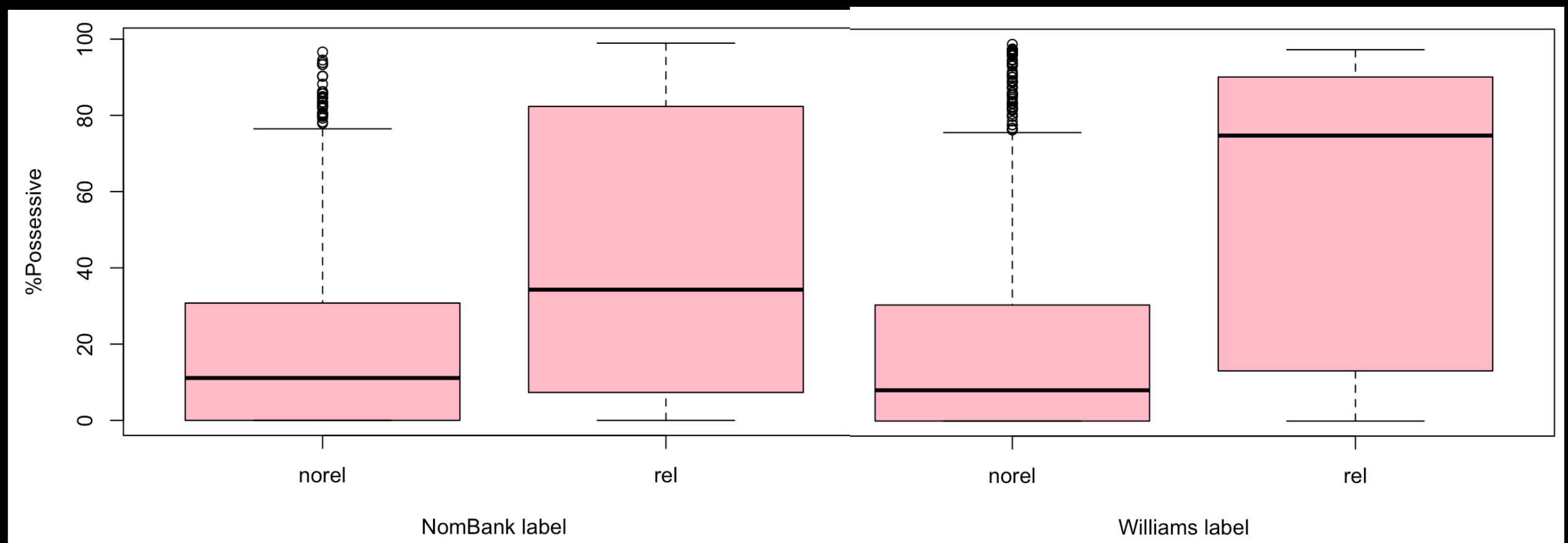


%Possessive — Validation



%Possessive — Validation

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



Meyers et al 2004 (NomBank), 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

Barker 1992, Heine 1997, Vikner & Jensen 2002, Grimm 2012, Peters & Westerståhl 2013, Levin et al 2019

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?

Nichols 1988, Barker 1992, Heine 1997, Vikner & Jensen 2002, Löbner 2011, Aikhenvald 2012, Karvovskaya 2018

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, Spence 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!

Frequency

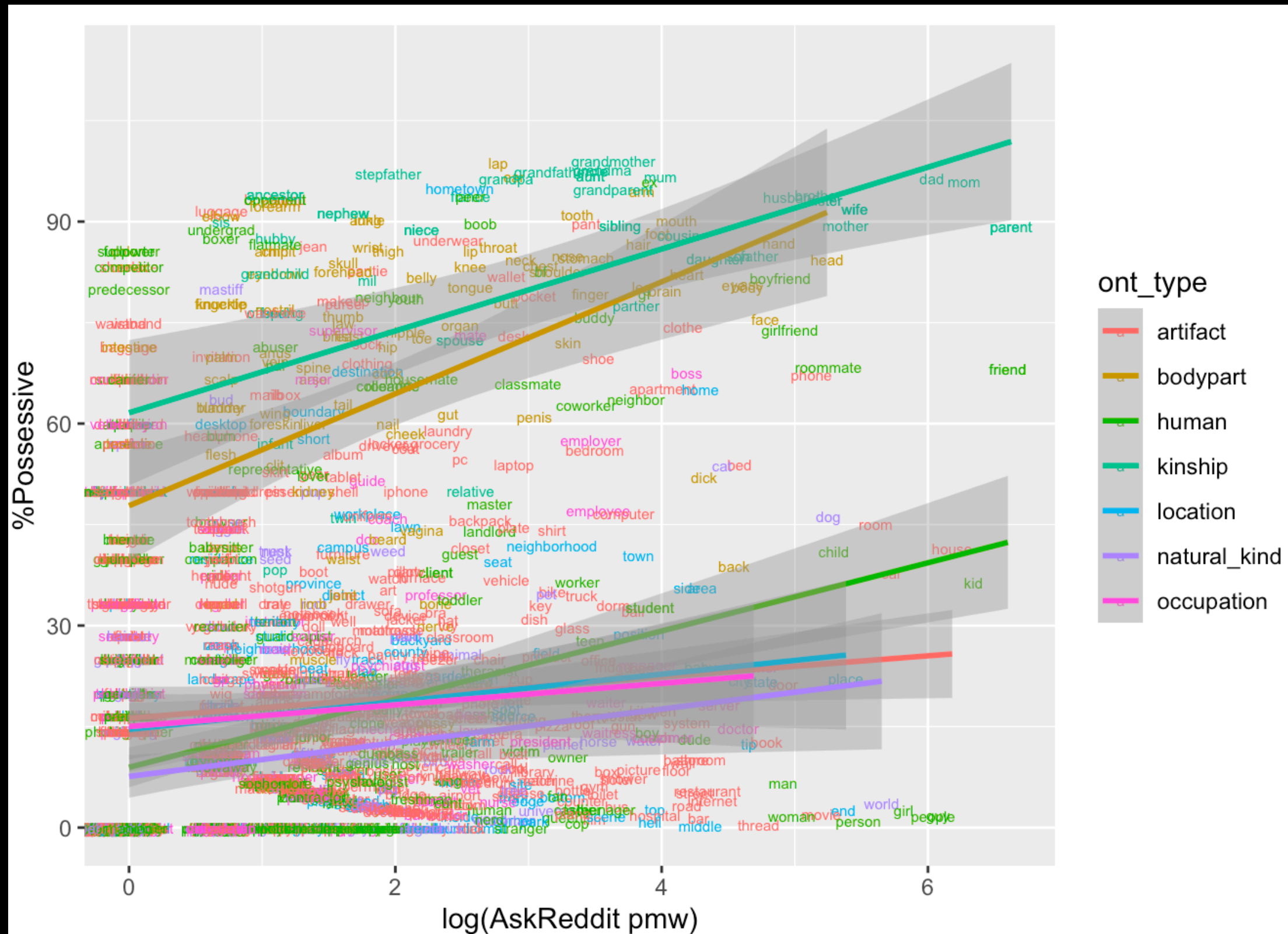
- A noun is *more relational* when human interaction with its referent is *more conventional*
- Predict:
 - More-%Possessive nouns...
 - *Proxy for: More relational*
 - ... should be more frequent
 - *Proxy for: More conventional interaction with their referent*
- TRUE across nouns, TRUE across communities!

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 <- lm(percent_poss ~ ont_class + pmw_count, 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 communities

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

	Possessive	Not Possessive
AskReddit	<p>N=4</p> <p><i>Go ahead and bring <u>your knife</u> to a gun fight</i></p>	<p>N=47</p> <p><i>yet another celeb who has gone under <u>the knife</u> to alter their appearance</i></p>
Cooking	<p>N=10</p> <p><i>Press the parsley stalks with the side of <u>your knife</u></i></p>	<p>N=15</p> <p><i>I never had [a peeler] before and usually did it with <u>a knife</u></i></p>

Freq. across communities

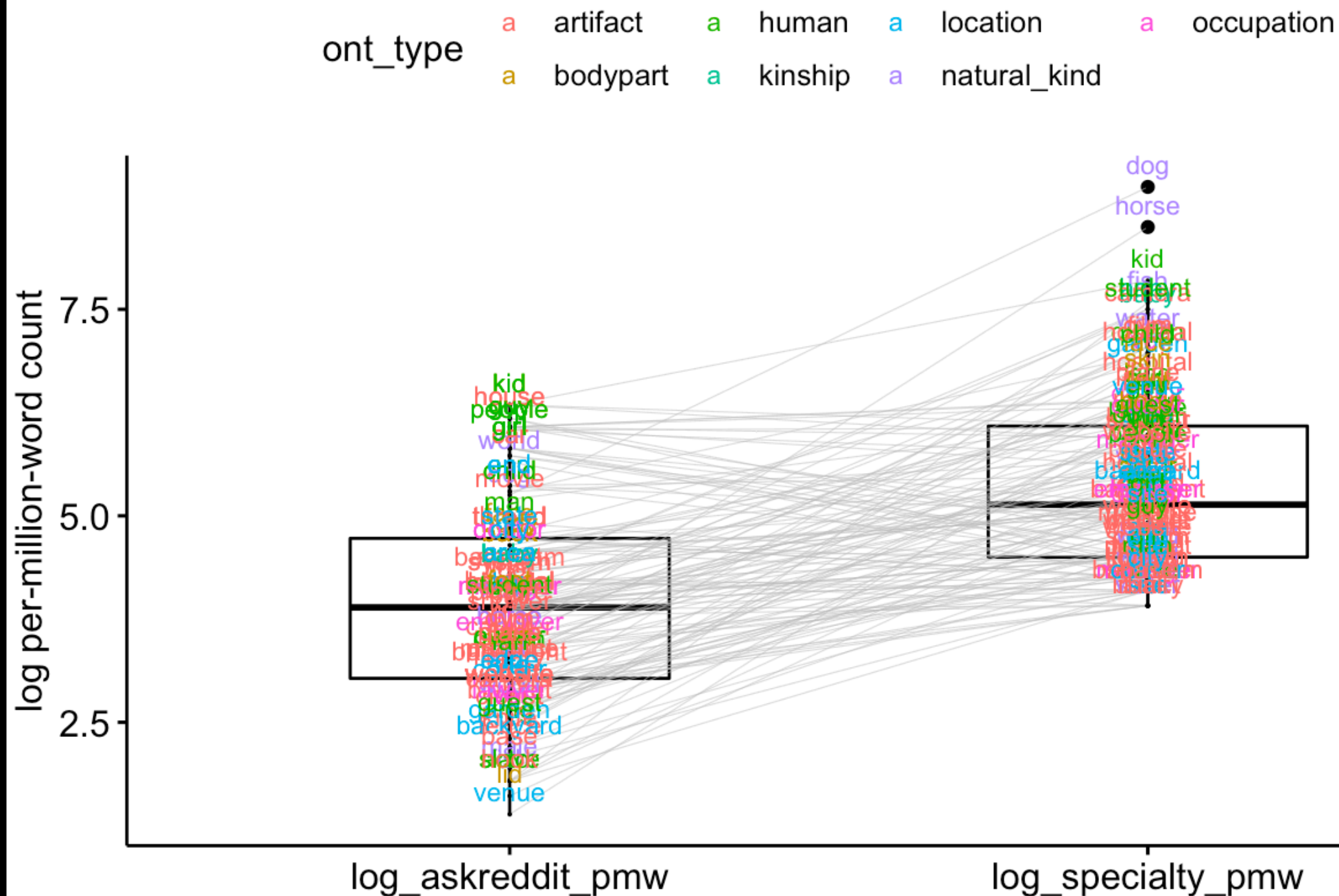
- Examples of nouns (136 total, 82 unique) used significantly more often as possessive in 32 specialist subreddits
 - r/Horses: blanket, horse, owner, vet....
 - r/Cooking: counter, kitchen, knife....
 - r/BabyBumps: baby, child, doctor, hospital, shower...

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

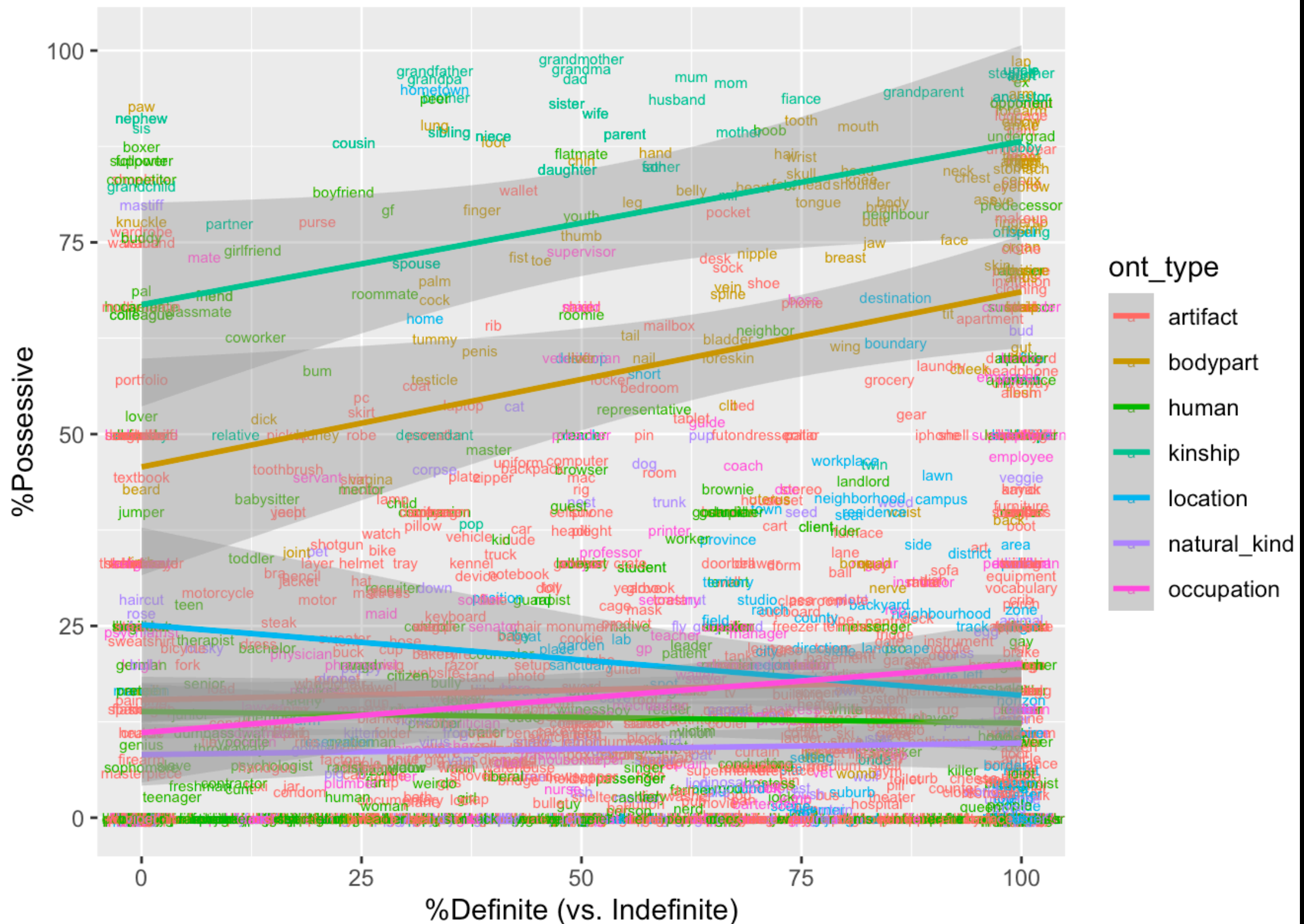


Definiteness ratio

- 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 communities

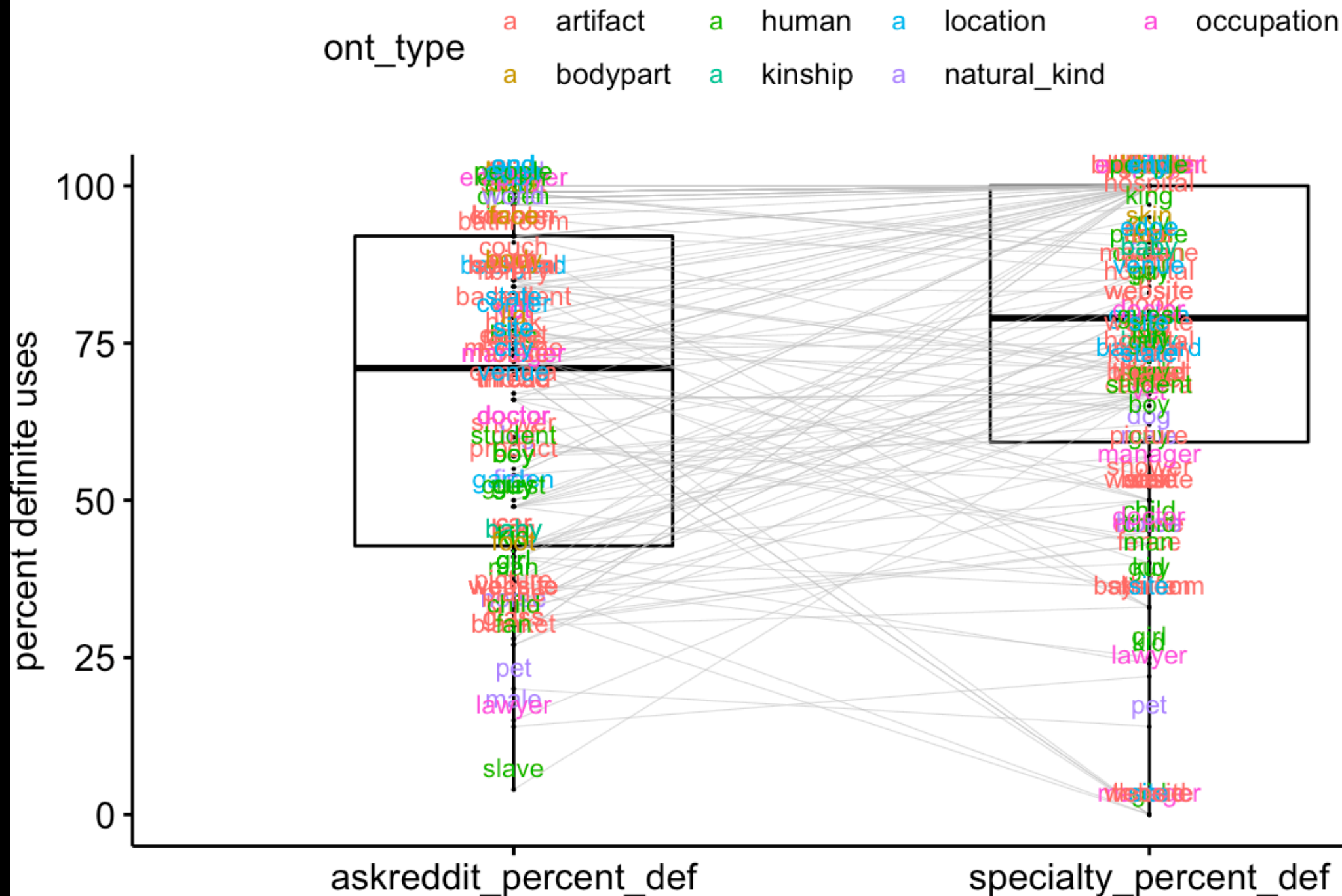
- Same 136 (82 unique) non-abstract nouns found in a Fisher Exact Test to be more often possessive in a specialty subreddit vs. AskReddit
 - r/Horses: blanket, horse, owner, vet....
 - r/Cooking: counter, kitchen, knife....
 - r/BabyBumps: baby, child, doctor, hospital, shower...

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



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 shouldnt 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?

Barker 1992, Partee & Borshev 1998, Payne et al 2013,
Peters & Westerståhl 2013, Kolkman 2016

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*)

Barker 1992, Partee & Borshev 1998, Payne et al 2013,
Peters & Westerståhl 2013, Kolkman 2016

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!

Barker 1992, Partee & Borshev 1998, Payne et al 2013,
Peters & Westerståhl 2013, Kolkman 2016

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

- Asmuth, Jennifer, & Gentner, Dedre. 2005. Context sensitivity of relational nouns.
- Aikhenvald, Alexandra Y. 2012. Possession and ownership: A cross-linguistic perspective.
- Barker, Chris. 1992. Possessive descriptions.
- Barker, Chris. 2000. Definite possessives and discourse novelty.
- Barker, Chris. 2011. Possessives and relational nouns.
- Clark, Herbert. 1975. Bridging.
- Clark, Herbert & Marshall, Catherine. 1981. Definite knowledge & mutual knowledge.
- Copestake, Ann, & Briscoe, Ted. 1995. Semi-productive polysemy & sense extension.
- De Bruin, Jos, & Scha, Remko. 1988. The interpretation of relational nouns.
- Del Tredici, Marco, & Fernandez, Raquel. 2018. Semantic variation in online communities of practice.
- Heine, Bernd. 1997. Possession.

Selected references

- Honnibal, Matthew, & Johnson, Mark. 2015. An improved non-monotonic transition system for dependency parsing.
- Karvovskaya, Lena. 2018. The typology and formal semantics of adnominal possession.
Kolkman, Julia. 2016. The pragmatics of possession: Issues in the interpretation of pre-nominal possessives in English.
- Levin, Beth, Glass, Lelia, & Jurafsky, Dan. 2019. Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds.
- Lichtenberk, Frantisek, Vaid, Jyotsna, & Chen, Hsin-Chin. 2011. On the interpretation of alienable vs. inalienable possession.
- Löbner, Sebastian. 2011. Concept types and determination.
- Meyers, Adam, Reeves, Ruth, Macleod, Catherine, Szekely, Rachel, Zielinska, Veronika, Young, Brian, & Grishman, Ralph. 2004. The NomBank project: An interim report.
- Miller, George, Beckwith, Richard, Fellbaum, Christiane, Gross, Derek, & Miller, Katherine. 1990. Introduction to WordNet: An online lexical database.

Selected references

- Mithun, Marianne. 1984. The evolution of noun incorporation.
- Nichols, Johanna. 1988. On alienable and inalienable possession.
- Partee, Barbara H., & Borschev, Vladimir. 1998. Integrating lexical and formal semantics: Genitives, relational nouns, and type-shifting.
- Payne, John, Pullum, Geoffrey K., Scholz, Barbara, & Berlage, Eva. 2013. Anaphoric 'ONE' and its implications.
- Peters, Stanley, & Westerståhl, Dag. 2013. The semantics of possessives.
- Pustejovsky, James. 1995. The generative lexicon.
- R Core Team. 2012. R: A Language and Environment for Statistical Computing.
- Roberts, Craige. 2003. Uniqueness in Definite Noun Phrases.
- Spenader, Jennifer. 2001. Presuppositions in spoken discourse.
- Vikner, Carl, & Jensen, Per Anker. 2002. A semantic analysis of the English genitive: Interaction of lexical and formal semantics.
- Williams, Adina. 2018. Representing Relationality,