

# Nominal Coercion in Space: Mass/Count Nouns and Distributional Semantics

**Manuela Hürlimann**

EM LCT/University of Trento

huerlimann.manuela@gmail.com

**Raffaella Bernardi**

**Denis Paperno**

University of Trento

{first.last}@unitn.it

## Abstract

**English** Theoretical linguists analyse all nouns as either mass or count, but admit that noun meanings can be shifted from one class to the other and classify these shifts. We use distributional semantic models to check how the theoretical analysis of mass-count meaning shifts relates to the actual usage of the nouns.

**Italiano** *In linguistica i sostantivi inglesi sono divisi in numerabili e non numerabili. È però riconosciuto che il significato nominale può passare da una classe ad un'altra seguendo determinati tipi di "spostamenti". In questo lavoro, usiamo i modelli semantici distribuzionali per verificare se le teorie linguistiche sugli spostamenti del significato nominale abbiano riscontro nei dati.*

## 1 Introduction

It is generally assumed that if a mass (count) noun is used in a count (resp. mass) context, its meaning changes. Compare example (1), where *wine* is used in a mass context (as a bare singular; denoting a substance) to (2), where the use of the determiner *three* indicates a count usage, shifting its interpretation to *types of wine*.

- (1) I like wine.
- (2) Three wines grow in this region.

The same phenomenon can also be observed for count nouns: in example (3), *apple* is used in its more frequent count sense, while its bare usage in example (4) constitutes a mass usage with a slightly changed meaning — the focus is not on individual, whole apples as in the countable example, but on their material/substance.

- (3) I bought five apples at the market.
- (4) There is apple in the salad.

Data-based approaches to the mass/count phenomenon include Baldwin and Bond (2003), who classify nouns into five countability types based on lexico-syntactic features and Ryo Nagata et al. (2005), who use context words to distinguish between mass and count nouns.

Katz and Zamparelli (2012) were the first to study mass/count elasticity using distributional semantic models. First of all, they dispelled the view that there is a clear count/mass dichotomy: like in the examples above, many nouns which appear frequently in count contexts also appear frequently in mass contexts. Hence, rather than making a binary distinction (count vs. mass nouns), we should speak of *predominantly count* (resp., *predominantly mass*) nouns, i.e., nouns which occur more frequently in count (resp. mass) contexts than in mass (resp., count) contexts. Moreover, Katz and Zamparelli (2012) take pluralisation as a proxy for count usage and conjecture that for *predominantly count* nouns the similarity between singular and plural is higher than for *predominantly mass* nouns since the latter undergo a shift whereas the former do not. This conjecture finds quantitative support in their data – the 2-billion word ukWaC corpus.<sup>1</sup> We wonder whether other factors, such as polysemy, have an impact on this quantitative analysis and we further investigate nominal coercion by also considering the abstract vs. concrete dimension and polysemy.

Katz and Zamparelli (2012) notice that while plurals are invariably count, singulars can be a mixture of mass and count usages, and propose to use syntactic contexts to disambiguate mass and count usages in future studies.

We take up their suggestion and look at coercion using vector representations of mass vs. count us-

<sup>1</sup>wacky.sslmit.unibo.it/doku.php?id=corpora

ages.

According to the linguistic literature (Pelletier (1975)), instances of coercion fall into several shift classes. In this view, coerced nouns move towards a particular “destination”:

- **Container shift:** Liquids (mass) are coerced into countable quantities contained in containers: “two beers, please!”
- **Kind shift:** Masses are coerced into a kind reading: “three wines grow in this region”
- **Food shift:** Animal nouns are coerced into a mass food meaning: “there was chicken in the salad”
- **Universal grinder:** Countables are coerced into a mass reading: “after the accident, there was dog all over the street”

We wonder whether these shift classes can be identified in the semantic space. Thus, we propose a simple experiment in which we assess whether the count usage vectors of typical mass nouns move towards (=become more similar to) these suggested destinations.

In sum, we address the following research questions: (1) Do nouns undergo noticeable shifts – and if so, what factors have an impact? (2) Can we interpret the destination of a shift in terms of standard shift classes?

## 2 Distributional Semantic Models

Distributional Semantic Models are based on the assumption that the meaning of a word can be captured by counting its co-occurrences in a corpus with other words in a given vocabulary. Hence, word meaning can be represented by a vector and semantic similarity between two words can be captured using the cosine similarity of the corresponding vectors Turney and Pantel (2010). The bigger the cosine similarity, the closer are the two words semantically.

**Core Vector Space** We collected co-occurrence statistics from the concatenation of ukWaC, a mid-2009 dump of the English Wikipedia, and the British National Corpus, a total of 2.8 billion words. For each target word, its co-occurrence with all context words in the same sentence was counted, with the top 20K most frequent content word lemmas being used as context items. We

furthermore used Positive Pointwise Mutual Information as a weighting scheme, followed by dimensionality reduction (Singular Value Decomposition) to 400 dimensions. In this space, all usages of a noun are collapsed for building its vector. The model distinguishes, however, between singular and plural nouns (i.e., *cat-sg* and *cat-pl* are two different vectors). We consider those vectors as representing an average or “core” meaning across different usages.

**Vector Space of Mass and Count Usages** Mass and count usages of nouns were defined using the following determiners: *much*, *less* for mass usages, and *a*, *an*, *every*, *many*, *each*, *fewer*, cardinals, *more* + plural noun, *enough* + plural noun for count usages. In order to reduce noise due to parsing errors, determiners had to be adjacent to the noun and their part of speech tag had to be adjective (not adverb). Based on these syntactic patterns, co-occurrence values were collected for both usages and their final vector representation were then obtained by projection onto the core vector space.

## 3 Datasets

In order to understand whether polysemy and abstractness have an impact on Katz and Zamparelli (2012)’s results, we create a data set of singular and plural nouns. We expand on Katz and Zamparelli (2012)’s methodology by annotating these nouns with information on concreteness/abstractness and polysemy.

Secondly, in order to avoid side effects of noisy data and to overcome the limitations of the singular/plural nouns as a proxy for the mass/count distinction, we create a second data set filtered by noun frequency and use the vector representations of the disambiguated mass/count usages of the nouns.

### 3.1 Singular-Plural Data

This dataset contains a total of 3960 singular-plural noun pairs. Only nouns that occur in the corpus at least 10 times in either a mass or a count context were considered.

These nouns have been annotated with information about abstractness/concreteness and polysemy. We required nouns to be unambiguously annotated as either *abstraction.n.06* or *physical\_entity.n.01* in WordNet.<sup>2</sup> Furthermore, for a

<sup>2</sup>wordnet.princeton.edu

more fine-grained measure of concreteness, we used the Ghent database Brysbaert et al. (2013) to assign a concreteness score (1=most abstract, 5=most concrete) to each noun. We used WordNet also to annotate polysemy, quantified as the number of different senses (synsets) for each noun.

### 3.2 Mass-Count Data

To overcome the ambiguity problems associated with the singular-plural data, we create an additional dataset of mass and count nouns and their *usage vectors*.

We use the output from the syntactic patterns of the singular-plural dataset above (see section 3.1) and take the intersection between the nouns that occur with count determiners and those that occur with mass determiners. We clean this list by excluding nouns which occur less than 10 times in a mass context, obtaining 2433 nouns.

## 4 Experiments

### 4.1 Exp. 1: Do nouns undergo shifts?

In this first experiment we use the vectors of the singular-plural dataset in order to verify the results by Katz and Zamparelli (2012) against our data and to furthermore check for effects of abstractness and polysemy. Our hypotheses are:

1. Mass nouns undergo greater singular-plural meaning shifts than count nouns.
2. The more abstract a noun (lower concreteness score), the greater its meaning shift between singular and plural.
3. Nouns with a higher degree of polysemy (greater number of synsets) show a greater singular-plural distance.

We then assess the correlations between these annotations and the singular-plural similarity using the cosine measure. In order to run the correlation analyses, we normalise the count and mass context frequencies, thus creating a continuous variable. We define an alternative measure, “massiness”. For count context frequency  $c$  and mass context frequency  $m$ ,  $massiness = \frac{m}{(m+c)}$ . Massiness can take values between 0 and 1.

Table 1 shows the Pearson correlations between each of the annotations and the cosine similarity measure. All correlations are highly significant (p-values between 2.2e-16 and 6.40e-05).

	Pearson correlation with cosine
concreteness score	0.167
massiness	-0.225
synsets	-0.266

Table 1: Pearson correlation coefficient between annotated variables and cosine similarity.

While the correlations between the annotated variables and similarity scores are not large, they do reveal tendencies which intuitively make sense:

- **concreteness score:** The meaning of concrete nouns shifts less when pluralised.
- **massiness:** Nouns used more frequently in mass contexts undergo greater meaning shifts when pluralised.
- **synsets:** Smaller number of synsets (less polysemy) correlates with greater similarity. Nouns with more unambiguous meanings shift less when pluralised.

### 4.2 Exp. 2: Where does a shift take a noun?

An important aspect of nominal coercion is the destination of a coerced noun — since we found above that noun meanings indeed change, it would be interesting to investigate *how* they change, or, speaking in terms of the semantic space, *where* they are taken to by coercion.

**Destinations of shift classes** We look at the container and the kind shifts, which are the most intuitive and least controversial ones among those discussed in the linguistic literature. We take *beer*, *coffee* and *tea* as examples of mass nouns that undergo the container shift and *flour* and *wine* as examples of mass nouns that undergo the kind shift. We run a small-scale experiment in which we compare the cosine similarity of the mass and count usages of these nouns to another word taken as a potential destination of the shift. The results are reported in Table 2 — we can see that the count usage vectors are more similar to the expected destinations than the mass usage vectors, which is in accordance with the container and kind shift explanations.

**How far does coercion take a noun?** We conclude the analysis of the destination of nominal coercion by visualising the distance of usage vectors with respect to their “core” representation. We

Usage vector	destination	cosine
beer- $n_c$	pint-n	<b>0.674</b>
beer- $n_m$	pint-n	0.548
coffee- $n_c$	cup-n	<b>0.559</b>
coffee- $n_m$	cup-n	0.478
tea- $n_c$	cup-n	<b>0.577</b>
tea- $n_m$	cup-n	0.486
flour- $n_c$	variety-n	<b>0.267</b>
flour- $n_m$	variety-n	0.140
wine- $n_c$	variety-n	<b>0.470</b>
wine- $n_m$	variety-n	0.177

Table 2: Container vs. kind shifts.

generate a plot in which, for each noun, we put the cosine similarity of the mass usage vector to the “core” noun vector on the x-axis and the similarity of the count usage vector to the “core” noun on the y-axis (see Figure 1). It is evident that there is no strong relation between the two similarities, as indicated by the red fit line.

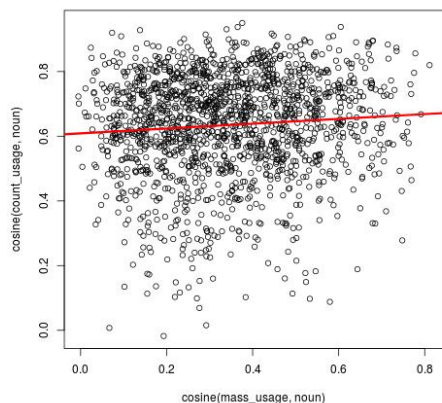


Figure 1: Similarity between mass usage and core noun plotted against similarity between count usage and core noun. Red line = linear fit.

Nouns in the lower left-hand corner (low-mass and low-count) are predominantly bare nouns; as they normally occur without determiners, their average meaning is not very similar to either the mass or the count usage. Words in the upper left-hand corner are nouns that are highly countable and do not seem to lend themselves much to mass usages. Contrary to the latter, words in the lower right-hand corner are nouns that are very “massy” and do not seem to be readily countable. The interesting cases (elastic nouns) are in the upper

	Low-count	High-count
Low-mass	diving, dissension	framework, diet
High-mass	importance, distress	love, fear

Table 3: Contingency table: examples

right-hand corner. For these nouns, both the mass and the count usage vectors are highly similar to the core noun vector. This corner seems to be where regular coercion, which is the subject of our study, lies. Many nouns in this corner shift from “abstract mental state” (mass) to “elements which elicit that state” (count), e.g. *love*, *fear*, *pleasure*. Similarly, *responsibility* shifts from a mental state to a list of concrete duties. Examples of nouns found in the four corners are reported in Table 3.

To sum up, regular coercion turns out to only slightly modify the meaning of the noun, so that neither the mass nor the count meaning shifts too far from the core meaning.

## 5 Conclusions

We have seen how Distributional Semantics Models (DSMs) can be applied to investigate nominal coercion. DSMs can capture some aspects of mass/count noun meaning shifts, such as the fact that predominantly mass nouns undergo greater meaning shifts than predominantly count nouns when pluralised. We also find that abstractness and polysemy have an impact on singular-plural distance: abstract nouns and highly polysemous nouns have a greater singular-plural distance than concrete and monosemous nouns, respectively. Furthermore, our second experiment shows that coercion lies mainly in cases where both the mass and count usage vectors stay close to the averaged noun meaning. However, as our toy evaluation of clear cases of container and kind coercion shows, the direction of the shift can be differentiated based on usage vectors.

## Acknowledgments

The first author was supported by the Erasmus Mundus European Masters Program in Language and Communication Technologies (EM LCT).

The other two authors were supported by COMPOSES (ERC 2011 Starting Independent Research Grant n. 283554).

We used the COMPOSES dissect toolkit (<http://clic.cimec.unitn.it/composes/toolkit/>) for our semantic space experiments.

We furthermore thank Roberto Zamparelli for sharing his huge knowledge of nominal coercion.

## References

- Baldwin, Timothy and Bond, Francis. 2003. "Learning the Countability of English Nouns from Corpus Data". In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pp. 463–470.
- Brysbaert, Marc; Warriner, Amy Beth and Kuperman, Victor. 2013. "Concreteness ratings for 40 thousand generally known English word lemmas". *Behavior research methods*, pp. 1–8.
- Katz, Grahahm and Zamparelli, Roberto. 2012. "Quantifying Count/Mass Elasticity". *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. Jaehoon Choi et al., pp. 371-379. Somerville, MA: Cascadilla Proceedings Project.
- Nagata, Ryo; Wakana, Takiro; Masui, Fumito; Kawai, Atsuo and Isu, Naoki. 2005. "Detecting Article Errors Based on the Mass Count Distinction". In: *Natural Language Processing IJCNLP 2005 Lecture Notes in Computer Science Volume 3651*, pp. 815–826.
- Pelletier, F.J. 1975. "Non-singular reference". *Philosophia* 5, pp. 1-14.
- Turney, Peter and Pantel, Patrick. 2010. "From frequency to meaning: Vector space models of semantics". *Journal of Artificial Intelligence Research* 37, pp. 141-188.