Underquantification: an application to mass terms

Aurelie Herbelot and Ann Copestake

University of Cambridge, Computer Laboratory,
J. J. Thomson Avenue, Cambridge, United Kingdom
ah433@cam.ac.uk,aac10@cam.ac.uk

1 Introduction

Mass terms are linguistically defined according to the following features: they do not occur with cardinals (**six rices*), they are not pluralisable (**rices*), they are quantified over using classifiers (*six grains of rice*), they accept determiners that do not occur with count nouns and, conversely, do not combine with some count noun articles (*much rice/*many rices*).

The particularities of mass terms with regard to quantification gives them an ontological status that also differentiates them from count nouns. In particular, the notion of individuation is not straightforward: we cannot retrieve the minimal constituents of an instance of *water* as we do for an instance of *cats* (by counting the number of individual cats in the group under consideration). Still, it is desirable to consider count and mass nouns in the same framework. That is, given the phrases *two kilograms of rice* and *two kilograms of apples*, we would expect the 'two kilograms' to have the same formal representation, i.e. to behave in a unique way whether quantifying over rice or over apples.

This paper claims that all subject noun phrases can be given a unique underspecified formalisation in terms of quantification (which we term **underquantification**). The claim has consequences for our ontological and linguistic understanding of mass terms. We suggest that it is possible to categorise mass terms in relation to **divisibility**: whether we can find elementary parts for them which are the same nature as their whole (a drop of water is water, so water is divisible; it is perhaps less clear what a part of progress is). We argue for an underspecified logical form which can be applied to both divisible and indivisible noun phrases (although the domain of quantification is not the same in both cases), satisfying our requirement for a unique quantificational formalisation of count nouns and mass terms.

2 The underquantification phenomenon

In Herbelot (to appear), we gave an account of ambiguously quantified noun phrases in terms of underspecification. We argued that sentences such as 1 and 2 feature an underspecified quantifier which must be resolved to get at the quantification semantics of the subject noun phrase.

1. Cats are mammals. (All cats...)
2. The reporters asked questions at the end of the press conference. (Some of the reporters...)

The underspecified quantifier has the following general form:

3. $X = \sigma^* x\, P'(x) \wedge \exists Y [Y \prod X \wedge Q(Y) \wedge \textit{quant-constraint}(X, Y)]$

The notation is borrowed from Link (1983) and assumes a lattice view of plurals. The sign σ is the sum operator. $\sigma^* xPx$ represents the **proper sum** of Ps, that is, the supremum of all objects that are proper plural predicates of P. The product sign $\prod$ expresses an **individual-part** relation. The $\cdot$ sign in combination with $\prod$ indicates a relation of **atomic part**. X denotes the Nbar referent for the noun phrase under consideration while Y denotes the whole NP referent. The 'quant-constraint' is a set relation that ensures the correct cardinality of Y for various quantifiers[1] and the predicate Q applies distributively or collectively depending on the semantics of the sentence.

The application of the underspecified quantifier to sentence 2 results in the following formalisation:

4. $X = \sigma^* x\, reporterAtPressConference'(x) \wedge \exists Y[Y \prod X \wedge \forall z[z \cdot \prod Y$
$\rightarrow askQuestions'(z)] \wedge (0 < |Y| < |X - Y|)]$

We can paraphrase the formalisation by saying that there is a set X of reporters at the press conference and there is a set Y which is a part of X and for every instance z in Y, z asked questions. The cardinality of Y is positive and less than the cardinality of $X - Y$ (this upper bound has for effect to distinguish *some* from *most* – due to space restrictions, we will not discuss our partitioning of the quantificational space; it is not important for this paper). A similar representation could be given of Sentence 1 where $Y = X$.

Our main claim is that every subject noun phrase can be formalised in terms of quantification (this includes generic sentences featuring a so-called 'kind reading'). We have demonstrated the validity of this claim by performing an annotation task where naturally occurring subject NPs were quantified by humans as *one*, *some*, *most* or *all* statements (see Herbelot and Copestake, 2010). Our empirical study included some mass terms but, their frequency being relatively low in our corpus, we could not draw general conclusions with regard to their quantificational behaviour. This paper takes a closer, theoretical look at the issues encountered when quantifying mass terms.

3 Mass terms and quantification

The formalisation of Section 2 is available to mass terms as long as they can be represented as lattices of 'non-overlapping parts'. See the examples below where we temporarily follow Link (1983)in his use of special operators for mass terms: μ is the **material fusion** operator and $\top$ denotes a **material part** relation. We posit that in combination with $\cdot$, $\top$ denotes an atomic part relation; this is not standard Link notation and we will come back later to the issue of whether we can assume the existence of atomic part in mass terms.

5. (Some) Water was dripping through the ceiling.
$X = \mu^* x\, water'(x) \wedge \exists Y[Y \top X \wedge \forall z[z \cdot \top Y \rightarrow dripThroughCeiling'(z)] \wedge$
$(0 < |Y| < |X - Y|)]$

The formalisation in 5 can be paraphrased as follows: X is the fusion of all water in the world and there is a Y which is a material part of X and for every minimal, or **atomic** part z of Y (let's say, a drop), z drips through the ceiling. Y is *some* of X.

The obvious problem of such an analysis is to define what counts as a 'part', and to elucidate whether all mass terms possess such parts. Consider, for instance, the following:

6. Most love is expressed through actions.

[1] Note that we follow Landman (1989) in his interpretation of pluralities as sets.

where it is not clear what 'unit' we are quantifying over.

Further, we can ask whether the relations between parts of mass terms should be formalised in a different way from the relations between count individuals (see the use of the material fusion operator μ against the sum operator σ in Link's treatment). This goes against the intuition that a single ontological framework is available to both count and mass terms. We discuss these questions in the next section.

4 Parts and atomic parts: what are they? (And do we really care?)

4.1 Linguistic representation: correlation with psychological evidence

It has been long debated whether natural language speakers conceptualise mass terms as a function of atomic parts and, regardless of the answer to this question, whether a linguistic formalisation of mass terms should include such parts or not. Link (1983) assumes a non-atomic view of mass terms: it is possible to talk of parts of water, but not of the smallest possible part of water (so each part of water is infinitely divisible). In contrast, Chierchia (1998) supports a formalisation which includes atomic parts. His representation of mass terms is a lattice of parts, where each part can be seen as a plurality of atomic parts.

It is difficult to prioritise either Link's or Chierchia's approach on the basis of psychology. Carey (1991), as well as Smith et al (2005), showed that, when presented with a piece of Styrofoam and asked whether the Styrofoam could be infinitely split into two parts, children under the age of 12 replied that there would be a point where there would be no Styrofoam left. This (incorrect) view of matter supports the existence of atomic parts in those children's conceptual frame. Older children, who have learnt a continuous view of matter, answer that the Styrofoam can be split ad infinitum. This fits with the non-atomic representation advocated by Link. Further still, children/adults that have been introduced to a particulate theory of matter correctly identify the molecular level as the point at which a substance cannot be divided any further and remain the same substance (Tirosh and Stavy, 1996).

4.2 The nature of atomic parts

If the notion of an 'atomic part' of water is easily conceivable (as a drop, or a molecule of H_2O), some instances of mass terms resist such straightforward interpretation. It is for example conceptually difficult to imagine what a part of 'freedom' or 'progress' might be, let alone an atomic part. And still, we can assume that humans would be able to answer questions about quantities involving those terms. Consider the following:

7. Who produces more rice? India or England?
8. Which group has made more progress? Dr Smith's or Dr Jones'?
9. Which report contains more information? The one-page report or the 500 pages report?

Although it may be impossible to visualise the whole rice produced by India, and there is no direct perceptual referent for the nouns *progress* and *information*, we can assume that humans would be able to answer those questions. We hypothesise that quantification still happens, but over items that are contingent to the actual mass terms in the questions. For instance, someone might decide that progress is measured in terms of publications and simply count how many articles the groups of Dr Smith and Dr Jones have published in the last year. Bloomfield (1933) observed this effect in abstract terms, saying that things denoted by such terms 'exist only as the demeanor (quality, action, relation) of other objects' – accordingly, he viewed abstract nouns and mass nouns as belonging to two separate categories.

Because of their linguistic behaviour, we wish to keep bare singular abstract nouns in the class of mass terms. Doing so, however, we must forego the idea of identifiable parts for them. Instead, we suggest that we can quantify over occurrences of their denotation.

Following Bloomfield's account, we can say that every time an abstract term is bound to an object, we observe an **eventuality** of the abstract concept. Such eventualities are frequent: abstract terms are often grounded in an individual or a situation (*the information in this report*, *Mary's love for Sally*). Eventualities can be quantified over, and they allow us to process statements such as *Most love is expressed through actions*, where we presumably consider instances of people expressing love for others. We further argue that the accepted use of quantifiers such as *most* in front of mass terms invites an analysis that involves atomic parts: assuming that these quantifiers can be represented as relations over the cardinality of sets, we need minimal units to 'count'.

So atomic parts can take various forms. Note that we are not interested in what those parts actually are, that is, at which level humans would stop when recursively dividing an instance of a mass concept. We have already seen that both drops and H_2O molecules may be taken as acceptable atomic parts for water, for instance. Further, when considering more abstract terms, it is even desirable to talk of eventualities rather than parts. We just assume the existence of minimal units – making our lattice similar to that proposed by Chierchia (1998) rather than to the original Link model. Following Chierchia's account, we also assume identical formalisations for plurals and mass terms. So we can do away with the μ and $\top$ operators introduced in 5.

4.3 Divisibility

In order to differentiate cases where a mass term can be broken down into parts of itself and cases where it must be bound to eventualities to be quantifiable, we will talk of the **divisibility** of a concept. A concept is divisible if its instances have identifiable parts which could still be described in terms of that concept. Water and Styrofoam are typical examples of divisible entities. A concept is indivisible if there are no obvious conceptual parts for it and it needs binding to eventualities to be quantified. Divisible mass terms can be represented as a lattice of non-overlapping parts. Indivisible mass terms remain single entities and quantification requires a mapping to a lattice of eventualities. It would be tempting to associate indivisible mass terms with abstract terms at the lexicon level, but we will refrain from doing so as the classification is not so clear: the noun *information* can, in certain contexts, easily be divided (imagine the information contained in a report, presented as a series of bullet points). Further, some abstractions are essentially count nouns (*idea*) or can be used as count nouns (*the loves of Romeo for Juliet and of Isolde for Tristan*). Note that we assume superordinates (*furniture*, *clothing*) to be divisible mass terms: they have clearly identifiable – although heterogeneous – parts, which can be counted.

5 Formalising indivisible mass terms

We argue that the perceived quantification in sentences such as *Most love is expressed through actions* is actually quantification over eventualities of a single entity. Although we prefer to talk of concepts and eventualities, we draw the parallel between our terminology and Carlson's notion of individual-level and stage-level properties (1977). It is possible to quantify over situations where a concept occurs as it is possible to quantify over situations involving an individual. The predicates appearing in conjunction with concept eventualities can be taken as equivalent to 'stage-level predicates'. We will show in the next section that some predicates apply to concepts directly rather than to eventualities thereof – those are equivalent to 'individual-level predicates'.

We assume the existence of an eventuality lattice, similar to the standard plural lattice proposed by Chierchia. Each indivisible concept has its own eventuality lattice, over which it can be quantified. So the formalisation of *Most love is expressed through actions* can be written as follows:

10. $X = \sigma^* e\ love'(e) \wedge \exists Y[Y \prod X \wedge \forall z[z \cdot \prod Y \rightarrow expressedThroughActions'(z)] \wedge (|Y - X| \leq |Y| < |X|)]$

where quantification happens over eventualities e of love.

6 Concepts

We note that quantification is not only contextual but also sense-sensitive. Consider the mass term *friction*:

11. Friction acting on the wheel slows down its rotation.
12. Friction is usually subdivided into dry friction, fluid friction, skin friction and internal friction.

The first sentence refers to particular eventualities of the friction phenomenon. We could paraphrase it with *Some frictional forces acting on the wheel slow down its rotation.* In the second sentence, the subject noun phrase denotes the concept, or phenomenon, of friction itself. The verbal predicate precludes plural quantification, as it makes no sense to say that particular eventualities of friction can be subdivided into dry friction, fluid friction, etc. In those cases, the sentence can be given a simple subject/predicate formalisation of the form $\psi(\phi)$, akin to the one given to proper nouns. The quantification of the subject is simply one.

Note that certain accounts regard sentences such as 12 as generic, in virtue of the subject noun phrase referring to a kind (see Krifka et al, 1995). But following Chierchia (1998) and Krifka (2004), we prefer to define a kind as a function that returns the greatest element of the extension of the property relevant to that kind, that is, as the supremum of all instances with property P'. We then argue for regarding *friction* in 12 as a concept (as opposed to an instance of that concept), that is, as a unique entity. A quantificational account of the sentence is possible as before:

13. $X = \sigma^* x\ friction'_{CONCEPT}(x) \wedge \exists Y[Y \prod X \wedge subdivided'(Y)] \wedge (|Y - X| = 0)]$ where $|X| = 1$

7 A final note on the mass/count distinction

The straightforward linguistic definition of mass terms is unfortunately unable to classify all nouns into two clear categories: consider nouns such as *scissors* or *brains* (as in *She has brains*). The former requires a classifier to be quantified over (*two pairs of scissors/*two scissors*) but does not occur with determiners normally associated with mass terms (**much scissors*). The latter, despite being pluralised, prefers mass term determiners (*You don't need much brains/*many brains to understand mass terms*). Further, the mass/count distinction is subject to regional variability: it is possible to utter *two scissors* in some dialects of English. Wisniewski (2010) argues that the difference between mass and count terms is generally conceptually motivated and mirrored in grammar, but that linguistic phenomena must be taken into account when trying to explain certain grammatical oddities: for instance, the fact that we refer to slices of bread (clearly individuated entities) as ‘toast’.

We suggest that the mass/count distinction is predominantly underspecified in English: it only becomes apparent in explicitly quantified noun phrases. It is however the case that humans are able to select the appropriate conceptual system when using and understanding NPs. We would expect that, when asked *How many beers did Tom drink last night?*, the hearer would try and remember the number of cans ingested by Tom, while the question *How much beer will that barrel hold?* implies a notion of volume. The utterance *Much drinking happened last night* involves yet another unit, based on eventualities rather than parts.

As far as formalising such noun phrases is concerned, the precise nature of the minimal units of quantification is irrelevant. We just need a representation for such minimal units, whatever they are. The domain of quantification, however, must be specified: it would be ontologically incorrect to represent things and events in the same way; we can similarly argue for a distinction between things and eventualities. Further work should therefore involve an empirical annotation study to show whether humans can distinguish direct quantification over individuals from quantification over eventualities.

References

1. Leonard Bloomfield. 1933. *Language*. New York: Holt, Rinehart and Winston.
2. Susan Carey. 1991. ‘Knowledge acquisition: Enrichment or conceptual change?’ In: Susan Carey and Rochel Gelman, Editors, *The epigenesis of mind: Essays on biology and cognition*, pages 257-291. Hillsdale, NJ: Erlbaum.
3. Gregory Carlson. 1977. *Reference to Kinds in English*. Ph.D. Dissertation. University of Massachusetts at Amherst. Published by Garland, New York, 1980.
4. Gennaro Chierchia. 1998. ‘Reference to kinds across languages’. *Natural Language Semantics*, 6:339–405.
5. Ann Copestake and Ted Briscoe. 1995. ‘Semi-productive Polysemy and Sense Extension’. *Journal of Semantics*, 12:15–67.
6. Aurelie Herbelot. To appear. *Underspecified quantification*. Ph.D. Thesis. Computer Laboratory. University of Cambridge, UK.
7. Aurelie Herbelot and Ann Copestake. 2010. ‘Annotating Underquantification’. In: *Proceedings of The Fourth Linguistic Annotation Workshop (LAW IV) at ACL’10*, Uppsala, Sweden.
8. Manfred Krifka, Francis Jeffry Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link and Gennaro Chierchia. 1995. ‘Genericity: An Introduction’. In Gregory N. Carlson and Francis Jeffry Pelletier, Editors, *The Generic Book*, pages 1–125. Chicago: Chicago University Press.
9. Fred Landman. 1989. ‘Groups, I’. *Linguistics and Philosophy*, 12(5):559–605.
10. Godehard Link. 1983. ‘The Logical Analysis of Plurals and Mass Terms: a lattice–theoretical approach’. In Rainer Bauerle, Christoph Schwarze and Arnim von Stechow, Editors, *Meaning, Use, and Interpretation of Language*, pages 302–323. Berlin: de Gruyter.
11. Carol L. Smith a, Gregg E.A. Solomon and Susan Carey. 2005. ‘Never getting to zero: Elementary school students’ understanding of the infinite divisibility of number and matter’. *Cognitive Psychology*, 51(2):101-140.
12. Dina Tirosh and Ruth Stavy. 1996. ‘Intuitive rules in science and mathematics: the case of ‘Everything can be divided by two’ ’. *International Journal of Science Education*, 18:653667.
13. Edward Wisniewski. 2010. ‘On Using Count Nouns, Mass Nouns, and Pluralia Tantum: What Counts?’ In: Francis Jeffry Pelletier, Editor, *Kinds, Things, and Stuff*. Oxford University Press.