A probabilistic, mereological account of the mass/count distinction

Peter Sutton
(with Hana Filip)

CLASP Seminar
University of Gothenburg
30^{th} November, 2016

SFB 991

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

The Mass/Count Distinction

- Found in many nominal systems of natural languages
- But tests to reveal distinction may vary
- Number marking languages (E.g. English, Swedish, Finnish)
  - The ‘signature property’ is the most general
  - Direct attachment of a numerical expression to a noun (without coercion)

(1) Alex bought three chairs

(2) ??Alex bought three muds.

- Classifier languages (E.g. Mandarin, Japanese) behave differently
  - No nouns display the signature property (need for intervening classifier).
  - But increasing evidence of a count/mass distinction e.g. Sudo (2016) for Japanese:

(3)

kinoo-no	jiko’de-wa	tasuu-no	sisha-ga	deta	yooda
yesterday-GEN	accident-LOC-TOP	many-GEN	fatality-NOM	came.out	EVID

‘It seems that the accident yesterday resulted in many fatalities.’

(4)

#Taro-wa	tasuu-no	ase-o	kaita
Taro-TOP	many-GEN	sweat-ACC	secreted

Int: ‘Taro sweated a lot.’

Mass/Count Variation: Data

- Languages differ in their count/mass lexicalization patterns.
- Even more variation in abstract and event-denoting Ns (we will focus exclusively on concrete Ns).
- There are stark similarities
  - ‘Substance’ denoting nouns are highly probably mass
    - *mud*, *blood*, *air*, *slime*
  - Ns denoting single discrete objects (esp. animate, large) are highly probably count
    - *woman*, *cat*, *car*, *chair*
- But many cases of cross- and intralinguistic variation

CROSSLINGUISTIC VARIATION

furniture$_{-C,\text{SG}}$	=	*huonekalu-t*$_{+C,\text{PL}}$ (Finnish)
footwear$_{-C,\text{SG}}$	=	*jalkine-et*$_{+C,\text{PL}}$ (Finnish)
kitchenware$_{-C,\text{SG}}$	≈	*küchengerät-e*$_{+C,\text{PL}}$ (German)
lentil-s$_{+C,\text{PL}}$, *linssi-t*$_{+C,\text{PL}}$ (Finnish)	=	*čočka*$_{-C,\text{SG}}$ (Czech), *lešta*$_{-C,\text{SG}}$ (Bulgarian)
bean-s$_{+C,\text{PL}}$	=	*bob*$_{-C,\text{SG}}$ (Bulgarian)

INTRALINGUISTIC VARIATION

meubel-s$_{+C,\text{PL}}$	vs.	*meubilair*$_{-C}$ (furniture, Dutch)
shoe-s$_{+C}$	vs.	*footwear*$_{-C}$
seed-s$_{+C,\text{PL}}$	vs.	*seed*$_{-C}$
oat-s$_{+C,\text{PL}}$	vs.	*oatmeal*$_{-C}$

Mass/Count Variation: Main question

- There are two distinct questions regarding count/mass variation
  1. Why is N_1 count in L_1 when its cognate, N_2, is mass in L_2?
     - E.g. Why is *furniture* mass, but *huonekalu* ('(item of) furniture', Finnish) count?
  2. Why is there much variation across languages for Ns that refer to X, but little variation in Ns that refer to Y?
     - E.g. Why do Ns that refer to furniture vary much in their count/mass lexicalization patterns when Ns that refer to cats do not?
- Although both are interesting questions, we will only really discuss 2.
  - Answering 1. would probably require detailed work in lexical semantics in every language.
  - We propose to try to answer 2. by examining more general properties of the interface between the world, language and cognition.
- However, also some speculations about Yudja (a language in which all notional mass nouns have been argued to be count nouns (Lima, 2014))

Outline

Introduction

Background: Pressures/constraints from learning and communication

Background: Mass/Count Literature

Analysis: Formally modeling the constraints of individuation and reliability

Lexical entries and deriving mass/count variation from the constraints

Piantadosi, Tily, and Gibson (2011): Ambiguity

- Zipfian Background
  - Competing principles of least effort for hearer and speaker.
- Piantadosi et al. (2011)
  - *Clarity* versus *Ease*
  - CLARITY: "A clear communication system is one in which the intended meaning can be recovered from the signal with high probability."
    - Pushes towards unequivocal signal system
  - EASE: "An easy communication system is one in which signals are efficiently produced, communicated, and processed"
    - Pushes towards simpler (smaller, more often used) system of equivocal signals.
- Ambiguity in language is a result of balancing these pressures/constraints

Sutton (2013, 2016): Vagueness

- Vagueness arises naturally as a result of pressure from communication and learning
- Similarly to Zipf:
  - Maximizing informational content would mean simple encoding and decoding of information in communication
    - E.g. tall_{487} means *greater that 176.3 cm in height*
  - But this would lead to instability (not enough speakers would be exposed to enough meanings). (Kirby and Hurford, 2002; Kirby, 2007)
  - For a stable, effective means of communication that is learnable, predicates in languages must be balanced between being informative (specific), and learnable in a relatively small learning phase (general).
- But learning predicates requires abstracting over information from a limited number of uses
  - Can give rise to graded representations
  - Graded probabilistic representations explain facts associated with the use of vague predicates.

Summary: Piantadosi et al and Sutton

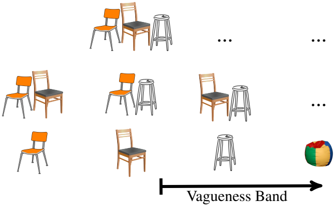
- General pressures on languages regarding learning and communication:
  1. Semantic learning: the meaning of an expression should be learnable from a limited number of instances.
  2. Communication: expressions should convey information sufficient amounts of information to be effective tools for communication.
- The way these pressures compete can explain the abundance of specific linguistic phenomena.
  - E.g. vagueness, ambiguity

Proposal Outline

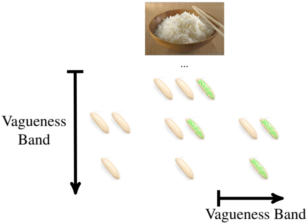
- We will apply a similar strategy to explain the distribution patterns of count/mass variation cross- and intralinguistically.
- Call a way of splitting up a number neutral predicate's denotation into individual entities an *individuation schema*
- For concrete nouns we will argue these two pressures translate into:
  1. **Reliability:** There should be an individuation schema that reliably predicts when to apply the predicate (a quantitative criterion of application).
  2. **Individuation:** Individuation schemas should convey sufficient amounts of information to be effective tools for identifying individuals.
- We will derive mass/count variation patterns in terms of when these two pressures can or cannot be jointly satisfied.

Chierchia 2010: Mass/Count is a matter of vagueness

- STABLE ATOMICITY of count Ns sanctions counting
  - Any count N has at least some elements that are atomic across all admissible precisifications

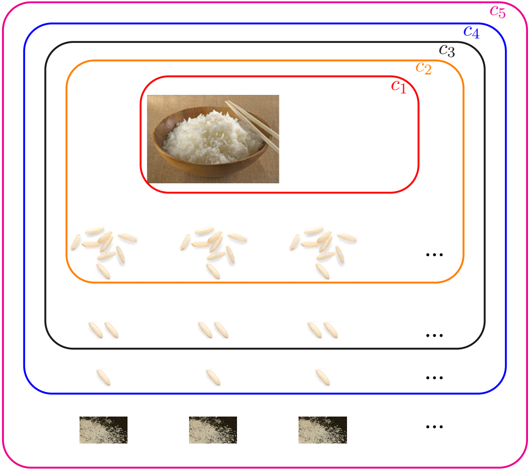


- VAGUENESS in mass Ns blocks counting
  - Mass Ns have *unstable individuals* in their denotation
  - Minimal individuals/atoms on some precisifications are not minimal/atomic on others



Chierchia (2010): Supervaluating over contexts

- Perhaps better thought of as a form of context sensitivity

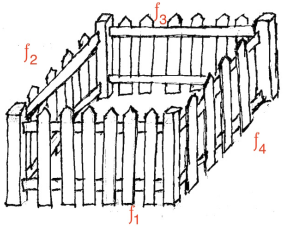


Chierchia (2010): Summary

- Take home message: Form some predicates P, there are many contexts in which single individuated units of P (grains etc.) are not sufficient in quantity to count as P.
- Our Analysis: The most natural individuation schema (in terms of grains) for some Ps provides an *unreliable* basis to identify when to apply P

Rothstein (2010) [+COUNT] means semantically atomic

- Atomicity relative to a context k, where k is a set of entities that ‘count as one’ in a given context.



$F = \{$

$f_1 \sqcup f_2 \sqcup f_3 \sqcup f_4$

$f_1 \sqcup f_2 \sqcup f_3$ $\quad f_1 \sqcup f_2 \sqcup f_4$ $\quad f_1 \sqcup f_3 \sqcup f_4$ $\quad f_2 \sqcup f_3 \sqcup f_4$

$f_1 \sqcup f_2$ $\quad f_1 \sqcup f_3$ $\quad f_1 \sqcup f_4$ $\quad f_2 \sqcup f_3$ $\quad f_2 \sqcup f_4$ $\quad f_3 \sqcup f_4$

f_1 $\quad f_2$ $\quad f_2$ $\quad f_2$

$\}$

- $N_{\text{count}} = COUNT_k(N_{\text{root}})$
  - $COUNT_k(F) = \{\langle d, k\rangle : d \in F \cap k\}$
  - $k_1 = \{f_1, f_2, f_3, f_4, g_1, g_2, ...\}, \quad k_2 = \{f_1 \sqcup f_2 \sqcup f_3 \sqcup f_4, g_1, g_2, ...\}$
  - $COUNT_{k_1}(F) = \{\langle f_1, k_1\rangle, \langle f_2, k_1\rangle, \langle\{f_3, k_1\rangle, \langle f_4, k_1\rangle\}$ $\Rightarrow$ *Four fences*
  - $COUNT_{k_2}(F) = \langle f_1 \sqcup f_2 \sqcup f_3 \sqcup f_4, k_2\rangle$ $\Rightarrow$ *One fence*

Rothstein (2010): Summary

- Take-home message
  - For some nouns (e.g. *fence*, *hedge*, *wall*), what counts as 'one' varies with context.
- Our Analysis: There is no single individuation schema that correctly identifies single fences, hedges, walls etc. in every context.

Landman (2011): Mass means Overlap

- Notional set of countable entities modeled as **generator sets**
  - Generator sets *generate* full N denotations under mereological sum $\sqcup$
  - A generating set for X is a set $\mathbf{gen}(X) \subseteq X - \{0\}$ such that: $\forall x \in X : \exists Y \subseteq \mathbf{gen}(X) : x = \sqcup Y$
- Chierchia's approach: *underdetermination* wrt what to count
- Landman's approach: *overdetermination* wrt what to count
  - For mass Ns, there are overlappping entities that count as one "simultaneously in the same context" in the selected generator set
- Count Ns (e.g. *cat*): No overlap in the generator set (the set of single cats) $\Rightarrow$ only one way to count
- Two kinds of mass Ns
  - Mess mass nouns (*mud*, *blood*, *air*)
  - Neat mass nouns (*furniture*, *kitchenware*)
- We focus on Landman's account of neat mass nouns

- Set of generators for *kitchenware*: What intuitively counts as one
  - This set overlaps
  - E.g. {pestle, mortar, pan, lid, pestle $\sqcup$ mortar, pan $\sqcup$ lid}
- Overlap resolved at *variants* (which form maximally disjoint subsets)
- But these variants form different cardinalities of countable entities
  - E.g. v_1 = {pestle, mortar, pan, lid} = 4 items
  - E.g. v_2 = {pestle $\sqcup$ mortar, pan $\sqcup$ lid} = 2 items
- Counting goes wrong, therefore *kitchenware* is mass

Landman (2011) Summary

- Take-home message
  - Denotations of certain mass nouns include multiple different ways of ‘splitting them up’ into countable units.
- Our Analysis: There is no single individuation schema that correctly identifies single items of furniture, kitchenware, jewellery etc. in every context.

Interim Analysis of Background Literature

Some important observations have been made in the recent literature:

- Nouns such as *fence*, *hedge*, *wall* AND nouns such as *kitchenware*, *furniture*, *jewellery*
  - There is no single individuation schema that correctly identifies single items of furniture, kitchenware, jewellery etc. in every context.
- Nouns such as *lentils*, *beans*, *rice*:
  - The most natural individuation schema (in terms of grains) for some Ps provides an *unreliable* basis to identify when to apply P

We will use these data in out anlysis.

Qualities and Measures

- Krifka (1989) distinguished QUALITATIVE and QUANTITATIVE factors in applying predicates
  - All nouns encode a qualitative criterion of application (MUD(x) and COW(x)] below)
  - Count nouns also encode extensive 'natural unit' measure functions
  - Measure phrases (e.g. *kilo of*) express extensive measure functions and encode a quantitative criteria of application

$$\begin{aligned} [\![\text{mud}]\!] &= \lambda x. & [\text{MUD}(x)] \\ [\![\text{cow}]\!] &= \lambda n.\lambda x. & [\text{COW}(x) \wedge \textbf{NU}(\text{COW})(x) = n] \\ [\![\text{kilo of}_{\text{simplified}}]\!] &= \lambda n.\lambda P.\lambda x. & [P(x) \wedge kg(x) = n] \end{aligned}$$

- We assume (closer to Landman (2011, 2016)) that all nouns encode quantitative and qualitative criteria of application
- Also like Landman – *Mass if counting base is not disjoint*

Lexical entries in mereological TTR

- A simplistic entry for a concrete noun like *cat*:

$$[\![\text{cat}]\!] = \lambda r : [x : \mathit{Ind}]. \left[\; s_{\mathit{cat}} \quad : \quad \text{cat}(r.x) \; \right]$$

- A slightly less simplistic entry for *cat*

$$[\![\text{cat}]\!] = \lambda r : [x : \mathit{Stuff}]. \begin{bmatrix} s_{\mathit{cat}} & : & \text{cat}(r.x) \\ s_{\mathit{cat\text{-}ind}} & : & \text{cat}_{\mathit{Ind}}(r.x) \end{bmatrix}$$

  - cat(x): Regular p-type (but assumed to be number-neutral)
  - Stuff: Basic type. No assumption of individuation.
  - $\text{cat}_{\mathit{ind}}$:
    - Specialized p-type for individuating cats
    - Builds in both quantitative and qualitative criteria for predicate application.

- Domain of entities structured as Boolean semilattice closed under sum
- More will be said about types such as $\text{cat}_{\mathit{ind}}$

P_{Ind} types

- Types that individuate Ps are a special case of a quantitative measure function on stuff with P qualities
  - Function from stuff which has some (to be specified) P properties, to a measure value i

$$\begin{bmatrix} s_{P_{qual}} & : & \begin{bmatrix} x & : & \textit{Stuff} \\ s_{P_{pptys}} & : & P_{pptys}(x) \end{bmatrix} \\ f_{P_{quant}} & : & (\begin{bmatrix} x & : & \textit{Stuff} \\ s_{P_{pptys}} & : & P_{pptys}(x) \end{bmatrix} \rightarrow \mathbb{N}) \\ i & : & \mathbb{N} \\ f_{p_{quant}}(s_{P_{qual}}) & : & \mathbb{N}_i \end{bmatrix}$$

- Abbreviation: When measure value is 1, this type is abbreviated to $P_{ind}(x)$

$$[\![\text{cat}]\!] = \lambda r : [x : \textit{Stuff}]. \begin{bmatrix} s_{\text{cat}} & : \text{cat}(r.x) \\ s_{\textit{cat-ind}} & : \begin{bmatrix} s_{cat_{qual}} & : \begin{bmatrix} s_{\text{cat}_{pptys}} & : \text{cat}_{pptys}(x) \end{bmatrix} \\ f_{\textit{cat-quant}} & : (\begin{bmatrix} s_{\text{cat}_{pptys}} & : \text{cat}_{pptys}(r.x) \end{bmatrix} \rightarrow \mathbb{N}) \\ f_{\textit{cat-quant}}(s_{cat_{qual}}) & : \mathbb{N}_1 \end{bmatrix} \end{bmatrix}$$

$$= \lambda r : [x : \textit{Stuff}]. \begin{bmatrix} s_{\text{cat}} & : \text{cat}(r.x) \\ s_{\textit{cat-ind}} & : \text{cat}_{Ind} \end{bmatrix}$$

Characterizing the pressures in prob-TTR

1. **Reliability:** There should be an individuation schema that reliably predicts when to apply the predicate (a quantitative criterion of application).
2. **Individuation:** Individuation schemas should convey sufficient amounts of information to be effective tools for identifying individuals.

- These pressures can be modeled more effectively in the probabilistic version of TTR
  - Linked to a learning model (Cooper et al., 2014, 2015) to express more/less reliable indictors of making type judgements.
  - Probabilistic formalism easily expressed in information theoretic terms.

1. **Reliability:** Maximize a (to be defined) pair of conditional probabilities .
2. **Individuation:** Minimize entropy wrt establishing a counting result.

Reliability: Maximize the conditional probabilities

(x : *Stuff* left out for brevity)

- Maximizing two conditional probabilities yields reliability.
- Find the $P_{Ind_j}(x)$ type such that

1. $Max_j(p(r : \left[\ s_P \ : \ P(x)\ \right] | r : \left[\ s_{P\text{-}ind} \ : \ {}^*P_{Ind_j}(x)\ \right]))$
   - So being a P-individual or a sum of P-individuals is a very strong indicator of being a P
   - Militates against over-inclusivity of $P_{Ind_j}(x)$
2. $Max_j(p(r : \left[\ s_{P\text{-}ind} \ : \ {}^*P_{Ind_j}(x)\ \right] | r : \left[\ s_P \ : \ P(x)\ \right]))$
   - So being a P is a very strong indicator of being a P-individual or a sum of P-individuals.
   - Militates against under-inclusivity of $P_{Ind_j}(x)$

- Open empirical question: What kind of Max function?
  - Absolute Max will return the trivial: $p(a : T|a : T)$

Individuation: Minimize Entropy for Individuation Schemas

- The property of being disjoint can be derived from more general informativeness constraints (cf Pientadosi et al)
  - Key idea: A disjoint individuation schema has minimum entropy with respect to determining a counting result.

A type T is disjoint relative to a probability threshold θ:

$T : Disj_\theta \leftrightarrow$ for all a, b such that $p(a : T) \geq \theta$ and $p(b : T) \geq \theta$, if $a \neq b$, then $a \sqcap b = 0$

- Equivalent to standard *disjointness* at the upper limit.
  - Intuitive idea: If we are certain enough that there are overlapping entities that are of type T, then we can judge that T is not disjoint.
  - A graded alternative would be possible.

Individuation: Minimize Entropy for Individuation Schemas

- For non-disjoint types, one can form maximally disjoint subtypes (variants).
- Disjoint types have only one variant (identity)
- Probability distributions over variants in terms of entropy.

$$Min_j(-\sum_{v_i \in V} p(v_i|P_{Ind_j}) \times \log p(v_i|P_{Ind_j}))$$

- Average amount of information needed to determine a specific counting result.
  - Assuming an equal distribution over variants

Number of variants	1	2	4	8	16
Entropy	0	1	2	3	4

- Effect: Pushes towards a disjoint Individuation schema.
- Can also model a cost to context sensitivity C:

$$Min_j(-(\sum_{v_i \in V} p(v_i|P_{Ind_j}) \times \log p(v_i|P_{Ind_j})) + C)$$

- Suggestion: The greater the number of potential disjoint schemas to resolve in context, the greater the cost.

Context specific and context general schemas

- Need some space of Individuation schemas to enter into probabilistic and information theoretic calculations for individuation and reliability.
- If a learner gets evidence of multiple schemas across contexts:
  - $P_{Ind_{c_i}}$: Schema for context c_i
  - $P_{Ind_{gen}}$: Generalized context independent schema.
- In principle, a different schema space could be fed into the system from e.g. a perceptual classifier.

Summary

- Both pressures on lexical entries for concrete nouns can be given probabilistic/information theoretic characterisations
  - Reliability: Pushes towards a general schema of individuation ($^*P_{Ind}$ as close as possible to the number neutral predicate P).
  - Individuation: Pushes towards a specific (and disjoint) schema of individuation P_{Ind}.

- Next: look at some specific cases.
- Conflict: Mass/count variation predicted.
- Unison: Stable mass predicted

Cat

- If cat_{Ind} is the p-type for single cats then:

Reliability: $p(r : [\ s_{cat} \ : \ \text{cat}(x)\ ] | r : [\ s_{\textit{cat-ind}} \ : \ {}^{*}\text{cat}_{Ind}(x)\ ]))$ = v. high

$p(r : [\ s_{\textit{cat-ind}} \ : \ {}^{*}\text{cat}_{Ind}(x)\ ]) | r : [\ s_{cat} \ : \ \text{cat}(x)\ ])$ $=$ v. high

Individuation: $-\sum_{v_i \in V} p(v_i | \text{cat}_{Ind}) \times \log p(v_i | \text{cat}_{Ind})$ $= 0$

- Result: cat_{Ind} satisfies both pressures well.

$$[\![\text{ cat }]\!] = \lambda r : [x : \textit{Stuff}]. \begin{bmatrix} s_{cat} & : \text{cat}(r.x) \\ s_{\textit{cat-ind}} & : \text{cat}_{Ind}(r.x) \end{bmatrix}$$

- Disjoint P_{Ind} type, so count

Constraints for lentil-like Nouns

Single grain individuation schema: lentil_{Ind}

Reliability: $p(r : [\ s_{\text{lentil}} : \textit{lentil}(x)\ ]\,|\,r : [\ s_{\textit{lentil-ind}} : {}^{*}\text{lentil}_{Ind}(x)\ ]))$ $= \text{lowish}$
(small quantities are not good predictors)

$p(r : [\ s_{\textit{lentil-ind}} : {}^{*}\text{lentil}_{Ind}(x)\ ])|r : [\ s_{\textit{lentil}} : \text{lentil}(x)\ ])$ $= \text{highish}$
(doesn't accommodate sub-grain parts)

Individuation: $-\sum_{v_i \in V} p(v_i|Ind_{\text{lentil}}) \times \log p(v_i|Ind_{\text{lentil}})$ $= 0$

Generalized individuation schema: $\text{lentil}_{Ind_{gen}}$

Reliability: $p(r : [\ s_{\text{lentil}} : \textit{lentil}(x)\ ]\,|\,r : [\ s_{\textit{lentil-ind}} : {}^{*}\text{lentil}_{Ind_{gen}}(x)\ ]))$ $= \text{high}$
(distribution of ${}^{*}\text{lentil}_{Ind_{gen}}$ approximates that of $\text{lentil}(x)$)

$p(r : [\ s_{\textit{lentil-ind}} : {}^{*}\text{lentil}_{Ind_{gen}}(x)\ ])|r : [\ s_{\textit{lentil}} : \text{lentil}(x)\ ])$ $= \text{high}$
(distribution of ${}^{*}\text{lentil}_{Ind_{gen}}$ approximates that of $\text{lentil}(x)$)

Individuation: $-\sum_{v_i \in V} p(v_i|\text{lentil}_{Ind_{gen}}) \times \log p(v_i|\text{lentil}_{Ind_{gen}})$ $= \text{v. high}$

- Neither of the two alternatives for individuation can satisfy both constraints.

Lentil versus *Čočka* ('lentil', Czech)

$$[\![\text{ lentil }]\!] = \lambda r : [x : \mathit{Stuff}]. \begin{bmatrix} s_{\mathit{lentil}} & : \text{lentil}(r.x) \\ s_{\mathit{lentil\text{-}ind}} & : \text{lentil}_{\mathit{Ind}}(r.x) \end{bmatrix}$$

- Prioritizes Individuation (minimizes entropy)
  - Disjoint P_{Ind} type, so count

$$[\![\text{ čočka }]\!] = \lambda r : [x : \mathit{Stuff}]. \begin{bmatrix} s_{\mathit{lentil}} & : \text{lentil}(r.x) \\ s_{\mathit{lentil\text{-}ind}} & : \text{lentil}_{\mathit{Ind}_{gen}}(r.x) \end{bmatrix}$$

- Prioritizes Reliability (maximizes conditional probabilities)
  - Not-disjoint P_{Ind} type, so mass.

Constraints for furniture-like nouns

Context-sensitive Individuation Schema:

Reliability $p(r : [\ s_{\textit{furn}} : \text{furn}(x)\ ] | r : [\ s_{\textit{furn-ind}} : {}^{*}\text{furn}_{\textit{Ind}_{c_i}}(x)\ ])) = 1$

$p(r : [\ s_{\textit{furn-ind}} : {}^{*}\text{furn}_{\textit{Ind}_{c_i}}(x)\ ]) | r : [\ s_{\textit{furn}} : \text{furn}(x)\ ]) = \text{lowish}$

(many context specific functions exclude some entities)

Individuation $-(\sum_{v_j \in V} \sum_{c_i \in C} p(v_j | \text{furn}_{\textit{Ind}_{c_i}}) \times \log p(v_j | \text{furn}_{\textit{Ind}_{c_i}})) + |C| = |C|$

(assumes cost equals number of different context-specific individuation schemas, each schema is disjoint, so has 0 entropy)

Context-general Individuation Schema:

Reliability $p(r : [\ s_{\textit{furn}} : \text{furn}(x)\ ] | r : [\ s_{\textit{furn-ind}} : {}^{*}\text{furn}_{\textit{Ind}_{gen}}(x)\ ])) \approx 1$

$p(r : [\ s_{\textit{furn-ind}} : {}^{*}\text{furn}_{\textit{Ind}_{gen}}(x)\ ]) | r : [\ s_{\textit{furn}} : \text{furn}(x)\ ]) \approx 1$

Individuation $-\sum_{v_i \in V} p(v_i | \text{furn}_{\textit{Ind}_{gen}}) \times \log p(v_i | \text{furn}_{\textit{Ind}_{gen}}) = \textit{high}$

- Neither of the two alternatives for individuation can satisfy both constraints (unless Cost is very high).

Furniture versus *Huonekalu* (‘(item of) furniture’, Finnish)

$$[\![\text{ huonekalu }]\!]^{c_i} = \lambda r : [x : \mathit{Stuff}]. \left[\begin{array}{ll} s_{\mathit{furn}} & : \text{furn}(r.x) \\ s_{\mathit{furn\text{-}ind}} & : \text{furn}_{\mathit{Ind}_{c_i}}(r.x) \end{array} \right]$$

- Prioritizes Individuation (minimizes entropy)
  - Disjoint P_{Ind} types at every context, so count

$$[\![\text{ furniture }]\!]^{c_i} = \lambda r : [x : \mathit{Stuff}]. \left[\begin{array}{ll} s_{\mathit{furn}} & : \text{furn}(r.x) \\ s_{\mathit{furn\text{-}ind}} & : \text{furn}_{\mathit{Ind}_{gen}}(r.x) \end{array} \right]$$

- Prioritizes Reliability (maximizes conditional probabilities)
  - Not-disjoint P_{Ind} type, so mass.

Constraints for mud-like nouns

Context-sensitive Individuation Schema:

Reliability $p(r : [\ s_{mud} : \text{mud}(x)\ ] | r : [\ s_{mud\text{-}ind} : {}^{*}\text{mud}_{Ind_{c_i}}(x)\ ])) = 1$

$p(r : [\ s_{mud\text{-}ind} : {}^{*}\text{mud}_{Ind_{c_i}}(x)\ ]) | r : [\ s_{mud} : \text{mud}(x)\ ]) = \text{lowish}$

(many context specific functions exclude some entities)

Individuation $-(\sum_{v_j \in V} \sum_{c_i \in C} p(v_j|\text{mud}_{Ind_{c_i}}) \times \log p(v_j|\text{mud}_{Ind_{c_i}})) + |C| = |C|$

(assumes cost equals number of different context-specific individuation schemas, each schema is disjoint, so has 0 entropy, very high number of possible schemas)

Context-general Individuation Schema:

Reliability $p(r : [\ s_{mud} : \text{mud}(x)\ ] | r : [\ s_{mud\text{-}ind} : {}^{*}\text{mud}_{Ind_{gen}}(x)\ ])) \approx 1$

$p(r : [\ s_{mud\text{-}ind} : {}^{*}\text{mud}_{Ind_{gen}}(x)\ ]) | r : [\ s_{mud} : \text{mud}(x)\ ]) \approx 1$

Individuation $-\sum_{v_i \in V} p(v_i|\text{mud}_{Ind_{gen}}) \times \log p(v_i|\text{mud}_{Ind_{gen}}) = \textit{high}$

- Countability of *mud* may depend on the cost of almost totally unconstrained context sensitivity.
  - Languages with measure systems: Very high cost
  - Languages without measure systems (e.g. Yudja): Lower cost

Blood versus *Apeta* ('blood' Yudja)

$$[\![\text{ apeta }]\!]^{c_i} = \lambda r : [x : \mathit{Stuff}].\left[\begin{array}{ll} s_{\mathit{blood}} & : \text{blood}(r.x) \\ s_{\mathit{blood\text{-}ind}} & : \text{blood}_{\mathit{Ind}_{c_i}}(r.x) \end{array}\right]$$

- Prioritizes Individuation by minimizing entropy (ONLY IF COST VALUE IS LOW ENOUGH)
  - Disjoint P_{Ind} types at every context, so count

$$[\![\text{ blood }]\!]^{c_i} = \lambda r : [x : \mathit{Stuff}].\left[\begin{array}{ll} s_{\mathit{blood}} & : \text{blood}(r.x) \\ s_{\mathit{blood\text{-}ind}} & : \text{blood}_{\mathit{Ind}_{gen}}(r.x) \end{array}\right]$$

- Prioritizes Reliability (maximizes conditional probabilities)
  - Not-disjoint P_{Ind} type, so mass.

Summary

- There are general cognitive pressures/constraints derived from a need for effective communication and learnability
- These can be characterized in probabilistic, information theoretic terms
- A highly suitable platform for this is a probabilistic mereological variant of TTR
- Variation in count/mass lexicalization patterns:
  - can be derived from an analysis of how these pressures either compete or act in unison.
- Context sensitivity:
  - Can be used to make communication more effective by providing a tailored individuation schema for the situation at hand.
  - Comes at a cost.
  - Almost unconstrained context sensitivity should only be expected if other strategies (e.g. measure phrases) are for some reason not available.

Selected References I

Gennaro Chierchia. Mass nouns, vagueness and semantic variation. Synthese, 174:99–149, 2010.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. A probabilistic rich type theory for semantic interpretation. Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics, 2014.

Robin Cooper, Simon Dobnik, Staffan Larsson, and Shalom Lappin. Probabilistic type theory and natural language semantics. LILT, 10(4), 2015.

Simon Kirby. "The Evolution of Meaning-Space Structure through Iterated Learning". In C. Lyon, C. Nehaniv, and A. Cangelosi, editors, *Emergence of Communication and Language*, pages 253–268. Springer, Verlag, London, 2007.

Simon Kirby and James Hurford. "The Emergence of Linguistic Structure: An overview of the Iterated Learning Model". In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language*, pages 121–148. Springer, Verlag, London, 2002.

Manfred Krifka. Nominal reference, temporal constitution and quantification in event semantics. In Renate Bartsch, Johan van Benthem, and Peter van Emde Boas, editors, Semantics and Contextual Expression, pages 75–115. Foris Publications, 1989. [In-Text: (Krifka 1989)].

Fred Landman. Count nouns – mass nouns – neat nouns – mess nouns. Baltic International Yearbook of Cognition, Logic and Communication, 6:1–67, 2011. doi: 10.4148/biyclc.v6i0.1579.

Selected References II

Fred Landman. Iceberg semantics for count nouns and mass nouns: The evidence from portions. The Baltic International Yearbook of Cognition Logic and Communication, 11, 2016.

Suzi Lima. All notional mass nouns are count nouns in Yudja. Proceedings of SALT, 24: 534–554, 2014.

S. Piantadosi, H. Tily, and E. Gibson. The communicative function of ambiguity in language. PNAS, 108(9):3526–3529, 2011.

Susan Rothstein. Counting and the mass/count distinction. Journal of Semantics, 27(3): 343–397, 2010. doi: 10.1093/jos/ffq007.

Yasutada Sudo. Countable nouns in japanese. Proceedings of WAFL 11, 2016. to appear.

Peter R. Sutton. Vagueness, Communication, and Semantic Information. PhD thesis, King's College London, 2013.

Peter R. Sutton. Towards a probabilistic semantics for vague adjectives. In Hans-Christian Schmitz and Henk Zeevat, editors, Bayesian Natural Language Semantics and Priagmatics, pages 221–246. Springer, 2016.