'Exoskeletal iceberg semantics': a best-of-both-worlds approach to the mass/count distinction

Hanna de Vries[1], George Tsoulas[1], Raffaella Folli[2], Agata Renans[2], Jacopo Romoli[2]

[1]University of York
[2]Ulster University

An intuitive approach to the mass/count distinction is to treat it as the grammatical counterpart to the human conceptual categories of substances and objects. These categories are prelinguistic and facilitate word learning in young children (Soja et al., 1991); conversely, children who have mastered the grammatical mass/count distinction use it in order to decide whether a new word refers to an object or a substance (Gordon, 1985; Bloom, 1990).

Lexical approaches to the mass/count distinction, that treat mass and count as lexical properties of nouns, capture this intuitive link between a noun's grammatical behaviour and human perception and categorisation of stuff and objects in the world. For example, the fact that English mass nouns cannot appear with numerals or be pluralised may be accounted for by assuming that they range over a 'stuff' domain that does not provide minimal elements, which means there is nothing to count and nothing to underpin a singular/plural distinction (Link, 1983). However, precisely because they have been designed to hardwire certain linguistic regularities into the lexical structure of nouns, such accounts are less well equipped to deal with the large amount of mass/count flexibility that is actually attested both within and between languages. Some languages have been claimed to treat all nouns as mass (Mandarin) or as count (Yudja; Lima 2014); some allow mass nouns to appear with plurals but not with numerals (Greek; Tsoulas 2009); many languages have 'object mass nouns' like *furniture* that are conceptually count but grammatically mass (Barner & Snedeker, 2005); and so on.

In contrast, constructionist approaches, that see the mass/count distinction as a purely morphosyntactic one, offer more tools to address crosslinguistic variation. Nothing in a noun's lexical meaning forces or blocks a certain syntax. Instead, meaning - including the distinction between object and stuff reference - is compositionally created by syntactic operations and, as such, subject to all kinds of parametric variation (for example, a plural feature may contribute different semantic or phonological effects depending on its position in the derivation). However, such morphosyntactic determinism leaves little room for conceptual factors, which makes it non-obvious how we could integrate (for example) the aforementioned acquisition data into our theory or account for wider crosslinguistic generalisations (e.g., no language expresses the concept CAT as a mass noun and WATER as a count noun).

In this talk, we will offer a new way of analysing the syntax and semantics of the mass/count distinction at the syntax-semantics interface, by synthesising the constructionist framework originating in Borer (2005) with the (lexicalist) 'iceberg semantics' proposed in Landman (2011, 2016). We believe that this synthesis has several conceptual and empirical advantages over existing frameworks. It combines the flexibility and morphosyntax-driven nature of constructionism with an explicit role for human conceptual categories such as INDIVIDUAL and SUBSTANCE. It allows

us to distinguish between different kinds of mass/count shifts and makes explicit why some of them are harder than others. Furthermore, it clarifies the distinction between countability and number neutrality, two distinct nominal properties that are often lumped together in lexical accounts that treat mass nouns as inherently plural, as well as the distinction between number neutrality and stuff reference.

1 Some new desiderata for a theory of the mass/count distinction

1.1 The right degree of flexibility

As is well known, mass and count are pretty flexible properties: in the right context and morphosyntactic environment, core mass nouns may be interpreted as count and vice versa[1] (Pelletier, 1975; Bunt, 1985, and many others). This flexibility is treated as a major argument in favour of constructionism by Borer (2005).

1. a. Three coffees, please. ('packaging')
   b. We offer three white wines. ('sorting')
   c. After the truck drove into my Halloween display, there was pumpkin all over the road. ('grinding')

In a lexical approach to the mass/count distinction, such flexibility is necessarily a matter of ambiguity or polysemy, that involves lexical reanalysis of the noun involved. On the other hand, in a constructionist approach, there is no such thing as lexical reanalysis; instead, all the mass/count shifts in (1) are treated as derivational.

The different approaches make different predictions. If mass/count shifts involve reanalysis, they should come with the associated processing cost; for example, processing sentence sentence (1a) should be more costly than processing "Three muffins, please". Moreover, grinding, packaging and sorting are all expected to be equally costly. On the other hand, if all count interpretations (whether 'packaged', 'sorted' or core count) involve exactly the same derivation and no lexical reanalysis, they should pattern together both in terms of grammatical behaviour and processing, and the same holds for mass interpretations (whether 'ground' or core mass).

However, the available evidence - though occasionally conflicting - seems to support a hybrid model, in which mass/count shifts differ in their availability and degree of complexity, sometimes involving reanalysis and sometimes syntactic derivation. In particular, the evidence supports (1) an asymmetry between packaging and sorting, and (2) an asymmetry between packaging and grinding. For example, Wiese & Maling (2005) show that packaging in Icelandic involves covert material while sorting doesn't; this was confirmed in an ERP study by Whelpton et al. (2014), who, however, also found that sorting was completely effortless (contra the lexicalist hypothesis). In Dutch (as well as in German, cf. Wiese & Maling 2005), the asymmetry between sorting and packaging is also reflected morphosyntactically. While sorting behaves as in English (i.e., the addition of any kind of count morphosyntax to a core mass noun turns it into a predicate over subkinds), unit interpretations of core mass nouns are extremely restricted in Dutch, and are not associated with count morphosyntax generally but with a special syntactic construction that seems to involve a covert classifier. Moreover, while some forms of mass/count flexibility in Dutch preserve the grammatical gender (e.g., stuff and subkind interpretations of core mass nouns always have the same

[1] We will use 'core mass' and 'core count' in a pretheoretical sense, to refer to nouns that are typically used to refer to stuff and to objects, respectively.

gender), others assign distinct genders to mass and count interpretations of the same noun (e.g., *draad* 'thread' is neuter in in its stuff interpretation and common when it refers to an individual thread). Finally, once we look beyond the restaurant and food-related contexts that are usually invoked in discussions of packaging, it turns out that packaging is fairly restricted in English as well, and seems to incorporate the (appropriately sized) container into the meaning of the resulting noun. Thus, it is extemely odd to say "I drank one wine" or even "I drank two wines" if I drank, from a beer glass, a quantity of wine corresponding to two standard units.

2. Asymmetry in mass/count flexibility

    - Sorting, grinding: easy; preserves gender; predictable meanin
    - Inherently flexible nouns: easy; mass/count may have different genders; predictable meaning
    - Packaging of core mass: difficult; unpredictable meaning / non-standard syntax or covert lexical material

All this suggests that some cases of mass/count flexibility are derivations involving the same noun, while others involve distinct nouns and lexical reanalysis.

1.2 Countability, number neutrality, stuff reference

Two related properties that distinguish mass nouns from count nouns, as well as different classes of count nouns, are countability (the ability to appear directly with numerals and count quantifiers like *each* or *several*) and number neutrality (including both singular and plural individuals, or, more generally, denoting a full mereology and not just a subset of it). Mass nouns like *sand* are uncountable (**each sand*) and number neutral (all possible quantities of sand including individual grains can be referred to as *sand*). English uninflected count nouns like *cat* are countable but not number neutral, while their pluralised counterparts are both countable and number neutral (it is generally agreed in the literature that plurals like *cats* include singular cats in their extension). In some languages, such as Hungarian (Farkas & de Swart, 2010), Indonesian (Chung, 2000), and Western Armenian, uninflected countable nouns are number neutral; English itself has a small class of countable number neutral nouns such as *staff* and *police*. In classifier languages like Mandarin, all nouns are uncountable number neutral.

3. Bezdig vaze.ts (Western Armenian; (Bale & Khanjian, 2008)
   child run.PAST(3, SG)
   'One or more children ran.'

4. Different classes of uninflected nouns across languages

	[+number neutral]	**[-number neutral]**
[+countable]	Count nouns in Hungarian, Western Armenian etc. In English: *police*, *staff*	Most uninflected count nouns in English and Dutch (*cat*, *bicycle* etc)
[-countable]	All nouns in Mandarin; (object) mass nouns in English	n/a

As the table shows, out of the 4 logically possible combinations, 3 are attested crosslinguistically while the 4th is not.[2] This means that the two properties can't be reduced to each

[2]Chierchia (2010) assumes that mass predicates originate as sets of uncountable atoms, which would make them [-countable,-number neutral], but he then needs to rely on a system of typeshifts in order to explain why mass nouns don't behave in this way (i.e., they always behave as if they range over full mereologies, not just minimal quantities).

other, but are not fully independent either. As it turns out, none of the lexical accounts in the Link/Chierchia/Landman tradition handles this in a satisfactory way. Chierchia (1998b,a), for example, treats them more or less like a package deal: all languages have uncountable number neutral nouns, and languages with a morphosyntactic singular/plural contrast additionally make countable non-neutral nouns available. Chierchia assumes that number neutral nouns are incompatible with plural number (because they already denote complete mereologies). This picture leaves no room for languages like Hungarian and Western Armenian in which uninflected count nouns are number neutral yet compatible with pluralisation. Chierchia (2010) treats non-countability and number neutrality as fully distinct properties where core mass nouns are concerned, but when it comes to the analysis of number neutrality and non-countability in object-referring nouns, he mostly falls back on his 1998 package deal approach, rooting both properties in reference to either kinds or maximal sums. This is somewhat counterintuitive since, as the table in (3) shows, it is precisely in the domain of countable nouns that number neutrality is optional, while uncountable nouns (including all core mass nouns) are also number neutral. Finally, in Landman (2011, 2016) the two properties are independent, but this also means that nothing rules out the existence of [-countable,-number neutral] nouns.

A constructionist account has the benefit of compositionality, which makes a pattern like that in table (3) fairly easy to model by assuming a sequence of grammatical operations. In Borer (2005), nouns start out as uncountable and number neutral; a dividing projection DivP optionally turns them countable; then a further NumP projection (which builds on DivP) optionally assigns them a non-neutral number feature (cf. Martí, 2017). However, unlike the lexical accounts, the exoskeletal account as put forward in Borer (2005) is unable to distinguish between uncountable number neutral reference to objects, and reference to stuff. Both are characterised by a lack of DivP and NumP and should therefore have the same range of meanings. However, as Cheng et al. (2008) observe, Mandarin bare object-referring nouns lack the 'ground' reading that characterises bare count nouns in English: they may be uncountable and number neutral but they still necessarily refer to objects and not to the stuff the objects are made of. Conversely, English bare count nouns are coerced to a stuff interpretation (as in (1c)); they cannot refer to objects in a number neutral way. And in Greek, bare count nouns are ambiguous between a number neutral and a ground interpretation.

Summing up, we need an account in which only [+countable] nouns can be [-number neutral], which does not reduce stuff reference to uncountable number neutrality.

2 'Exoskeletal iceberg semantics'

Our analysis of the mass/count distinction adapts the 'Iceberg Semantics' of Landman (2011,2016) to fit a constructionist or 'exoskeletal' syntax based on Borer (2005). Landman analyses nouns and NPs not as sets but as *i-sets* (iceberg sets): a pair $\langle \mathbf{body}, \mathbf{base} \rangle$ where both **body** and **base** are sets of mereological sums, and **base** *generates* **body**: all the elements in **body** are (sums of) elements in **base**, and **body** and **base** have the same maximal sum ($\mathbf{body} \subseteq {}^*\mathbf{base}$ and $\sqcup\mathbf{body} = \sqcup\mathbf{base}$, where * is semantic pluralisation à la Link 1983). The body expresses the extension of the noun in the usual way (e.g. **body**(CAT) = the set of individual cats; **body**(*CAT) = the (finite) set of all possible cat-sums; **body**(WATER) = the (possibly infinite) set of all possible water-sums). However, Landman does away with atomicity as a crucial distinction between mass and count predicates: what distinguishes individual cats from individual quantities of water is not that the former correspond to atoms and the latter to sums, but that the former can't overlap while the latter can. In other words, **body**(CAT) is *disjoint* while **body**(WATER) is overlapping.

The mass/count distinction, then, boils down to overlap versus disjointness: certain grammatical operations, such as counting, presuppose disjointness while others, such as measuring, do not. This is where the base comes in: since the body is affected by operations such as pluralisation, which turns it into an overlapping set, the mass or count status of a noun or NP is determined by the properties of its base, which is not affected in the same way. Both *cat* and *cats*, for example, have the same disjoint base CAT (the set of individual cats).

In our adaptation of Landman's system, rather than being provided lexically, both body and base are derived in syntax. We will assume:

5. a. Lexical items are classless *roots*, which we will model as *root concepts*: sets of mereologies. For example, the root $\sqrt{}$TRUMPET is associated with a 'physical' mereology of trumpet parts and sums, but also with an event mereology of trumpeting events and subevents, a more abstract mereology of trumpet sound, and whatever other ontological domains the word *trumpet* may draw its possible meanings from.

   b. Roots are turned into grammatical objects by an intersective filtering operation, which selects from the root concept only those sums that match a criterion imposed by the *filter head*. The available filter heads in the physical domain are INDIVIDUAL and PHYS_STUFF:

      – INDIVIDUAL: $\lambda C \lambda x$ [C(x) $\wedge$ INDIVIDUAL(x)].
      For example, INDIVIDUAL($\sqrt{}$STONE) is the set of all stone-sums that qualify as individuals.
      – PHYS_STUFF: $\lambda C \lambda x$ [C(x) $\wedge$ PHYS_STUFF(x)].
      For example, PHYS_STUFF($\sqrt{}$STONE) is the set of all stone-sums.

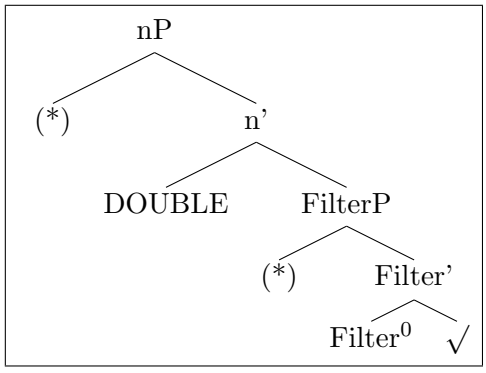
      We will call the result of applying a filter to a root concept a *filter set* or *f-set*. Note that f-sets derived with INDIVIDUAL are semantically singular while f-sets derived with PHYS_STUFF are complete mereologies, i.e. number neutral.

   c. F-sets are subsequently turned into i-sets by a 'doubling' operation: DOUBLE(F) = $\langle$F,F$\rangle$.

   d. Covert semantic pluralisation with * may obligatorily or optionally apply either immediately after derivation of the f-set (e.g. Mandarin), or after derivation of the i-set (e.g. Hungarian, Indonesian). (This is a locus of crosslinguistic variation.)

3 Back to our desiderata; conclusions

Implementation-wise, the present account is fully constructional; following Borer, there is no structural difference between 'mass roots' and 'count roots'. However, even though the distinction between objects and stuff is not built into the lexical roots themselves, it plays a large role in the filtering process. We cannot define a filter head such as INDIVIDUAL, which takes a root concept and returns the set of objects embodying that concept, without some kind of notion of what constitutes an individual. In a sense, this approach takes psycholinguistic findings on the prelinguistic nature of such notions more seriously than a lexical approach to the mass/count distinction does, since it relies on a truly independent notion of objecthood.



This does not mean that all filters are equally compatible with all roots, or that a concept like WATER can just as easily be embedded in a count DP as it may a mass one. Filtering out the individuals from a root concept is only possible if this root concept is in fact associated with individuals. For example, while it is technically possible to derive an f-set **water** by applying the filter INDIVIDUAL to the root $\sqrt{}$WATER, this f-set will end up empty since water-sums do not correspond to individuals as such. As we have seen, even in highly flexible languages like English, a count nP *wine* does not refer just to particular quantities of wine, but to a particular quantity of wine along with an appropriately sized container; it seems reasonable to treat this as a distinct root concept.

On the other hand, a root like $\sqrt{}$STONE can be selected by either an INDIVIDUAL or a PHYS_STUFF base with meaningful (i.e., non-empty) results in both cases, since there are some stone-sums that clearly match our prelinguistic objecthood criteria. Finally, since individuals are made out of physical stuff and kinds can be defined on the basis of its instantiations, we predict that it will always be possible to derive (at some point in the tree) an f-set over stuff or subkinds of individuals if INDIVIDUAL(C) is nonempty, and an f-set over subkinds of stuff if PHYS_STUFF(C) is nonempty. If we assume, in addition, that nominal gender in Dutch depends on the choice of filter head (cf. Arsenijevic 2016), the pattern in (3) neatly falls out.

The system is also able to distinguish number neutrality and countability while predicting the crosslinguistic lack of [-countable,-number neutral] nouns.

6. a. Uncountable number neutral stuff reference (universal): [DOUBLE [PHYS_STUFF [$\sqrt{}$WATER]]] → ⟨WATER, WATER⟩
   b. Uncountable number neutral object reference (Mandarin): [DOUBLE [* [INDIVIDUAL [$\sqrt{}$CAT]]]] → ⟨*CAT, *CAT⟩
   c. Countable number neutral object reference (Hungarian): [* [DOUBLE [INDIVIDUAL [$\sqrt{}$CAT]]]] → ⟨*CAT, CAT⟩
   d. Countable singular object reference (English): [DOUBLE [INDIVIDUAL [$\sqrt{}$CAT]]]] = ⟨CAT, CAT⟩
   e. Uncountable singular object reference = ⟨CAT, *CAT⟩ = underivable

And unlike Borer (2005), the system formally distinguishes uncountable number neutral object reference from stuff reference. In Landman's terminology, the former are *neat mass* (they have an overlapping base that is generated by a disjoint set) while the latter are *mess mass* (their base is not generated by a disjoint set).

References

Bale, A., & Khanjian, H. (2008). Classifiers and number marking. In T. Friedman & S. Ito (Eds.), *Proceedings of SALT XVIII* (p. 73-98).

Barner, D., & Snedeker, J. (2005). Evidence that mass nouns count. *Cognition*, *97*, 41-66.

Bloom, P. (1990). *Semantic structure and language development* (Unpublished doctoral dissertation). MIT.

Borer, H. (2005). *In name only (structuring sense vol. i)*. Oxford: Oxford University Press.

Bunt, H. (1985). *Mass terms and model-theoretic semantics.* Cambridge: Cambridge University Press.

Cheng, L., Doetjes, J., & Sybesma, R. (2008). How universal is the Universal Grinder? *Linguistics in the Netherlands*, *25*, 50-62.

Chierchia, G. (1998a). Plurality of mass nouns and the notion of 'semantic parameter'. In S. Rothstein (Ed.), *Events and grammar.* Dordrecht: Kluwer.

Chierchia, G. (1998b). Reference to kinds across languages. *Natural Language Semantics*, *6*, 339-405.

Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthese*, *174*, 99–149.

Chung, S. (2000). On reference to kinds in Indonesian. *Natural Language Semantics*, *8*, 157-171.

Farkas, D., & de Swart, H. (2010). The semantics and pragmatics of plurals. *Semantics and Pragmatics*, *3*.

Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition*, *20*, 209-242.

Landman, F. (2011). Count nouns - mass nouns - neat nouns - mess nouns. In B. H. Partee, M. Glanzberg, & J. Škilters (Eds.), *Formal semantics and pragmatics: discourse, context and models. the Baltic international yearbook of cognition, logic and communication, vol. 6.* Manhattan (KS): New Prairie Press.

Landman, F. (2016). Iceberg semantics for count nouns and mass nouns: classifiers, measures and portions. In S. Rothstein & J. Škilters (Eds.), *Number: Cognitive, semantic and crosslinguistic approaches. the Baltic international yearbook of cognition, logic and communication, vol. 11.* Manhattan (KS): New Prairie Press.

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

Link, G. (1983). The logical analysis of plurals and mass terms: a lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and interpretation of language* (p. 302-323). Berlin: De Gruyter.

Martí, L. (2017). *Inclusive plurals and the theory of number.* (unpublished ms., Queen Mary University of London)

Pelletier, J. (1975). Non-singular reference: some preliminaries. *Philosophia*, *5*, 451-465.

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

Tsoulas, G. (2009). On the grammar of number and mass terms in Greek. In C. Halpert, J. Hartman, & D. Hill (Eds.), *Proceedings of the 2007 workshop in Greek syntax and semantics at MIT.* MIT Working Papers in Linguistics.

Whelpton, M., Trotter, D., þórhalla Guðmundsdóttir Beck, Anderson, C., Maling, J., Durvasula, K., & Beretta, A. (2014). Portions and sorts in Icelandic: An ERP study. *Brain and Language*, *136*, 44-57.

Wiese, H., & Maling, J. (2005). Beers, Kaffi, and Schnaps: Different grammatical options for restaurant talk coercions in three Germanic languages. *Journal of Germanic Linguistics*, *17*, 1-38.