

Johanna Marie Poppek, Tibor Kiss, Francis Jeffrey Pelletier

Kinds, Containers, Instances: Mass Nouns and Plurality

Abstract The existence of formally realized plurality in the domain of mass nouns is a major challenge, especially if the hypothesis is taken that mass nouns possess some kind of “built in” plurality as their main distinguishing feature compared to count nouns. To address this issue, we performed a large-scale corpus study on the plural occurrences of mass nouns and dual life nouns using the OANC corpus and a database of noun-sense pairs annotated in terms of their countability class. Results showed that not only do pluralizations of mass terms occur frequently in the corpus, the nature of their meaning shifts differs with regards to their specific countability class, providing a deeper insight into the semantic and pragmatic nature of the count and mass continuum.

Keywords Countability, plural, mass terms, corpus study

1 Introduction

The existence of formally realized plurality in the domain of mass nouns is a major challenge, especially if the hypothesis is taken that mass nouns possess some kind of “built in” plurality as their main distinguishing feature compared to count nouns, noting that nouns that possess a (morphological) plural are usually considered count (e.g., Chierchia 1998). Other approaches stress the general similarity of mass nouns and plural expressions, leaving out the field of plurality of mass nouns (e.g., Lasersohn 2011).

In this article, we will present a large-scale corpus study as an approach for a systematic analysis of mass terms and plurality and their implications. It is based on a fine-grained nominal classification resource (*Bochum English Countability Lexicon* Kiss et. al. 2014 and 2016) that eschews both a binary distinction and a lemma-based approach to countability.

1.1 Data

Since the analysis of the general phenomenon of countability is usually described as a binary feature dividing the domain into only two realms of countable and uncountable nouns (e.g., Borer 2005) and only addressed with a small set of staple nouns, we created a database that allows the study of countability on a larger scale and in a more fine-grained way. The database consists of approximately 12,000 English noun-sense pairs that were enriched with their WordNet definitions for every sense and annotated in terms of countability by four native speaker annotators using a set of six pattern test questions to test their semantic and syntactic behavior. The resulting 18 subclasses are grouped in four major classes that represent the general complexity of the countability issue. Table 1 shows the general distribution of major classes for the resulting pairs consistently annotated by at least two annotators and provides examples for each class in terms of WordNet lemma, POS-Tag and sense number. Note that the names of the subclasses are an artifact of the initial classification process carried out in R and were kept as neutral captions for the respective classes.

It should be stressed here that the annotation for every sense did apply at the type level without any access to corpus data, not at the token level. Therefore, a classification e.g., as *both mass and count* does imply a deviant position in the count-mass continuum or a dual life nature, while *neither mass nor count* contains senses where the whole distinction does not seem to apply, e.g., unique entities. It should further be noted that although most noun-sense pairs are classified as *regular count* or *regular mass* as accounted for in the literature, there is a relevant amount of data that does not fit into the binary scheme, showing that the issue of countability resembles more a continuum or a spectrum than a distinction.

Table 1: Major Classes of BECL

Major Class	Frequency	Subclasses	Examples
Regular Count	8,434	235, 721, 371, 73	<i>animal.n.01; childhood.n.01; manners.n.01; making.n.03</i>
Regular Mass	2,427	528, 519, 531	<i>knowledge.n.01; adaptability.n.01; lingerie.n.01</i>
Both Mass and Count	699	510, 726, 729, 513	<i>glue.n.01; superstition.n.01; theft.n.01; china.n.04</i>
Neither Mass nor Count	315	523, 37, 190, 514, 199, 28, 353	<i>doomsday.n.02; infinite.n.01; midline.n.01; provenance.n.01; heyday.n.01; midst.n.01; hamlet.n.02</i>
Total	11,875	18	

To further support research on the issue, the resource is made publicly available via <http://www.count-and-mass.org>.

1.2 Approach

Although the nominal classification of the resource in terms of countability already allowed insight into the type level of countability, there are several approaches to address the phenomenon of mass-to-count or count-to-mass shifts at the *token level* (e.g.; De Belder 2008b; Nicholas 2002 among others) Mass-to-count shifts are usually determined by a “deviant” behavior of a noun that is usually classified as mass. “Deviant behavior”, in case of mass terms, could occur with an indefinite article in the singular or with a morphologically realized plural. This does not include cases like pluralia tantum (e. g., *scissors*) that fall into a different countability category, but rather to genuine mass terms that occur in a plural form.

To determine the distribution of plural occurrences of apparent mass nouns, we have used the Stanford NLP system¹ to parse sentences from the *Open American National Corpus* (OANC, <http://www.anc.org>) containing nouns from three mass noun classes of the database (528, 510 and 726) and extracted sentences that showed plural occurrences despite the nouns being classified as *mass nouns exclusively* (528) and *dual use nouns* (510, 726). For more information on the classes and their annotation pattern cf. Table 2.

2 Corpus Study on Plural Mass Terms

The phenomenon addressed here takes place at the token level of a specific lemma. Since our data is annotated at the sense level, we took only completely annotated lemmata into account (meaning that all senses of a lemma that WordNet provides must be present in our data) that consistently belong to one subclass with respect to all their senses. The general hypothesis is that mass nouns of class 528 should not possess a morphological plural, while plural occurrences of mass terms from class 510 and 726 should be accompanied by a meaning shift (cf. Borer 2005; Chierchia 1998 on plural meaning shifts on mass terms).

The sentences extracted from the OANC corpus contained approximately 1,900 plurality examples for class 528 (167 lemmata), approximately 5,400 examples for class 510 (241 lemmata) and approximately 1,500 plural occurrences (64 lemmata) for class 726. Most lemmata contained in all three classes showed

1 Included in the parser software package (<http://nlp.stanford.edu/software/nndep.shtml>).

Table 2: Annotation Patterns

528 (regular mass)	510 (both mass and count)	726 (both mass and count)
Can be combined with <i>more</i> , the resulting sentence uses a mode of measurement other than number	Can be combined with <i>more</i> , the resulting sentence uses a mode of measurement other than number	Can be combined with <i>more</i> , the resulting sentence uses a mode of measurement other than number
<i>more</i> + morphological plural is <i>not applicable</i>	<i>more</i> + morphological plural is possible and semantically equivalent to a sentence with an explicit classifier	<i>more</i> + morphological plural is possible and semantically equivalent to a sentence with an explicit classifier
Singular form can be subject of a classification or definition without, but not with an indefinite determiner (*A <sense> is a kind of X)	Singular form can be subject of a classification or definition without, but not with an indefinite determiner (*A <sense> is a kind of X)	Singular form can be subject of a classification or definition with and without an indefinite determiner (A <sense> is a kind of X)
Example: <i>flexibility</i>	Example: <i>punishment</i>	Example: <i>friendship</i>

several plural occurrences in the corpus, so we can assume that mass plurals are not a rare phenomenon. Besides these generally high frequencies, all three classes showed a behavior that can be described as mass-to-count type shifting. Type shifting, for this matter, would indicate an arising interpretation as a *kind*, a *unit* or an *instantiation* of an act, event or result (cf. Table 3 for examples from OANC).

Table 3: Type Shifting Examples

Unit Interpretation: Three carboxy-terminal tyrosines (positions 624-6), hypothesized to play regulatory roles, were replaced by <i>phenylalanines</i> .
Kind Interpretation: The universe, in short, is breaking <i>symmetries</i> all the time by generating such novelities, creating distinctive molecules or other forms which had never existed before.
Instantiation Interpretation: The reaction products were purified by means of three repeated gel <i>chromatographies</i> using water-saturated Sephadex G-50 in Millipore/ Multiscreen filtration plates according to the instructions provided by the supplier and dried under vacuum.

Kind interpretations could also be labelled *species interpretation* or *type interpretation* and imply an interpretation shift from a (bare) mass reading towards an element from a greater variety, meaning a *kind of something* or creating a *class of objects*. By this process of interpretation, a mass term can obtain a countable interpretation.

- (1) Experts tend to implicate increased environmental exposure to *carcinogens*.

Unit interpretations describe the general phenomenon of *containering* a mass term into certain bits or quantities thus allowing it to be counted, usually by using a specific measure phrase. However, the corpus study showed that unit interpretations do not require a measure phrase, but can also be contextually derived:

- (2) The PCR products of the ITS were resolved as single bands on 1 % agarose *gels*. (without measure phrase)
- (3) Using an experimental group and a control group, researchers would compare levels of *pesticides* found in settled dust, on children's hands, and in their blood, urine, or hair. (with measure phrase)

It should be noted here that especially unit interpretations without a specific measure phrase can be easily confused with kind interpretations. Since the example (2) provides contextual information that the plural of *gel* refers to the same type of object, the plural is interpreted as *portions* of something, not *kinds*, in contrast to example (1) where the context strongly suggests a number of different kinds of *carcinogens*. Nonetheless, both categories show a certain amount of overlap in some cases depending on the nature of the noun.

We observed another kind of type shift we call the *instantiation interpretation*. In those cases, nouns are coerced into a countable noun by an interpretation as an act, an event or a result.

- (4) In most places, heavy *snowfalls* are considered a troublesome (albeit picturesque) natural phenomenon.

This type of mass-to-count shifting is rarely discussed in the literature, and when it is, it is usually described as a restricted extension to certain categories of nouns, and as neither regular nor predictable (e.g., by Payne & Huddleston 2002).

While those kinds of mass term pluralization have been partly described in the literature (cf. e.g., De Belder 2008a and 2008b or Payne & Huddleston 2002), the phenomenon has, to our knowledge, not been addressed on a large scale in terms of observing general frequencies and implications of plural mass terms.

Although all three classes showed all three kinds of meaning shifts, the distribution of shifting interpretations strongly differs, resulting in a stronger preference for a *unit interpretation* or an *instantiation interpretation* for dual use nouns and as an *instantiation* or a *kind* for proper mass nouns.

These empirical results provide impulses for two observations. First, mass term plurals seem to occur with regularity and show a certain variation which is

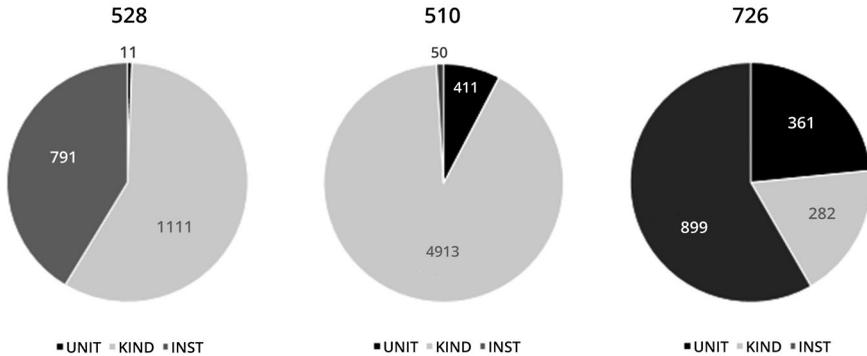


Figure 1: Type-Shifting Distribution.

greater than that which is accounted for by other researchers. In particular, the frequent observation of instantiation type-shifting formerly classified as rather rare (cf. Payne & Huddleston 2002) implies that the phenomenon might be a lot more common than hitherto thought. Second, the countability subclass seems to have a strong effect on the general distribution of the type-shifting classes (and also their frequency), implying that there might be a semantic effect that is revealed though a large-scale analysis (Figure 1).

3 Conclusion and Further Work

This first corpus study showed that although they are neglected by a large amount of current research, mass term plurals frequently occur in actual language data. In addition to this, our observations imply that they also follow certain regularities. Since the kind of type shift also seems to be based on the general semantic nature of the noun (only abstract nouns can undergo an instantiation type shift, for example, cf. Payne & Huddleston 2002), the general distribution of the coercion examples also allows a closer look at a general semantic pattern that might influence the position of a noun inside the countability continuum and to clarify to what extent those phenomena could be the result of a systematic polysemy. The variation inside the data of the different subclasses also implies that a more fine-grained view of the count and mass spectrum can provide a deeper insight into mechanisms that might be overlooked in a broader classification.

The data extracted thus provides the basis for an account of the varying effects of plurality within the class of “mass terms” and shows how large-scale corpus studies are able to address a basically underresourced phenomenon. Further research will extend to similar countability classes as well as analyzing the general semantic and pragmatic nature of pluralization of mass nouns.

References

- Alexiadou, Artemis. 2011. Plural Mass Nouns and the Morphosyntax of Number. In *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 33–41.
- Borer, Hagit. 2005. *Structuring Sense. Vol. I: In Name Only*. Oxford: Oxford University Press.
- Chierchia, Gennaro. 1998. Plurality of Mass Nouns and the Notion of the “Semantic Parameter”. In Susan Rothstein (ed.), *Events in Grammar*. Dordrecht: Kluwer, 53–103.
- De Belder, Marijke. 2008a. Size matters: Towards a syntactic decomposition of countability. In Natasha Abner and Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Somerville: Cascadilla Proceedings Project.
- De Belder, Marijke. 2008b. Sizing up countability: Towards a more fine-grained mass-count distinction. Talk at ConSOLE, Paris, 10–12 January 2008. [https://lirias.kuleuven.be/bitstream/123456789/409023/1/abstract+Console+De+Belder+\(%2Bname\).pdf](https://lirias.kuleuven.be/bitstream/123456789/409023/1/abstract+Console+De+Belder+(%2Bname).pdf).
- Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Katz, Graham and Roberto Zamparelli. 2011. Meaning-shifting plurality and the Count/Mass Distinction. In: *QITL-4 – Proceedings of Quantitative Investigations in Theoretical Linguistics 4* 29.03.2011–31.03.2011, Berlin, Humboldt-Universität, 43–46.
- Kiss, Tibor, Francis Jeffrey Pelletier, Halima Husic, Johanna Marie Poppek and R. Nino Simunic. 2016. A Sense-Based Lexicon for Count and Mass Expressions: The Bochum English Countability Lexicon. In *Proceedings of LREC 2016*. Portorož, Slovenia.
- Kiss, Tibor, Francis Jeffrey Pelletier and Tobias Stadtfeld. 2014. Building a Reference Lexicon for Countability in English. In *Proceedings of LREC 2014*. Reykjavik, Iceland.
- Lasersohn, Peter. 2011. Mass Nouns and Plurals. In: *Semantics: An International Handbook of Natural Language Meaning*. Berlin: de Gruyter.
- Nicolas, David. 2002. Conversions of count nouns into mass nouns in French: the roles of semantic and pragmatic factors in their interpretations. https://halshs.archives-ouvertes.fr/ijnl_00000623v2/document.
- Nicolas, David. 2008. Mass nouns and plural logic. In *Linguistics and Philosophy*, 31 (2): 211–244.
- Payne, John, Rodney Huddleston 2002. Nouns and Noun Phrases. In *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press. 323–524.

