

ABSTRACT MASS NOUNS AND MASS-COUNT ELASTICITY

Roberto Zamparelli
University of Trento

Mass/Count 101

- **Mass/Count:** a lexical distinction:
dog, apple, city **Mass nouns**
water, blood, furniture **Count nouns**

...reflecting a semantic contrast (notional approach)

mass nouns denote stuff
count nouns denote things

...with clear distributional reflexes

one dog *one blood *one furniture
*much dog much blood much furniture

Semantic Contrast?

Classic semantic account (Quine 1960, Link 1983, Krifka 1998)

- **Mass nouns** denote properties with non-quantized reference (divisibility: if $p(x)$ and y is part of x then $p(y)$ perhaps, down to a point)
 - A part of water is water *(modulo molecules)*
- **Count nouns denote properties with quantized reference** (if $p(x)$ and y is part of x then $\neg p(y)$)
 - A part of a horse is not a horse
- **Mass nouns are homogenous** (Quine 1960)
- If $P(x)$ and $P(y)$ then $P(x+y)$
 - Some water plus some water is water
- **Singular count nouns are not**
 - A horse plus a horse is not a horse

Semantic Contrast?

Classic problems with it (Chierchia 1998b, Rothstein 2010)

- **Similarity of denotation, different**
 - Wheat, rice vs. lentils, peas
- **Difference across languages**
 - Hair/capelli (it), spaghetti, broccoli
- **Existence if intuitive (if heterogenous) minimal elements**
 - Footwear, mail, furniture, carpeting, luggage
- **Existence of count nouns with non-quantized reference:**
 - Wall, fence, object, splinter,

Elasticity

- Mass/Count Elasticity (Chierchia 2010; Pelletier 1979)
 - Traditional coercion case
 - (count → mass)
After the impact, the meteorite was all over the area ("grinding")
 - We tasted alligator in Australia ("food")
(mass → count)
 - Can I have one water please ("packaging")
 - We enjoy these wines ("kind formation")
- Are these switches:
 - The result of a finite set of linguistic transformations (akin to zero derivational affixes) which apply over the original root meaning?
 - The result of pragmatic operations to which possibly any noun could be (more or less forcibly) subjected?

Linguistic factors

- 'Food coercion' can be blocked (Ostler&Atkins, 1991):
 - I will eat ??pig/pork tonight (cf. «the hero's *gloriosity/glory»)
- 'coercion' seems able/forced to apply after pluralization
 - Two grams of apple / two kilos of apple??(s)
 - There was apple/*pea/*bean in the soup (vs. *bean paste*, *pea soup*)
- coercion can be stacked
 - The chef can prepare two lambs, one with onions, one with tomato.
(animal_{count} → food_{mass} → kind_{count})

Linguistic factors

- Some coercions are impossible:
 - The shark ate *(a) school of fish
 - I put on a *footware/shoe,
 - he assembled a new IKEA *furniture/table
- It might be impossible to establish the direction of the process (elasticity, not coercion)
 - Hope / hopes, fear / fears, work/works, building/buildings
- Productivity/Lexicality trade-off

Questions with coercion

- Which types of mass/count elasticity exist?
- In which direction do they go?
- To which kind of nouns can they apply?

- Perhaps we answer by looking at the frequency of various noun meanings in actual usage (Katz & Zamparelli, 2010)

K&Z 2010: Hypothesis

- If Mass/Count is lexical, nouns should have primary use and secondary use:

much water > many waters
 much blood >> many bloods
 many apples > much apple
 many cars >> much car

- Corpus study:
 - Run on the UKWAC English corpus (Baroni et. al. 2006)
 - 2.8 billion words, web data (not clean!);
 - tagged for parts-of-speech, lemmatized
 endowed/VVN with/IN <much/JJ wisdom/NN>
 because/IN <many/JJ system/NNS> cannot/NN run/VVP
 - Processed with CorpusWorkBench (Evert et. al 2009)

Mass/Count: some indicators

- **Count nouns** don't appear in the "bare" singular, mass noun can.
 - *Restaurant/wine is good
 This restaurant is good
- **Mass nouns** don't appear in the plural:
 - Rice is good
 *Two rices are good
- Certain quantifiers select for **mass** or **count**:
 - Much of the water evaporated.
 - * Much of the people left.
 - * Each of the water evaporated.
 Each of the people left.

K&Z 2010: Quantificational Patterns

Quantifiers and modifiers associated to Mass/Count uses

- Mass indicators
 - *enough, much, lots of, plenty of, sufficient, considerable, boundless, ample, limited*
 - CQP query:

```
[word!="-a|an"] ("lots|plenty" "of" | "much"
|"more"|"less"|"enough"|"most"|"sufficient"|"considerable"|"boundless"|"ample"|"limited")@[pos="NN"] [pos!="-VVN|JJ."|N.*"]
```
- Count indicators
 - *an, every, each, one, first, another*
 - CQP query:

```
[lemma="a|an|one|every|first|each|another"] [pos="J.*"]* @[pos="NN"] [pos!="N.*"];
```

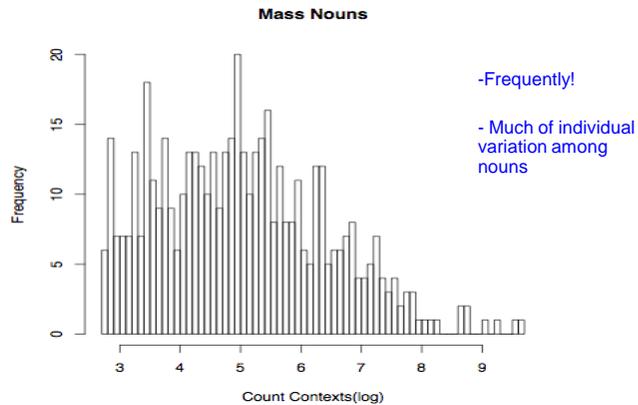
K&Z (2010) "Mass Nouns"

Indeed, nouns occurring in the mass contexts **look like mass nouns**:

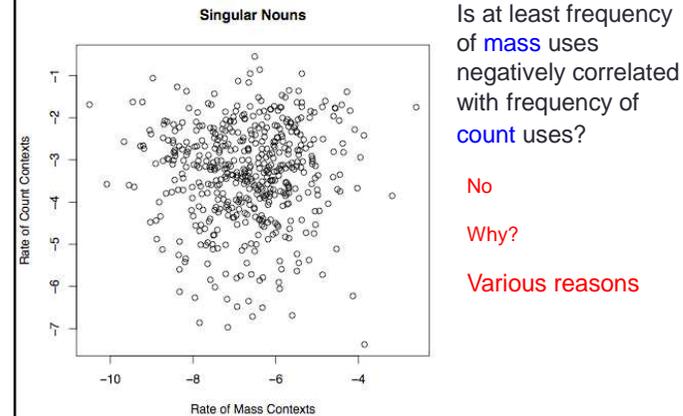
information	8031	energy	342	...	
time	2687	room	337	
money	1280	power	313	magnification	2
detail	946	water	307	lettuce	2
info	708	support	248	consumer	2
fun	597	use	243	stage	2
space	525	effort	240	permanence	2
part	470	research	218	rockabilly	2
attention	450	emphasis	218	algae	2
work	437	choice	201	respect.	2
experience	393	sense	200	grist	2
interest	377	...		market	2
evidence	372	...			

- But how often do these "mass nouns" occur in **count** contexts?

Count Uses of "Mass Nouns"



K&Z: Mass vs. Count Uses



Search Issues

Searching a (web) corpus can be tricky

- **Compounds:**
 - **Much computer** power, **enough computer**-based work
 - **Much cat** and dog-style fighting
- **Misparsing:**
 - I didn't do **much Boy** was I tired!
 - A cute **little dog**
- Bare nouns of all kinds common in section titles
- Faulty Part-Of-Speech tags

The rest: bona fide coercion

But not only the one we would probably expect:

- *this gadget, **more computer** than smartphone,*
- *How **much car** you can afford?*
- *Phantom 87s were **too much ski** for me*
- *I always practice **too much piano***
- *Don't use **too much battery***
- **Quotational uses:**
 - *"more properly described as 'computer' than as 'smartphone', too much "ski", that sound is so much '70ies"*
- **Compound-like semantics:**
 - *I hear **too much piano** (sound)*
 - *A lot of New York (themes/atmosphere) in his writings.*

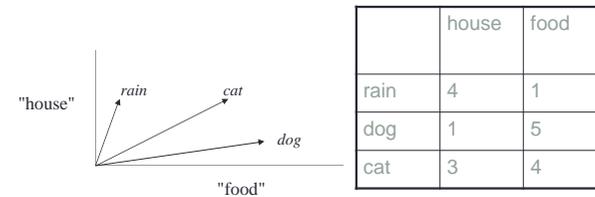
Number Related Meaning Shifts

One aspect that we can reliably extract from corpora is the plural/singular distinction.

- Assuming that masses do not pluralize without changing meaning, plurals of masses (i.e. Ns occurring with "much", etc.) will **always** have some meaning shift.
 - Wine = mass (poured w.) / kind (is alcoholic) / portion (have a w.)
 - Wines = kind / portion
- Can we measure this meaning shift?
- In K&Z2010, we tried to answer this question using Vector Space Semantic models

Vector Space Models

- Using **lexical co-occurrence** to induce a quantitative model of meaning similarity (Schütze 1993; Lund & Burgess 1996; Landauer & Dumais 1997; Sahlgren 2006; Baroni et al, in press)
 - Similarity of distribution → similarity of meaning
 - Surprisingly good correlations with humans semantic similarity judgements (Landauer & Dumais 1997)

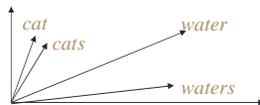


Note: typical measure of similarity is cosine of angle between vectors

Vector Space Comparison

Using UKWAC, we built meaning vectors for Nouns tagged SING or PLUR

Computed the similarity between singular and plural form lemma



- Expectation**
 - Basic Count Nouns (cat, dog) → similar singular/plural distribution
 - Basic Mass Nouns (water, beer) → different distributions (because of Mass-Count shifts)

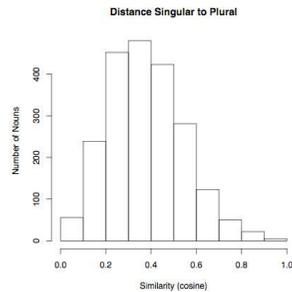
Model Details

- COALS (Rohde et al 2005) algorithm using **airhead-software java package**
 - Gather co-occurrence counts, ignoring closed-class neighbors in a **ramped, 4 word window**
 - Maintain co-occurrence counts with 14,000 most frequent content words
 - Convert counts to correlations, set negative values to 0, and take square roots of positive ones.
- Model trained on first part of UKWAC corpus
 - stopwords excluded (571 frequent word types)
 - words were coded as lemma-POS tokens**
the-DT three-CD dog-NNS
- Model contains 2131** nouns in both singular/plural
 - contains 20,000 words total (with 14000 dimensions)

Singular/Plural Similarity (1)

Lots of word-to word variation in singular to plural distance

No evidence of shifter (mass) and nonshifter (count) classes



Similarity Examples

Semantically Closest Sing to Plur

phone - phones
therapist - therapists
resort - resorts
impairment - impairments
reaction - reactions
list - lists
century - centuries
speaker - speakers
cookie - cookies
engine - engines
locomotive - locomotives
pool - pools
sector - sectors
cancer - cancers

Semantically Furthest:

leave - leaves
make - makes
creator - creators
con - cons
humanity - humanities
extreme - extremes
good - goods
disadvantage - disadvantages
toddler - toddlers
strength - strengths
fortune - fortunes
horizon - horizons

Observationally classes look different

Mass/Count?

• Semantically Furthest

leave - leaves
make - makes
creator - creators
con - cons
humanity - humanities
extreme - extremes
good - goods
disadvantage - disadvantages
toddler - toddlers
strength - strengths
fortune - fortunes
horizon - horizons

• Several of these are mass:

much humanity
much strength
much fortune
much good
much disadvantage

• Does singular/plural semantic distance correlate with rate of mass context uses?

Count/Mass Rate and Similarity

Using a slightly different method (neighbor count):

Rate of **mass** contexts is significantly **higher** for "far" nouns than for "near" nouns

0.0028 vs. 0.0014, $p = 0.003041$

Rate of **count** contexts is significantly **lower** for "far" nouns than for "near" nouns

0.192 vs. 0.149, $p = 7.09e-09$

"Near" nouns are used in more predominantly count (and less predominantly mass) contexts than "far" nouns

Katz Observations:

'Anna Karenina' effect:

"Happy families are all alike; every unhappy family is unhappy in its own way." Leo Tolstoy, *Anna Karenina*

- "Near" sing/plur nouns are all near for the same reason
→ They are normal count nouns (*chair/chairs*)
- "Far" sing/plur nouns are far for many different reasons
 - Mass nouns with meaning shift: *water/waters*
 - Ambiguous nouns (cf. Acquaviva, 2008): *arm/arms, game/games*
 - we sell arms, we eat game
 - Differential Idiomatic usage: *eye/eyes*
 - can see it with the naked eye/?naked eyes

Limitations of Method

- Singular/plural splits can only detect count-to-mass meaning shifts for nouns whose singular use is predominantly mass
 - Many abstract nouns do not have this characteristic
belief/beliefs
hope/hopes
 - And abstract nouns make up the vast majority of mass nouns: 88% in our hand sample
- Can we somehow get at this kind of elasticity?

Beyond Karenina

Can we at least find regular classes of noun elasticity within the data?

- A sample of our mass data (=N frequently occurring with mass Dets)

action, activity, advice, attention, bandwidth, beer, caffeine, care, cash, chance, choice, comfort, content, control, credit, damage, depth, detail, effort, energy, evil, exercise, experience, fish, flavour, flexibility, food, force, fun, guidance, hair, harm, heat, help, housing, info, information, input, involvement, kindness, knowledge, light, luck, money, noise, pain, passion, point, power, proof, publicity, question, ram, reality, reason, rent, respect, room, sense, service, skill, space, staff, success, support, tax, time, tolerance, training, trouble, value, water, wealth, weight, wit, wood, work

Beyond Karenina

- The vast majority of case are not objects like «water», nor aggregates like «furniture», but rather **abstract objects**.
- Can we find patterns in this domain?

action, activity, advice, attention, bandwidth, beer, caffeine, care, cash, chance, choice, comfort, content, control, credit, damage, depth, detail, effort, energy, evil, exercise, experience, fish, flavour, flexibility, food, force, fun, guidance, hair, harm, heat, help, housing, info, information, input, involvement, kindness, knowledge, light, luck, money, noise, pain, passion, point, power, proof, publicity, question, ram, reality, reason, rent, respect, room, sense, service, skill, space, staff, success, support, tax, time, tolerance, training, trouble, value, water, wealth, weight, wit, wood, work

Abstracts Ns: preliminaries

- Multiple ways/criteria to be «abstract»
 - Orbit, mammal, hope, divisibility
 - Material objects > events > psych. states > notions
- Typically no specific measures, except time:
 - A lot of happiness / *two hedons of happiness
 - A moment of happiness / two hours of intense communication
- Cumulativity? Divisibility? Subparts?
 - Difference + difference = difference?
 - A subpart of happiness / change / challenge?
 - **But:** He could do with half of that happiness / challenge

Abstract Meanings elasticity

- No clear direction of 'coercion'
 - Belief / beliefs, depth / depths
- Plural version is typically *less abstract*
 - Comfort / comforts
- Special behaviour with the singular indefinite
 - A ??(profound) happiness, a ??(great) peace
- Special behavior with certain conjunctions
 - His patience and tenacity were remarkable

Abstract Meanings Shifts: some cases

caution	cautions	N-plur: eventualities which exemplify instantiate N-sing-mass
challenge	challenges	
chance	chances	
change	changes	Sometimes «social objects» (in the Sense of DOLCE, Gangemi et al. 2002: <i>duties, Comforts, ...</i>)
choice	choices	
comfort	comforts	
commitment	commitments	
communication	communications	
conflict	conflicts	
debate	debates	
difference	differences	
difficulty	difficulties	
duty	duties	

Abstract Meanings: further shifts

contact	contacts	(results from contacting episodes)
control	controls	(e.g. to control a game)
choice	choices	(«this car was my choice»)

Psych. attitudes and their direct object content N-plur:

belief	beliefs	But not:		
doubt	doubts			
desire	desires		shame	*shames
hope	hopes		courage	*courage
fear	fears		reluctance	*reluctances
feeling	feelings			

Abstract Meanings: further shifts

«Kind» meaning shift?

content	contents
context	contexts
background	backgrounds
colour	colours

'Extent' shift (cf. Acquaviva 2008): water/waters, sand/sands

depth	depths
length	lengths

Disentangling the meanings

- «Notion» reading, not a mass; probably a Carlsonian kind (the proper name of a property)
 - Calm is a virtue / (*Enough) difference is important in modern society
- But typical kind predicates give mixed results:
 - True calm is {rare /?common}. ?Choice appeared on Earth with internet
- What about the mass reading?
 - We need more choice, more calm, more change

Abstract Ns: a proposal

• General idea:

- Abstract nouns such as challenge, change, choice ... (and maybe others) denote the sets of all events that properly contain subevents of challenging, changing, or choosing something.

• $\text{Choice}_{\text{mass}} = \lambda e[\exists e'[e' < e \ \& \ \text{CHOICE}(e')]]$

- While events have countable atoms («three good choices»), superevents can be extremely heterogeneous. The only thing they have in common is that they contain a certain type of subevent -- all the rest is left open.
- This gives the superevent the atom-level vagueness required by some theories of mass (e.g. Chierchia 1998, Rothstein, 2010)

Evidence: Abstract Ns inside DPs

- Italian and most (European) languages, but not English, allow singular N&N coordination under a single Det only when N&N can apply to a single entity (Heycock&Zamparelli 2005): (1), but not (2)

- 1) a. Una amica e collega era qui (joint, 1 person)
b. A friend and colleague was here
- 2) a. *Un padre e figlio erano qui (split, 2 people)
b. A father and son were here

Evidence: Abstract Ns inside DPs

One apparent exception is abstract nouns like (1), which can appear under a single DET in many languages etc. (provided N gender and number are matched).

- 1) La sua pazienza e determinazione furono esemplari
his patience and determination were exemplar_{plur}

This language difference is reflected in that fact that English has about always twice as many noun types appearing in this construction than Italian (data from UKWAC and ITWAC).

Evidence: Abstracts inside DPs

This conjunction is impossible with concrete mass nouns, unless understood as substance mix (singular verb) (1)

- 1) *La sua birra e limonata erano calde
the his beer and lemonade were warm.

Also impossible as coordination of two «notions»

- 3) La pazienza e ??(la) determinazione sono virtù importanti
the patience and (the) determination are important virtues

Proposal: Merging compatible events

- «patience and determination» must be a joint/intersective reading, agreement notwithstanding.
- «Patience» ≠ «determination»: the set of patience-events is different from the set of determination-events
- But an event containing «patience» as a subevent could also contain «determination» as a subevent.
- **Hypothesis**: intersection over eventualities acts as **unification** (events can be unified if their subevents are *compatible*, not *identical*)

Consequences

- «John's patience and determination» does not mean that John had a quality which can simultaneously be described as 'patience' and 'determination'
- Rather, it makes reference to a set of events which has John as agent and which partly involved his being patient and/or determinate.
- So, it is a joint reading as in «friend and colleague»

Outlook

- Much remains to be done in lexical semantics for the vast majority of the nominal lexicon: the abstract part.
- If we can extract reasonable examples of various subtypes of meaning shifts we can use computational semantic methods (in particular, functional vector space semantics, see Baroni et. al. In press) to categorize the lexicon.
- These results could then be examined 'by hand', to arrive at a more precise characterization, and better formal theories.

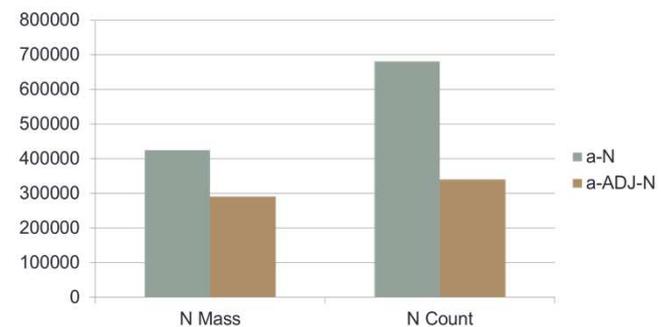
Thanks!

Appendix: modified indefinites

- Non-notional readings are masses, yet can often appear with «a», but require (or imply) some kind of modification.

“a silence”	249 tokens (BNC)
“a ADJ silence”	385 tokens

Proportion of modified indefinites



BNC data, «mass» defined simply by low # of plural tokens

Why modified mass indefinites?

- There was a certain peace/silence/discussion (cf. Himmelweise)
- Kind-reading? Unlikely.
 - A {certain *amount* / *quantity* /??*kind*} of *peace*
 - A *(smiling) Barack Obama entered the senate
- Probably a covert measure reading