Extension, ontological type, and morphosyntactic class: Three ingredients of countability

David Clausen, Alex Djalali, Scott Grimm, Sven Lauer, Tania Rojas-Esponda and Beth Levin
Stanford University
Extended abstract — May 2010

The literature presents two perspectives on the status of nouns naming certain entities as mass or count. On one view, this choice is predetermined by the nature of the entity named. On the other, this choice is arbitrary, though there may be some regularities or at least tendencies in lexicalization as mass or count. Proponents of arbitrariness cite doublets like *mail* and *letters*, *leaves* and *foliage*, or *rice* and *lentils*, as well as the pervasiveness of Universal Grinder and Packager effects (Pelletier 1979), which suggest that mass/count status is not tied to a lexical item. However, proponents of the alternate view note that even purportedly arbitrary minimal pairs are less arbitrary than they appear. For instance, Zwicky (1997) points out that although *petunias* and *ice plant* can both be used to cover areas of ground in a garden, the former is count as it is easily divisible into individual plants, while the latter is mass, precisely because this is not the case: a single plant covers an extended area, while separate plants are hard to distinguish. In this paper we reconcile the two views: we argue that there are cognitive–perceptual and cultural factors which generally determine whether the noun naming an entity will be morphosyntactically mass or count, while recognizing that these factors apply to conceptualizations of entities, so that entities that are open to multiple conceptualizations might indeed be lexicalized as either a mass noun or a count noun. Thus, there is some apparent arbitrariness in classification, but it is constrained by precisely the factors which give rise to mass/count classification in the first place. We also argue that there is a seeming residue of arbitrariness in mass/count assignment that arises for historical reasons, for instance, due to phonological change or meaning drift, as proposed most explicitly by Wisniewski (2010).

A key property of our account is recognizing that the process of determining whether a noun is mass or count morphosyntactically is not based on the properties of the real world referent of that noun, i.e. its extension, but on a conceptualization of that referent in terms of its status on a scale of individuation, which encompasses a richer ontology than simply mass vs. count, ranging from substances to aggregates to collectives to individuals. This intermediate conceptual representation of an entity, which is determined by perceptual and cognitive factors interacting with cultural ones, allows for systematicity in the assignment of a morphosyntactic classification to a noun depending on its position on the individuation scale. Available morphosyntactic classifications vary across languages and are reflected in number marking systems. Drawing on data from several languages, we argue that while English shows a two-way morphosyntactic classification (mass vs. count), some other languages make additional distinctions. These classification systems all respect the individuation scale, with a language's morphosyntactic classes picking out contiguous portions of this scale. Languages always provide distinct morphosyntactic treatments of 'substances' and 'things'—the endpoints of the scale; entities falling between the endpoints are either assimilated to one endpoint morphosyntactically or put in additional classes.

The picture proposed here contrasts with previous research on the semantics of mass and count, which while proposing analyses which differ in various respects, assumes a direct relation between the extensional realm and morphosyntactic class (e.g. Quine 1960, Link 1983, Chierchia 1998, 2010, Krifka 2007). On our picture the mapping is not direct, but involves three levels, adding a level of conceptualization, which mediates between the extensional realm and the morphosyntax. We will argue that we can better account for the larger empirical realm, including some of the key properties and puzzles raised in prior work, in this way. Such a picture is not new: in his work, Manfred Bierwisch (e.g. 1983) proposes that a conceptual system mediated between language and the real-world entities it referred, though based on rather different phenomena.

1 Introducing the conceptual level

A fruitful line of research on the semantics of number has proposed analyses which, in various ways, explore a direct relation between the extensional realm and the morphosyntax of number marking (e.g. Link 1983, Chierchia 1998, Krifka 2007). This work, however, must confront a set of difficult examples where two terms, one count and one mass, ostensibly overlap in their reference. Typical doublets include *letters/mail*, *coins/change*, *leaves/foliage*, *tiles/tiling*, and *curtains/drapery*.[1] These examples pose a challenge: if a theory of mass/count is based on the external world, how can one entity give rise to two word forms belonging to distinct morphosyntactic classes? Only considering the extensional and morphosyntactic levels forces the conclusion that the mapping between entities and the morphosyntax is, at least partially, arbitrary—a claim explicitly reiterated in the literature (e.g. Chierchia 1998, Borer 2005). The approach advocated here disposes of a different option: while these doublets may often be referentially interchangeable, they are also conceptually distinct.

Over and above permitting arbitrariness in the grammar, which from many perspectives is theoretically undesirable, considering these doublets from a purely extensional view breaks down upon closer inspection. For instance, *mail* describes the set of objects that one receives via the post, which may include letters, but may also include magazines, packages, postcards, and the like. *Letters* names a far narrower class of entities, which, in addition, need not actually have been mailed. Similarly, *change* designates the leftover money received after a sale, which may include but certainly is not limited to coins, whereas *coins* refers to a narrower class of entities, whether or not they have been received after a particular financial transaction. Additionally, when observing a tree, a speaker may freely choose to talk about its leaves or its foliage. In contrast, when discussing leaves raked into a pile, *leaves* is appropriate, whereas *foliage* is not. While these doublets often overlap in extension, they actually name distinct perspectives on the entity at hand.

What are these distinct naming perspectives picking up on? Or to rephrase this, is there something that sets that member of each pair with mass morphosyntax apart from the member with count morphosyntax? *Mail* and *change* have in common that they name sets of entities that participate together in an event or arise together as a result of this event: *change* is a result of a monetary transaction, while *mail* names a set of entities that travel through the postal system. In contrast, the noun *foliage* provides a holistic perspective on a coccurring, contiguous and normally connected aggregate of leaves. Thus, *foliage* emphasizes the collectivity and the interconnectedness of the leaves with one another rather than the individual leaves. A comparable noun is *plumage*, which contrasts with *feathers*, in referring to the ensemble of feathers on a bird, but not say to the contents of a down pillow, which may be referred to as *feathers*.

The use of a noun conveys more than its extension. The above instances demonstrate nouns making reference to spatial and temporal contiguity, *mail* and *foliage* respectively, as well as to similarity in form (*foliage, plumage*) and function (*mail, change*) of the members of the aggregate. In her exploration of lexical semantics of mass and count nouns, Wierzbicka (1985) recognizes the importance of conceptual and cultural factors in determining a noun's classification; among these she points to the importance of mode of interaction with the relevant entity, as well as the ease of distinguishability of any constituent elements, a notion that is influenced by their size and contiguity. Experimental support for the reality of these factors is provided by Middleton et al. (2004). The doublets discussed above are remarkable, not because they illustrate supposed arbitrariness, but because they demonstrate the availability of multiple perspectives on certain entities, which allows them to be lexicalized as either morphosyntactically mass or count.

---

[1]We ignore superordinate terms, such as *furniture* or *silverware*, though they are often included in discussions of mass nouns. These are very different from the typical entities lexicalized as morphosyntactically mass, as reflected in the label 'object-level mass' nouns for superordinates used by Barner and Bale 2009. The reasons that mass morphosyntax is recruited for their expression in many languages deserves further study.

2 The structure of the conceptual level: Evidence from morphosyntax

Arbitrariness in the morphosyntactic mass or count status of nouns in two languages with the same extension could also arise even when the perspective on the extension is the same, since it could reflect difference among languages as to how they divide up the same conceptual space into morphosyntactic categories. In particular, this situation arises because some languages make more morphosyntactic distinctions than English does. We will make this point, and at the same time, we will show that there is structure to the conceptual space that emerges from the attested morphosyntactic options, which we will describe in terms of an individuation scale; see section 3.

We begin with English. As many have noted, in English liquids such as *water* and *oil* and substances such as *granite* or *wood* have mass morphosyntax, while individuated things such as *apple* or *pencil* have count morphosyntax (cf. Chierchia's 2010 Mapping Property). That is, the former occur only in the unmarked form and have no contrasting plural form, while the latter have both forms; further, only the latter can be counted (*three apples/*wood*). Also treated as having mass morphosyntax are granular aggregates, such as *flour*, *rice*, *sand*, and *sugar*. English, then, makes a two-way split among nouns in terms of their morphosyntactic type.[2]

In contrast, Welsh demonstrates a tripartite division based on its number marking possibilities: singular/plural, collective/singulative and mass (Stolz 2001). The singular/plural class covers typical animates and individuals, while the mass class covers canonical mass entities such as liquids. The domain of the collective–singulative marking pattern includes collective aggregates such as small animals and insects, vegetables/grains/fruits, inherently plural body parts (e.g. *ribs*), as well as granular aggregates (e.g. *turf*, *sand*). The singular/plural and collective/singulative classes differ in the direction of markedness: for the former, the singular is morphologically unmarked, for the latter, the collective is unmarked and the singulative is marked.

Dagaare, a Gur language spoken in Ghana and Burkina Faso, permits a division into four number marking classes. At the heart of Dagaare's nominal system is a rare type of number marking known as 'inverse' or 'polarity' number marking (Baerman 2007). The marker *–ri* marks the plural for some nouns (*child, dog*), while for other nouns (*seeds, insects*) for which the plural is not marked, *-ri* marks the singular. Grimm (2009) shows that nouns for which *–ri* marks the plural align with typical individuals, while nouns for which *–ri* marks the singular align with typical collective aggregates, such as vegetation, insects, or inherently plural body parts (cf. Acquaviva's 2008 'inherent plurals'). In addition to this distinction, Dagaare also possesses a singulative marker *–ruu*, which is primarily restricted to granular mass terms (*pepper*, *straw*, *grass*), designating 'a piece of'. Core mass terms, such as liquids, are yet a separate category, showing no number distinctions.

Pulling these observations together, we find that entities of a given ontological type may receive distinct treatments in the three languages. The table below, which aligns the morphosyntactic patterns of the three languages, makes this point clearly.

Language	liquids/ substances	granular aggregates	collective aggregates	individual objects
English	0	0	0/Plural (*–s*)	0/Plural (*–s*)
Welsh	0	0/Singulative (*–yn*)	0/Singulative (*–yn*)	0/Plural (*–od*)
Dagaare	0	0/Singulative (*–ruu*)	0/Singular (*–ri*)	0/Plural (*–ri*)

For example, collective aggregates are morphosyntactically assimilated to count nouns in English,

---

[2]This English picture is somewhat simplified. English has a morphosyntactic class of 'pluralia tantum', nouns occuring in a plural form with no contrasting singular, yet unlike typical plurals such nouns cannot be counted (e.g. #*sud*, #*three suds*). These include some granular aggregates such as *oats* and *curds*, as well as representatives of smaller lexical semantic classes. We ignore this complication, simply pointing out that Koptjevskaja-Tamm and Wälchli's (2001) cross-linguistic study identifies recurring semantic classes among pluralia tantum, which suggest that they could be accommodated within a refined theory of ontological types; see also Wisniewski (2010).

but not in Welsh and Dagaare. More important, although the three languages impose distinct morphosyntactic categories on the ontological individuation types, each does this in a way that suggests the ontological types form a scale (Grimm 2010). A particular ontological type can be assigned to a unique morphosyntactic class (e.g. Dagaare liquids); otherwise, it is associated with the same class as the type to its left, right, or both. The morphosyntactic classes respect the structure of the scale: no morphosyntactic class spans two individuation types that are not contiguous on the scale. We now suggest, following Grimm (2010), that this scale can be viewed as reflecting individuation.

3   A scale of individuation

We consider the arrangement of ontological types given in the table to be a scale of individuation in accordance with research from various fields which claim that the categorization of entities into mass or count morphosyntactic classes is due to the individuation level of the relevant entity.

(1) liquids/substances < granular aggregates < collective aggregates < individual objects

Individuation is a cover term for conceptual and perceptual factors which characterize the propensity for an entity to appear as an individual unit.

A variety of perspectives in the philosophical, linguistic and psycholinguistic literatures (e.g. Quine 1960, Mufwene 1980, Wierzbicka 1985, Bloom 1994, Zwicky 1997, Middleton et al. 2004) have isolated various factors which are the result of perceptual and cognitive processes, including ease of distinguishability of elements, size of elements, spatial and/or temporal contiguity among elements, as well as canonical mode of interaction. These factors are all relevant to whether entities are classified as, for instance, individual objects or aggregates.

The poles of the scale are liquids/substances vs. individual objects, a fundamental opposition which appears early in child development (Soja et al. 1991). This opposition in turn corresponds to minimally and maximally individuated entities: the minimal elements of liquid and substances are continuous and not distinguishable: one does not interact with individual elements at all; for individual objects, the inverse holds for all the relevant properties. Granular aggregates have individuation properties similar to liquids, and tend to pattern with them morphosyntactically. Granular aggregates often have minimal elements (*a grain of sand*), which are small, not easily distinguishable, and one does not canonically interact with them. Collective aggregates represent an intermediate category: the minimal elements are more accessible, typically being larger than for granular aggregates; interaction with their minimal elements is also more frequent. The scale that emerges from examining these morphosyntactic classes sheds light on the types of factors that have been appealed to in the literature on conceptualization. Finally, while we discuss the ontological types as forming a scale of individuation, we leave open the possibility that the interrelations between the types is better described not as a single scale, but as the result of several interacting dimensions.

4   Further ramifications

Arbitrariness revisited: This view of individuation yields insight into the doublets discussed in section 1. The perceptual and conceptual factors underlying individuation allow the entities described by *letters* and *mail* to be categorized in two different ways: as individuals or as a granular aggregate, a set of entitites functionally related by their journey through the postal system. These divergent mappings from the extensional to the conceptual level are illustrated in the figure:

⟦mail⟧   ⟦letters⟧
   | |
liquids < granular aggregates < collective aggregates < individual objects

4

A second type of mismatch often taken as evidence of arbitrariness at a cross-linguistic level involves lexical items which surface in different morphosyntactic classes in different languages. For instance, *cherries* is realized with singular/plural marking in English, but collective/singulative marking in Welsh. As these morphosyntactic categories align differently with the individuation scale, as delineated in the earlier table, such cross-linguistic variation is expected.

'Universal' Grinder and Packager effects: Our approach makes predictions about operations which apparently invert the morphological class of a noun. It has been claimed that every count noun, given the right context, can have a mass interpretation, as exemplified by *There is apple in the salad*, a conversion known as the Universal Grinder. The 'inverse' operation, the Universal Packager, results in count interpretations for typically mass nouns, as in *Three beers, please*. Such data are often taken as evidence that a noun's status is not tied to the lexical item itself but is necessarily computed at the NP level (Allan 1980, Bunt 1985).

If these effects were truly universal, they should apply uniformly across all nouns, yet these operations are restricted, as also noted by Chierchia (2010). Further, the entity itself contributes more to its classification than typically recognized. Packaging is largely restricted to those nouns whose referents are already associated with conventionalized units of packaging: (2) is obviously ungrammatical under the interpretation that rices are packaged into measurable units. Grinding is also restricted. In particular, it is difficult to 'grind' highly individual objects, especially artifacts, as in (3). Although foodstuffs generally permit grinding, there are exceptions, as shown for *pea* in (4).

(2) #Rices adorn the altar.
(3) #There is mug/toaster on the table.
(4) #Would you care for some more pea? (Fillmore 1989: 49)

These shifts involve not only the extensional and morphosyntactic levels, but also the conceptual level. The phrase *apple in the salad* involves bits of apple, a re-conceptualization which picks out an entity extensionally distinct from the individual object *apple*, related to the 'granular aggregate' position on the scale. According to the picture developed here, if the extensional shift can be successfully related to the needed ontological type, then the change in the noun's morphosyntax class from count to mass follows directly. Yet, shifts require considerable context to make them approach plausibility and, thus, be successful, and as the limiting cases show, the link between the re-conceptualized entity and the ontological class cannot be always accommodated.

Apparent exceptions: The morphosyntactic mass/count status of some nouns does not match their individuation status, but as Wisniewski (2010) notes, these exceptions have explanations. Thus, *aspirin* is often used with mass morphosyntax (e.g. *a bottle of aspirin*), though it comes in pill form; the reason is that originally aspirin was a powder, and as a granular aggregate appropriately had mass morphosyntax. The morphosyntax is now catching up, with count uses of *aspirin* increasingly found (e.g. *Take two aspirins before bed*). A slightly different example is *pea*, a count noun, backformed from the mass noun *pease*, which referred to a cooked porridge made of peas. Here the presence of the final /s/ allowed a reinterpretation of the form as a plural, which required its referent to be an individual object—a single pea. Finally, *chad*, originally a mass noun applied to the detritus from computer punch cards, became a count noun at the time of the US presidential election in 2000, when each individual bit of paper punched out of a hole—a proxy for a vote—counted; this shift reflects a change in mode of interaction. None of these instances truly turns out to challenge the proposed larger picture once the conditions that govern a noun's individuation status are understood.

Concluding remarks: The larger picture sketched here recognizing three levels—extensional, morphosyntactic, and in particular conceptual—helps make sense of some of the empirical challenges that have confronted prior research on the mass/count distinction, as well as the cross-linguistic diversity of mass/count-related morphosyntactic distinctions. As the brief discussion of ramifications suggests, we believe that this picture opens fruitful avenues for further research.

References

Acquaviva, P. (2008) *Lexical Plurals: A Morphosemantic Approach*, Oxford University Press, Oxford.

Allan, K. (1980) "Nouns and Countability", *Language* 56, 541-567.

Baerman, M. (2007) "Morphological Reversals", *Journal of Linguistics*, 43, 33-51.

Bale, A.C. and D. Barner (2009) "The Interpretation of Functional Heads: Using Comparatives to Explore the Mass/Count Distinction ", *Journal of Semantics* 26, 217-252.

Bierwisch, M. (1983) "Semantische und konzeptuelle Representation lexikalischer Einheiten", W. Motsch and R. Ruzicka, eds., *Untersuchungen zur Semantik*, Akademie-Verlag, Berlin, 61-99.

Bloom, P. (1994) "Possible Names: The Role of Syntax-Semantics Mappings in the Acquisition of Nominals", *Lingua* 92, 297-329.

Borer, H. (2005) *Structuring Sense I: In Name Only*, Oxford University Press, Oxford.

Bunt, H.C. (1985) *Mass Terms and Model-Theoretic Semantics*, Cambridge University Press.

Chierchia, G. (1998) "Plurality of Mass Nouns and the Notion of 'Semantic Parameter'", in S. Rothstein, ed., *Events and Grammar*, Kluwer, Dordrecht, 53-103.

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

Fillmore, C.J. (1989) "Grammatical Construction Theory and the Familiar Dichotomies", in R. Dietrich and C.F. Graumann, eds., *Language Processing in Social Context*, North-Holland, Amsterdam, 17-38.

Grimm, S. (2009) "Number Marking and Individuation: A View from Dagaare", unpublished ms., Stanford University.

Grimm, S. (2010) "Number and Individuation", dissertation proposal, Stanford University.

Koptjevskaja-Tamm, M. and B. Wälchli (2001) "The Circum-Baltic Languages: An Areal-Typological Approach", in Ö. Dahl and M. Koptjevskaja-Tamm, eds., *Circum-Baltic Languages 2: Grammar and Typology*, John Benjamins, Amsterdam, 615-750.

Krifka, M. (2007) "Masses and Countables: Cognitive and Linguistic Factors", slides, The Syntax and Semantics of Measurement Workshop, CASTL, University of Tromsø.

Link, G. (1983) "The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach", in R. Bäuerle, C. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language*, Walter de Gruyter, Berlin, 302-323.

Middleton, E.L., E.J. Wisniewski, K.A. Trindel, and M. Imai (2004) "Separating the Chaff from the Oats: Evidence for a Conceptual Distinction between Count Noun and Mass Noun Aggregates", *Journal of Memory and Language* 50, 371-394.

Mufwene, S. (1980) "Number Countability, and Markedness in Lingala *li–/ma–* Noun Class", *Linguistics* 18, 1019-1052.

Pelletier, F.J. (1979) "Non-Singular Reference: Some Preliminaries", in F.J. Pelletier, ed., *Mass Terms*, Reidel, Dordrecht, 1-14.

Soja, N.N., S. Carey, and E.S. Spelke (1991) "Ontological Categories Guide Young Children's Inductions of Word Meaning: Object Terms and Substance Terms", *Cognition* 38, 179-211.

Stolz, T. (2001) "Singulative-Collective: Natural Morphology and Stable Classes in Welsh Number Inflexion on Nouns", *Sprachtypologie und Universalienforschung* 54, 52-76.

Wierzbicka, A. (1988) "Oats and Wheat: The Fallacy of Arbitrariness", in A. Wierzbicka, *The Semantics of Grammar*, John Benjamins, Amsterdam, 499-560.

Wisniewski, E.J. (2010) "On Using Count Nouns, Mass Nouns, and Pluralia Tantum: What Counts", in F.J. Pelletier, ed., *Kinds, Things and Stuff*, Oxford University Press, Oxford, 166-190.

Zwicky, A.M. (1997) "Count versus Mass in English: How to Talk about Plants", unpublished ms., Stanford University and Ohio State University.