Count/Mass Variation
Objects and the grammar of countability

Peter R. Sutton
Universitat Pompeu Fabra

Hana Filip
HHU Düsseldorf

Morphosyntax and Semantics Colloquium
Universität Potsdam
14. May 2024
`https://peter-sutton.github.io/talks`

Outline

Types of count/mass variation

- Focus: properties/concepts being lexicalized differently across and within languages

Empirical Landscape. A connection between:

- Variation in count/mass lexicalization patterns
- Non-canonical reflexes of countability

Semantics

- object-centred semantics for countability
- addition of structure to the lexicon

A system of constraints

- Why are there mass nouns with objects in their extensions?
- A corpus study to support the system of constraints

Count/Mass distinction in English

		Canonical Count Nouns	Canonical Mass Nouns
a.	Plural morphology	✓	✗
b.	Numeral constructions (no CL)	✓	✗
c.	Bare singular	✗	✓
d.	Distributive determiners	✓	✗
e.	Stubs	✓	✗

(1) a. Alex described the cats/#muds.
     b. three cats/#muds; one cat/#mud
     c. Mud/#Cat was on the floor.
     d. Every cat/#mud was hidden under the rug.
     e. big/small/round/square cat(s)/#mud

Three types of variation

1. Anti-universal variation?
   - One language has a grammatical count-mass distinction, but the other does not

2. Morphosyntactic variation
   - Two languages have a grammatical count-mass distinction, but the distinction is reflected differently in their grammars

3. Lexicalization pattern variation
   - Two languages have a grammatical count-mass distinction, but the mapping from concepts/properties to count/mass nouns is different

Anti-universal variation?

Languages without any countability distinctions?

- Mandatory classifier languages
  - All nouns are mass
    - Muromatsu 1995 for Japanese
    - Chierchia 1998 for Mandarin
  - Now more-or-less widely seen to be false (Doetjes, 1997)

- Yudja (Tupi)
  - All 'notional mass nouns' are count (Lima, 2014a,b)
  - Still an open question
  - Some doubt based on Deal 2017 (for a relevantly similar language Nez Perce, and also, with slightly weaker conclusions for Yudja)

Morphosyntactic variation

		Canonical Count Nouns	Canonical Mass Nouns
Plural morphology	Eng	✓	✗
	Fin	✓	✗
	Man	(N/a)	(N/a)
	Yud	(N/a)	(N/a)
	Grk	✓	✓
Numeral constrc (no CL)	Eng	✓	✗
	Fin	✓	✗
	Man	✗	✗
	Yud	✓	✓
Bare singular	Eng	✗	✓
	Fin	✓	✓
	Man	✓	✓
	Yud	✓	✓

Lexicalization pattern variation

	Count	Mass
Functionally combinatorial		
`furniture`	*huonekalu-t* (Fi); *meubel-s* (Dutch);	*furniture* (En); *meubilaire* (Dutch)
`jewellery`	*koru-t* (Fi); *joya-s* (Spa)	*jewellery* (En); *Schmuck* (Ger)
Granular		
`lentil`	*lentil-s* (En); *linssi-t* (Fin)	*čočka* (Cz); *lešta* (Bul)
`bean`	*bean-s* (En); *papu/-vut* (Fin)	*fasole* (Rom); *bob* (Bul)
Interconnected		
`fence`	*fence-s* (En); *plot-y* (Cz)	*fencing* (En); *oplocení* (Cz)
`shrub`	*shrub-s* (En); *Strauch/-äucher* (Ger)	*shrubbery* (En); *Strauchwerk* (Ger)

Overview: Two claims

Empirical claim: A strong correlation between:

- properties (concepts) that display variation in their count/mass lexicalization patterns and
- the properties underpinning nouns that have **non-canonical** grammatical reflexes of countability

Theoretical claim: The centrality of objects

- A broad notion of object that includes e.g., grains of sand and jigsaw puzzles.
- objects as a necessary condition for count lexicalization (at least for concrete properties)
  - This in turn explains the observes non-canonical reflexes of mass nouns

Overview: impact for lexical semantics

Context sensitivity (skated over today):

- At least some count nouns have context-dependent individuation conditions (Rothstein, 2010)

Structure in the lexicon:

- Common nouns specify their truth-conditions, and
- their individuation/counting criteria
  - Needed to be able to distinguish co-intensional mass-plural count pairs (e.g., *meubels-meubilaire*)
- bi-partite lexical entries

Overview: problems and challenges

Constraint problem:

- Only some properties that have objects in their extensions display variation in their count/mass lexicalization patterns
  - e.g., cat vs. `lentil` and `furniture`

Exception problems:

- Why do some mass nouns that have non-canonical reflexes of countability never have count counterparts crosslinguistically: *dust*, *pollen*, *sand*.

- More generally, why are there large differences in the probabilities of some properties being lexicalized as count (mass) compared to others

Overview: Proposal

A system of constraints. For a property P

- Indistinguishability: the objects in the extension of P are, perceptually speaking, indistinguishable relative to average human perceptual acuity.

- Collective uses of Instruments: At least in many cases, objects in the extension of P are used together as instruments in eventualities typically associated with P

- Object splitting: Objects in the extension of P are often used in such a way that requires them to (first) be split apart or broken up.

If P has objects in its extension and satisfies at least one of the constraints, P can be lexicalized as mass.

Corpus study: evidence for the constraints

- Accounting for tendency of properties to be lexicalized as count or mass based on the system of constraints

Object mass nouns (OMNs)

ammunition, apparel, armor, art, artillery, artwork, autumnwear, baggage, bakeware, beachwear, bedding, change, china, clothing, clutter, coinage, crockery, cutlery, decoration, dishware, equipment, earthenware, freight, furniture, footwear, gear, glassware, hardware, inventory, jewelry, knitwear, ladieswear, laundry, legwear, lingerie, loot, luggage, mail, menswear, merchandise, [...], outerwear, packaging, paperwork, plasticware, rigging, seating, shapewear, silver, silverware, software, sportswear, [...], stock, swag, tackle, teaware, tupperware, underwear, weaponry (Erbach, 2021, p. 201)

OMNs as a focus of count/mass semantic accounts:

- A notional/grammatical mismatch: denote countable objects, but are grammatically mass
- This shows up in their grammatical reflexes

OMNs: Cardinality comparison readings
(Barner and Snedeker, 2005)

(Who has more silverware?)   (Who has more shoes?)   (Who has more toothpaste?)
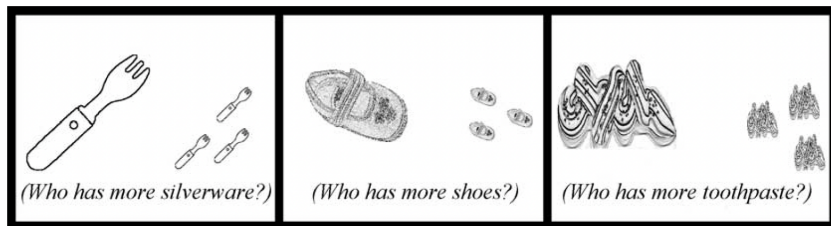
Object mass nouns have cardinality comparison readings, but canonical mass nouns do not

- Also have measure readings (e.g., Rothstein, 2017)

OMNs: Felicitous when modified by Stubs

Stubbornly distributive predicates (Stubs) (Schwarzschild, 2011; Rothstein, 2010)

- *big*, *small*, *round*, *square*

(2)   Alex moved the round/small tables/furniture/#oil.
    ⇒ Each of the tables/pieces of furniture are round/small

OMNs: count/mass variation

	Count	Mass
`furniture`	*huonekalut* (Fi); *meubels* (Dutch);	*furniture* (En); *meubilaire* (Dutch)
`jewellery`	*korut* (Fi); *joyas* (Spa)	*jewellery* (En); *Schmuck* (Ger)
`cutlery`	*ruokailuvälineet* (Fi);	*cutlery* (En); *príbor* (Cz)
`kitchenware`	*Küchengeräte* (Ger); *keittiövalinnet* (Fi)	*kitchenware* (En); *nádobí* (Cz)

Functionally combinatorial nouns

- A cover term for OMNs and their count counterparts

Beyond OMNs

Arguably too much attention

- Present in Germanic, Romance and Slavic languages
- Very rare/absent in others (Greek, Finnish)

Much focus on cardinality comparison readings

- Led to less attention on other nouns

Granular and Filament nouns

Variation

- Granular and filament mass nouns widely attested in many languages
- Frequently lexically simply (cf object mass nouns)

	Count	Mass
`lentil`	*lentil-s* (En);	*čočka* (Cz)
	linssi-t (Fin)	*lešta* (Bul)
`bean`	*bean-s* (En);	*fasole* (Rom)
	papu/pavut (Fin)	*bob* (Bul)
`cabbage`	*cabbage-s* (En);	*cabbage* (En)
	kaali-t (Fin)	*Kohl* (Ger)
`asparagus`	*asperge* (Fr);	*asparagus* (Rom)
	?*Spargel* (Ger)	?*Spargel* (Ger)
`bamboo`	*Bambusrohr-e* (Ger);	*bamboo* (En)

But some limits on variation

- E.g., `dust`, `pollen` always lexicalized as mass

Granular and Filament nouns: Stubs

Observation: GF mass nouns are felicitous with Stubs, just like OMNs

(3) Dieser sehr feine **runde Reis** kann viel Flüssigkeit aufnehmen ohne aufzuweichen.

(4) eine Lage **langes Stroh oder Heu** mit Lehm beschmiert

Even with nouns like *Sand*, *Staub* and *Pollen*

(5) Dieser **Sand ist rund** und enthält kaum Kalk,...

(6) Der soll auch den **kleinsten Staub** mit aufnehmen und wegfeudeln beim Putzen.

(7) Darunter ist der **kleine Pollen** des Vergißmeinnichts ein echter Marker Berliner Honige
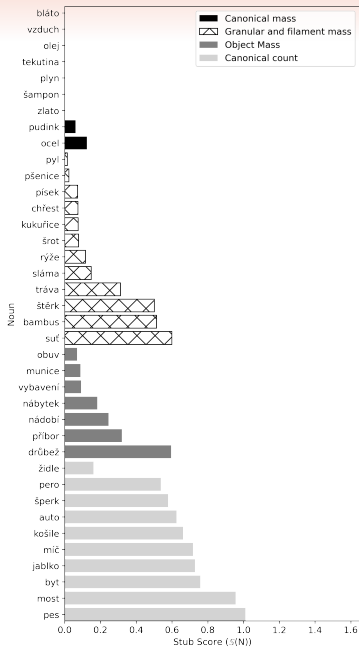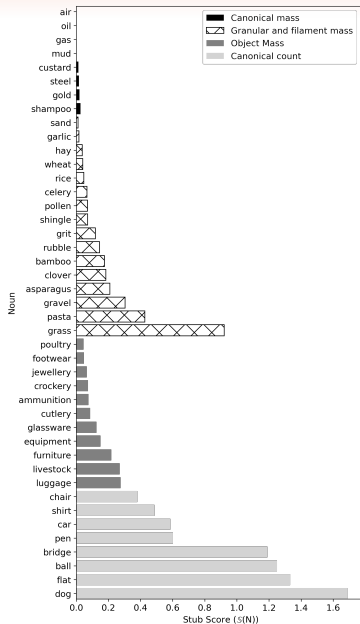
Granular and Filament nouns: Cardinality comparison readings?

Mostly only under very heavy context-setting

- Landman (2021, (Sutton p.c.)): a rice grain hunting competition in which contestants must find as much rice as possible in an allotted time. The intuition is that the winner would have the most grains, regardless of whether they found the most by weight or volume.)

But at least one case: *pollen*

- supposing that last month, most of the pollen in the air was small grass pollen, and this month, larger tree pollen, if the numbers of pollen grains in the air are the same, it is not clear to us that 'There is more pollen in the air this month' is true.

Granular and Filament nouns in the literature

Often treated as in some sense mass by default e.g., Chierchia 2010; Landman 2020

- Chierchia:
    - stably atomic properties have a set of atoms shared across all contexts. E.g., what counts as a minimal chair (a chair atom) is stable
    - granular properties are not stably atomic: e.g., what counts as the smallest *rice* entities varies with context.
    - A standardized partition operator needed to account for count granulars
- Landman
    - 'Neat mass' nouns have a disjoint set of atoms in their extensions
    - 'Mess mass' nouns do not
    - Granular mass nouns are mess mass (like canonical mass nouns)

Interconnected nouns: Variation

Mass nouns commonly derived morphologically from count nouns

	Count	Mass
`fence`	*fence-s* (En); *plot-y* (Cz)	*fencing* (En); *oplocení* (Cz)
`shrub`	*shrub-s* (En); *Strauch/-äucher* (Ger)	*shrubbery* (En); *Strauchwerk* (Ger)
`wall`	*wall-s* (En); *zed'/-i* (Cz)	*walling* (En); *zdivo* (Cz)

Does not always result in a mass noun:

(8) Einige AktivistInnen nutzten den Raum zwischen zwei
    several activists   use   the space between two
    Umzäunungen zu einem "Atomwaffenfreien Picknick".
    fencing.PL   to an   atomic.weapon.free picnic
    "Several activists use the space between two (boundary) fences
    for an 'atomic weapon free picnic'."

Interconnected mass nouns: stubs

A little rare, but some instances attested:

(9)   a.   They sell large fencing, small fencing - any size.
                [enTenTen21]
         b.   For a nicer and cleaner finish, add some small fencing
                around your garden [enTenTen21]

(10)   For smaller hedging or trimmed shrubs, bays and rosemary
        are both extremely hardy, water savvy and aromatic.

Interconnected count nouns: pseudopartitives
(Filip and Sutton, 2017)

Count nouns in English cannot be used in the bare singular

- Including as the 'downstairs' NP in a pseudopartive construction

(11)   a. #6 kilograms of baby
       b. #You can find a heavy piece of baby in the nursery.

But interconnected count nouns can:

(12)   Thick woolen drapes of red and gold covered every inch of wall. (COCA)

(13)   Thus a cm dry length of twig increased in dry weight by 0.047g. (*Community Ecology of a Coral Cay*, Heatwole et al. p.152)

(14)   The cages were 1 foot in diameter and enclosed a 3-foot length of branch. (*California Agriculture*. Mar-Apr, 1989 p.7)

(15)   155 kilometers, or 96 miles, of wall encircled West Berlin (CNN "Berlin wall secrets")

	Count diagnostic environments			Mass diagnostic environments	
	NNCs	Card. comp.	Stubs	SG NP Meas	Bare SG
Mass Ns					
Canonical	✗	✗	✗	✓	✓
Interconnected	✗	✗	✓	✓	✓
Granular & Filament	✗	✗/✓	✓	✓	✓
Func. Combinatorial (object mass)	✗	✓	✓	✓	✓
Count Ns					
Interconnnected	✓	✓	✓	✓	✗
Func. Combinatorial	✓	✓	✓	✗	✗
Granular & Filament	✓	✓	✓	✗	✗
Canonical	✓	✓	✓	✗	✗

Hypotheses and observations

Observations:

- When the properties that display variation in their count/mass lexicalization patterns are lexcalized as mass, these mass nouns are felicitous with stubs.

- In general variation implies some kind of non-canonical grammatical reflexes of countability

Hypotheses:

- For concrete properties, only properties with objects in their extensions can be lexicalized as count.

- If such properties are lexicalized as mass, these objects can be accessed by the grammar (e.g., the semantics of stubs)

Corrolaries

Variation: We should not be asking why some properties are lexicalized as count and some as mass.

- E.g., Why can `furniture` be lexicalized as mass? Why can `lentil` be lexicalized as count?
- Leads to positing different mechanisms in the lexicon to derive countable predicates from supposedly mass-like properties and non-countable predicates from count-like properties

The unifying question: Why can properties with objects in their extensions be lexicalized as mass nouns at all?

- Default: If there are objects, these can be grammatically counted
- This can be overidden
- Need to develop a system of constraints

Spelke Objects

Spelke objects

- "bodies that are cohesive, bounded, spatiotemporally continuous, and solid or substantial; they move as connected wholes, independently of one another, on connected paths through unoccupied space" (Soja et al., 1991, p. 183).

Too narrow

- Entities that are not cohesive or bounded
- Entities that are connected to others but are different objects

Other kinds of objects

Non-cohesive objects

- multi-mixers, potato head toys, jigsaw puzzles

Connected objects

- fences, rooms in a house, houses in a terrace

Spelke Objects

Entities that fulfil functional roles

Non-cohesive objects multi-part toys, mixers etc.

apples, cats, grains of rice, strands of hair etc.

chairs, hammers etc.

Connected objects rooms, flats, storeys etc.
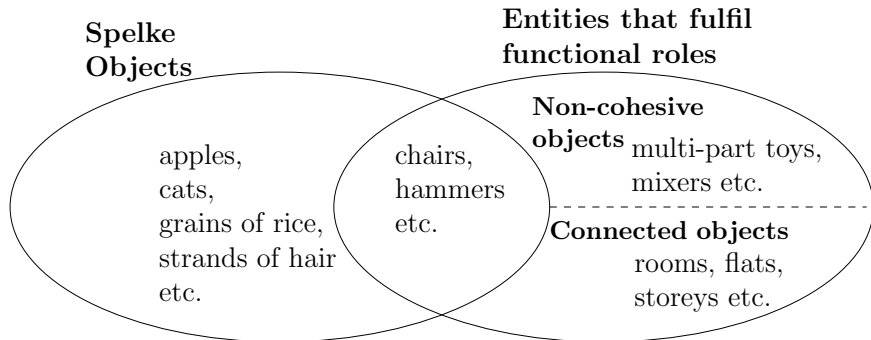
Objects do not form a non-quantized set

- Quantized: At least one entity in the set is a proper part of another $\{a, a \sqcup b\}$
- $QUA(P) \leftrightarrow \forall x, y[(P(x) \wedge P(y)) \rightarrow \neg x \sqsubset y]$
- The set of objects is not quantized
    - jigsaw puzzles and pieces
    - multimixers and parts
    - fences and certain parts of fences

Object-centred contextualist semantics

Contextual domain restriction

- Nouns are context sensitive, and their domains are routinely restricted in DPs

- Not part of the semantics of quantifiers! (Stanley, 2002)

  (16)   The tallest person is nice.

    - Can mean e.g., the tallest person in the room is nice

- If CDR encoded by the determiner, we cannot derive this reading

- 'tallest person' would always pick out the tallest person in the world

First pass: Kaplanian semantics in TY_3

(17) $\llbracket \text{person} \rrbracket = \lambda c.\lambda w.\lambda x.\mathfrak{f}_c(\text{person}_{Obj})(w)(x)$

    a. $\text{person}_{obj}(w)$: The intersection of the number neutral set of people with the set of objects in w

    b.   contexts: $\langle c_{\text{utt}}, c_{\text{time}}, c_{\text{loc}}, c_{\text{utt}}, c_{\text{dom}} \rangle$

    c. $c_{dom} := \lambda P_{\langle s,et \rangle}.\lambda w.\lambda x.\text{P}(w)(x) \land \text{Q}(w)(x)$

    d.   There is a $w' \in \mathcal{D}_s$ where $\llbracket \text{P}(w') \rrbracket \cap \llbracket \text{Q}(w') \rrbracket \neq \varnothing$ and Q is salient/relevant to P.

E.g., if in c_0 Q is in_the_room:

(18) $\llbracket \text{person} \rrbracket^{c_0} =$
       $\lambda w.\lambda x.\text{person}_{obj}(w)(x) \land \text{in\_the\_room}(w)(x)$

Problems & Challenges

- Demarcating count from mass nouns (with objects in their extensions)
- Why can stubs access the set of objects when e.g., numerals cannot

The cointensionality problem

(P1)   The count/mass distinction is reflected in the semantics of common nouns.

(P2)   The only relevant locus for a semantic countability distinction is the extension of a common noun at a world and in a context. E.g. the extension is generated from a quantized/disjoint/stably atomic set etc.

(P3)   There are plural count nouns and object mass nouns that are coextensional.

(C1)   There can be no coextensional plural count nouns and object mass nouns.   (P1, P2)

(C3)   ⊥   (P3, C1)

The cointensionality problem: Chierchia's solution

Deny (P3)

- Not: There are plural count nouns and object mass nouns that are coextensional.

- E.g., $[\![\text{furniture}]\!]^w$ = The singleton set containing only the sum of all furniture items in w.

The cointensionality problem: Our solution

Deny (P2)

- Not: The only relevant locus for a semantic countability distinction is the extension of a common noun at a world and in a context.
- Landman 2011, 2016; Sutton and Filip 2016, 2017
- Bi-partitite lexical entries ⟨Extension, Counting base⟩

Second pass:

(19) ⟦huonekalu-t⟧ $=$
$\lambda c.\lambda w.\lambda x.\langle {}^{*}\mathfrak{f}_{c}(\texttt{furniture}_{obj})(w)(x),\ \lambda y.\texttt{furniture}_{obj}(w)(y)\rangle$

    a.   Extension: The set of furniture objects closed under sum

    b.   Counting base: The set of single furniture objects (Quantized)

(20) ⟦furniture⟧ $=$
$\lambda c.\lambda w.\lambda x.\langle \mathfrak{f}_{c}(\texttt{furniture})(w)(x),\ \lambda y.\texttt{furniture}(w)(y)\rangle$

    a.   Extension: The number neutral set of furniture entities

    b.   Counting base: The number neutral set of furniture entities (Not Quantized)

Mappings from properties

(21) $\mathrm{P} \mapsto \lambda c.\lambda w.\lambda x.\langle^* \mathfrak{f}_c(\mathrm{P}_{obj})(w)(x), \ \lambda y.\mathrm{P}_{obj}(w)(y)\rangle$

- Object-centred lexicalization
- Result: Count noun

(22) $\mathrm{P} \mapsto \lambda c.\lambda w.\lambda x.\langle \mathfrak{f}_c(\mathrm{P})(w)(x), \ \lambda y.\mathrm{P}(w)(y)\rangle$

- Object-neutral lexicalization
- Result: Mass noun

One constraint already:

- Object-centred lexicalization only available for P if P_{obj} is non-empty

Stubs

(23) $[\![$ round $]\!]^{M,g} = \lambda\mathfrak{P}_{:\mathfrak{P}(w)\cap\mathbb{O}(w)\neq\varnothing}\lambda c\lambda w\lambda x.$
$\langle\pi_1(\mathfrak{P}(c)(w)(x)) \wedge \forall z[z \sqsubseteq x \wedge \pi_1(\mathfrak{P}(c)(w)(z) \wedge \mathbb{O}(w)(z)) \rightarrow Round(z)],$
$\pi_2(\mathfrak{P}(c)(w)(x))\rangle$

(24) $[\![$ round rice $]\!]^{M,g} = \lambda c\lambda w\lambda x.$
$\langle\mathfrak{f}_c(\mathtt{rice})(w)(x) \wedge \forall z[z \sqsubseteq x \wedge \mathtt{rice}(c)(w)(z) \wedge \mathbb{O}(w)(z)) \rightarrow Round(z)],$
$\lambda y.\mathtt{rice}(w)(y)\rangle$

The right result

- A mass NP ($\mathtt{rice}(w)$ is not a quantized set)
- The objects in the extension (i.e, the grains) must be round
- e.g., *round mud* would be infelicitous

Indistinguishability

(C1) **Indistinguishability:** For a property P, objects in the
extension of P are, perceptually speaking, indistinguishable
relative to average human perceptual acuity. By this we mean
that they are too small, alike in their perceptual properties or
are clustered together in such a way that make them hard for
us to track as individuals.

Properties affected include:

- dust, pollen, rice, lentil, bean

Graded:

- pollen > rice > potato

Collective uses of Instruments

(C2) **Collective uses of Instruments:** There is a typical
associated eventuality with a property P such that in many
cases, there is a least upper bound for the fulfilment of
INSTRUMENT role in this eventuality, namely (usually
heterogeneous) sums of objects in the extension of P. I.e.,
any proper part of these sums of instruments would not
normally facilitate bringing that eventuality about.

kitchenware: One of the most typical eventualities associated
with kitchenware is *preparing a meal*.:

(25) $\{\langle e, x, y \rangle | \texttt{prepare}_w(e) \wedge \textsc{theme}(e, x) \wedge \texttt{meal}_w(x)$
             $\wedge \textsc{instr}(e, y) \wedge \texttt{kitchenware}_w(y)\}$

- E.g., a knife, a cutting board, a pan, a wooden spoon etc. are
  used collectively as the instruments for making a sauce

A case where (C2) is not satisfied: `chair`

The typical eventuality associated with `chair` is to seat someone (chairs facilitate seating people):

(26) $\{\langle e, x, y \rangle \,|\, \texttt{seat}_w(e) \wedge \text{THEME}(e, x) \wedge \texttt{person}_w(x)$
            $\wedge \text{INSTR}(e, y) \wedge \texttt{chair}_w(y)\}$

The (C2) constraint does not apply to `chair` since a single chair typically counts as a least upper bound for fulfilling the instrument role in seating someone.

Property	Assoc.Eventuality	C2 applies
`kitchenware`	preparing a meal	Yes. E.g. knife and pan etc. needed for making a sauce
`jewellery`	adorning different parts of a body	Yes. E.g., earrings and necklace needed in some contexts
`furniture`	furnishing some space	Yes. E.g., bed, closet etc. needed for a furnished bedroom
`chair`	seating someone	No. Single chairs are sufficient to seat someone

Table: Examples of Constraint (C2) and whether it applies.

C2 and interconnected nouns

Count and mass interconnected nouns

- *fence*, *hedge*
- *fencing*, *hedging*

(27)   The property backs onto adjoining farmland and is
         enclosed by fencing and hedging. [ukWaC]

Typical eventuality

- Enclosing and/or partitioning spaces
- They involve, as instruments, multiple objects in the extension
  of fence/hedge.

Interconnected count nouns and psudopartitives

(28)   three hundred metres of fence/wall/hedge/road

Quantization and pseudopartitives (Krifka, 1989)

- The 'downstairs' NP cannot be quantized
- At least one entity in the set is a proper part of another $\{a, a \sqcup b\}$
- $QUA(P) \leftrightarrow \forall x, y[(P(x) \wedge P(y)) \rightarrow \neg x \sqsubset y]$

C2 and quantization

- It is partly because the fencing around, say, a property can count as both one fence or as more than one that allows for properties such as fence to satisfy (C2)
- Another way of articulating the observation that the set of objects in the extension of fence fails to be quantized
- Motivated why singular count interconnected nouns are felicitous in pseudopartitives

Object splitting

(C3) **Object splitting:** Objects in the extension of P are often used in such a way that requires them to (first) be split apart or broken up.

Properties affected include:

- $\texttt{melon}, \texttt{carpet}, \texttt{potato}, \texttt{cabbage}, \texttt{apple}$

Graded:

- $\texttt{cabbage} > \texttt{potato} > \texttt{apple}$

Food nouns and coercion

Most count nouns can be coerced if used in mass syntactic environments, but not all such environments are equal (Fillmore and Kay, 1987)

(29)   a.   There is apple/pear in the fruit salad.
       b.   ?We ate apple/pear for dessert.

Nouns such as *melon* and *mango* predominantly have count uses in English, they are also natural when used bare in the direct object position.

(30)   a.   There is melon/mango in the salad.
       b.   We ate melon/mango for dessert.

Suggests that this is not just systematic coercion

Corpus model

Basic idea:

- Estimate the degree to which concepts satisfy *indistinguishability* and *object splitting*
- The way we conceive of properties may be reflected in the uses of the nouns that lexicalize them

Limitations:

- A single mono-lingual corpus (enTenTen 20, >43 billion tokens)

Lemmas tested

(31)   Set of test lemmas:
        apple | ball | bean | berry | cabbage | car | chair | dust |
        equipment | furniture | grape | gravel | jewelry |
        kitchenware | lentil | sand | pebble | pollen | potato | rice |
        seed

Of these, on a rational basis, we assigned (C2) satisfaction scores
of 1 to equipment, furniture, jewelry and kitchenware, and 0 to all
other lemmas.

C1 and C3

(32)   Groups of (C1) indicating lemmas

    a.   heap | mound | pile

    b.   cluster | scatter | scattering | shower | sprinkle |
        sprinkling | smatter | smattering

    c.   cloud | dusting | mist

    d.   homogenous | indistinguishable | uniform

    e.   flake | granule | particle

(33)   Groups of (C3) indicating lemmas

    a.   chop | cut | dice | slice

    b.   break | grate | grind | mince | powder

    c.   divide | halve | quarter

    d.   fragment | sliver | wedge

Scoring model

- $Pr(M|N)$: probability of a modifier given a noun
- $\mathbb{E}(Pr(G|N))$: entropy score for the diversity of modifiers by group
  - E.g., if N is only used with *chop, cut, slice* and never with e.g., *break, halve, fragment*, $\mathbb{E}(Pr(G|N))$ is low

Fr is a frequency function. $Mod_s(M, N)$ means *M* modifies *N* in a sentence in the sample. $Co_s(M, N)$ means *M* co-occurs with *N* in a sentence in the sample:

(34)   Probability of *M* modifying *N*, given *M* co-occurs with *N* in the sample:

$$Pr_s(Mod_s(M, N)|Co_s(M, N)) = \frac{Fr(Mod_s(M, N))}{sample\ size}$$

(35)   Estimation of frequency of any *M* modifying *N* in the corpus. The co-occurrence of *M* with *N* in the whole corpus, factored by the probability that *M* modifies *N* in the sample:

$$Fr(Mod_c(M, N)) \approx$$
$$Fr(Co_s(M, N)) \times Pr_s(Mod_s(M, N)|Co_s(M, N))$$

(36)   Estimated probability of *M* modifying *N*, given *N* in the corpus. The estimated frequency of any *M* modifying *N* in the corpus as a proportion of total occurrences of *N*:

$$Pr_c(M|N) \approx \frac{Fr(Mod_c(M, N))}{Fr(N)}$$

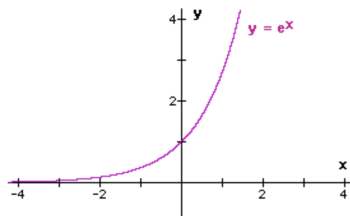(37)   Probability of a group G, given M modifies N in the sample:

$$Pr_s(G|Mod_s(M,N)) = \frac{Fr(Mod_s(G,N))}{Fr(Mod_s(M,N))}$$

(38)   Entropy of $Pr_s(G|Mod_s(M,S))$:

$$\mathbb{E}(Pr_s(G|Mod_s(M,S))) =$$
$$\sum_{G \in \mathcal{G}} Pr_s(G|Mod_s(M,S)) \times Ln(Pr_s(G|Mod_s(M,S)))$$

(39) $\mathbb{S}_{C1,C3}(N) =$
   $1 - exp(-Pr_c(M_C|N) \times \mathbb{E}(Pr_s(G_C|Mod_s(M_C, N)))) \times 100)$
   Such that $\mathbb{S}_{C1,C3}(N) \in [0, 1]$

(40) $\mathbb{S}_{C2}(N) \in \{0, 1\}$

The weighted mean model is then a simple weighted sum of the scores for all constraints. Where weights ω_{C1}, ω_{C2}, ω_{C3} sum to 1:

(41) $\mathbb{S}(N) = \sum_{Cn \in \{C1, C2, C3\}} \omega_{Cn} \times \mathbb{S}_{Cn}(N)$

(42)   Success criteria for the model
    a.   Scores for dust, pollen and sand should be higher
       than all others.
    b.   Scores for car, chair, and ball should be lower
       than all others.
    c.   Scores for properties underpinning collective artefact
       nouns (e.g., furniture) should be low (but above
       those of car, chair, and ball).

(43)   Null Hypothesis: There are no values for ω_{C1}–ω_{C3} that
    return scores that fulfil the success criteria.

(44)   Ordering hypothesis: The ordering of ω_{C1}–ω_{C3} values that
    return successful outcomes according to (42) will be:
    $\omega_{C1} > \omega_{C3} > \omega_{C2}$

Results: Scores

Property	$\mathbb{S}_{C1}$	Property	$\mathbb{S}_{C2}$	Property	$\mathbb{S}_{C3}$
dust	0.997	equipment	1	potato	0.874
sand	0.773	furniture	1	cabbage	0.787
pollen	0.666	jewelry	1	bean	0.658
gravel	0.459	kitchenware	1	apple	0.491
pebble	0.403	apple	0	berry	0.220
rice	0.278	ball	0	pebble	0.128
berry	0.228	bean	0	grape	0.115
seed	0.208	berry	0	pollen	0.066
potato	0.152	car	0	ball	0.049
cabbage	0.096	cabbage	0	kitchenware	0.024
bean	0.087	chair	0	chair	0.022
lentil	0.074	dust	0	sand	0.018
grape	0.071	grape	0	car	0.016
apple	0.065	gravel	0	lentil	0.000
furniture	0.063	lentil	0	jewelry	0.000
ball	0.030	pebble	0	gravel	0.000
kitchenware	0.017	pollen	0	dust	0.000
chair	0.016	potato	0	equipment	0.000
jewelry	0.011	rice	0	rice	0.000
equipment	0.010	sand	0	furniture	0.000
car	0.006	seed	0	seed	0.000

Results: Runs of the model

Ordering of scores with equally weighted constraints: $\omega_{C1}, \omega_{C2}, \omega_{C3} = 0.33$

```
furniture > kitchenware > potato > jewelry >
equipment > dust > cabbage > sand > bean > pollen >
apple > pebble > gravel > berry > rice > seed > grape >
ball > lentil > chair > car
```
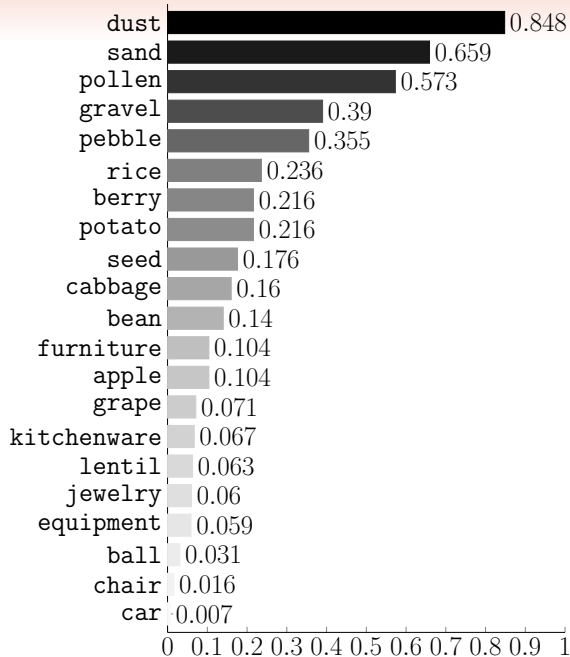
Ordering of scores weighted where C1 has twice the weight of C3 and C3 has twice the weight of C2: $\omega_{C1} = 0.57$, $\omega_{C2} = 0.14$, $\omega_{C3} = 0.29$:

```
dust > sand > pollen > potato > cabbage > pebble >
gravel > bean > berry > apple > furniture > rice >
kitchenware > jewelry > equipment > seed > grape >
lentil > ball > chair > car
```

Results: most successful run

Ordering of scores weighted where C1 has 8-9 the weight of C3 and C3 has twice the weight of C2. $\omega_{C1} = 0.85$, $\omega_{C2} = 0.05$, $\omega_{C3} = 0.1$:

dust $>$ sand $>$ pollen $>$ gravel $>$ pebble $>$ rice $>$ berry $>$ potato $>$ seed $>$ cabbage $>$ bean $>$ furniture $>$ apple $>$ grape $>$ kitchenware $>$ lentil $>$ jewelry $>$ equipment $>$ ball $>$ chair $>$ car

dust	0.848
sand	0.659
pollen	0.573
gravel	0.39
pebble	0.355
rice	0.236
berry	0.216
potato	0.216
seed	0.176
cabbage	0.16
bean	0.14
furniture	0.104
apple	0.104
grape	0.071
kitchenware	0.067
lentil	0.063
jewelry	0.06
equipment	0.059
ball	0.031
chair	0.016
car	0.007

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Discussion

- empirical means of approximating the extent to which properties underpinning nouns satisfy certain conceptual constraints
- confirmed our hypothesis that the strongest predictor of mass lexicalization are perceptual properties of entities in the extensions of the relevant properties
- We were surprised regarding how dominant this effect is

Follow-ups

- Replication with other English corpora
- Replication with corpora from other languages
- A large scale crosslinguistic study to estimate the actual likelihood of object extensional properties being lexicalized as mass across languages
- collective uses of instruments scores on a rational and categorical basis is non-optimal.

Order in chaos?

Variation in count/mass lexicalization patterns is messy

- Taken to be evidence that countability is purely grammatical (Rothstein, 2010)

But variation is not random

- Importance of objects
- Restriction to certain kinds of objects
  - perceptual properties
  - how we interact with those objects

Lexical entries

The Co-intensionality Challenge

- Does this mean that the count/mass distinction is not semantically encoded in the lexicon? (Pelletier, 1975; Borer, 2005)

- But, again, variation is not random

- We do need to add structure to the lexicon (bi-partite lexical entries)

Extensions: Non-number marking languages

- Extensions to Mandarin and Yudja
- Bi-partite lexical entries
- Extensions as kinds (Chierchia, 2015)

Extensions: Abstract nouns

- Objects as anchors for counting (Grimm, 2014)
- Propositions have the grammatical reflexes of objects
- But the set of propositions is radically non-quantized
  - Predicts context sensitivity in counting

(45) Vielleicht **ist eine Information** für unsere treuen Besucher noch
interessant: Alle Ensemblemitglieder sind davon úberzeugt, sich
persönlich von Jahr zu Jahr gesteigert zu haben und meinen, dennoch
genug Potenziale zu verspüren, die noch gehoben werden können.

(46) Vielleicht **sind zwei Informationen** für unsere treuen Besucher noch
interessant: Alle Ensemblemitglieder sind davon úberzeugt, sich
persönlich von Jahr zu Jahr gesteigert zu haben und meinen, dennoch
genug Potenziale zu verspüren, die noch gehoben werden können.

One piece/two pieces of information *is/are perhaps of interest to our
loyal visitors: All cast members are satisfied that they have improved year
on year, but nevertheless think that they still have enough potential for
enhancement.*

Thanks

Thank you for listening!

Many thanks to Frank Grüneisen and Nina Haslinger for providing German judgements and to Markus Hippi for assistance with Finnish judgements and translations.

References I

Barner, D. and J. Snedeker (2005). Quantity judgments and individuation: evidence that mass nouns count. *Cognition 97*, 41–66.

Borer, H. (2005). *Structuring Sense I: In Name Only*. OUP.

Chierchia, G. (1998). Plurality of mass nouns and the notion of "semantic parameter". In *Events and Grammar: Studies in Linguistics and Philosophy Vol. 7*, pp. 53–103. Kluwer.

Chierchia, G. (2010). Mass Nouns, Vagueness and Semantic Variation. *Synthese 174*, 99–149.

Chierchia, G. (2015). How universal is the mass/count distinction? three grammars of counting. In A. Li, A. Simpson, and W.-T. D. Tsai (Eds.), *Chinese syntax: A cross-linguistic perspective*, pp. 147–177. OUP.

Deal, A. (2017). Countability distinctions and semantic variation. *Natural Language Semantics 25*(2), 125–171.

Doetjes, J. (1997). *Quantifiers and Selection: on the distribution of quantifying expressions in French, Dutch and English*. Ph. D. thesis, University of Leiden.

Erbach, K. (2021). *Object Mass Nouns: A Frame Based Analysis*. Ph. D. thesis, Heinrich Heine Universität Düsseldorf.

Filip, H. and P. R. Sutton (2017). Singular count NPs in Measure Constructions. *SALT 27*, 340–357.

References II

Fillmore, C. J. and P. Kay (1987). Construction grammar. Lecture notes from University of California, Berkeley.

Grimm, S. (2014). Individuating the abstract. *Proceedings of Sinn und Beduetung 18*, 182–200.

Krifka, M. (1989). Nominal Reference, Temporal Constitution and Quantification in Event Semantics. In R. Bartsch, J. van Benthem, and P. van Emde Boas (Eds.), *Semantics and Contextual Expression*, pp. 75–115. Foris.

Landman, F. (2011). Count Nouns – Mass Nouns – Neat Nouns – Mess Nouns. *The Baltic International Yearbook of Cognition 6*, 1–67.

Landman, F. (2016). Iceberg Semantics for Count Nouns and Mass Nouns: Classifiers, measures and portions. *The Baltic International Yearbook of Cognition 11*, 1–48.

Landman, F. (2020). *Iceberg Semantics for Mass Nouns and Count Nouns: A New Framework for Boolean Semantics. Studies in Linguistics and Philosophy 105*. Springer.

Landman, F. (2021). Iceberg semantics for count nouns and mass nouns: How mass counts. In H. Filip (Ed.), *Countability in Natural Language*. Cambridge University Press.

Lima, S. (2014a). All notional mass nouns are count nouns in Yudja. *Proceedings of SALT 24*, 534–554.

References III

Lima, S. (2014b). *The Grammar of Individuation and Counting*. Ph. D. thesis, UMass Amherst.

Muromatsu, K. (1995). The classifier as a primitive: Individuation, referability and argumenthood. *University of Maryland Working Papers in Linguistics 3*.

Pelletier, F. J. (1975). Non-singular reference: Some preliminaries. *Philosophica 5*(4), 451–465.

Rothstein, S. (2010). Counting and the Mass/Count Distinction. *Journal of Semantics 27*(3), 343–397.

Rothstein, S. (2017). *Semantics for Counting and Measuring*. Cambridge University Press.

Schwarzschild, R. (2011). Stubborn distributivity, multiparticipant nouns and the count/mass distinction. *Proceedings of NELS 39*, 661–678.

Soja, N., S. Carey, and E. Spelke (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition 38*, 179–211.

Stanley, J. (2002). Nominal restriction. In *Logical Form and Language*, pp. 365–390. Oxford University Press.

Sutton, P. R. and H. Filip (2016). Counting in Context: count/mass variation and restrictions on coercion in collective artifact nouns. *SALT 26*, 350–370.

Sutton, P. R. and H. Filip (2017). Restrictions on subkind coercion in object mass nouns. *Proceedings of Sinn und Bedeutung 21*, 1195–1213.