Acta Universitatis Wratislaviensis
No 3434
Anglica Wratislaviensia L
Wrocław 2012

Jacek Woźny
University of Wrocław

Calculating Countability — A Corpus-Based, Mereological Study of the Count/Mass Distinction of a Group of English Nouns

Abstract: A mereological, part-whole perspective is applied in a corpus-based study of noun countability to explain why some English nouns like *peas*, *flowers* and *pebbles* are countable and others like *maize*, *grass* and *gravel* are not, despite the fact that the size and other physical qualities of their referents are practically equal. The countability of a group of English nouns is linked with the average quantity of their referents in random British National Corpus samples.

1. Introduction

Frank Joosten (2003) divides the linguistic enquiry into count/uncount distinction into four major schools: grammatical (Bloomfield 1933), ontological (Quine 1960), conceptual-semantic (for example: Wierzbicka 1988; 1991; Langacker 1987; Berezowski 1999) and contextual (for example: Ware 1979). He summarizes his paper as follows:

> … the count-mass distinction cannot be reduced to an exclusively grammatical, ontological, conceptual-semantic, or contextual issue. Instead, it should be analyzed as a multidimensional phenomenon … (Joosten, 227)

While I agree that an enlightened, wide-scope approach should be paramount for any scientific endeavour, I think Joosten underestimates the conceptual-semantic approach, by not noticing that it is in fact based on an intricate connection of grammar, ontology, conceptualization and context, and therefore already multidimensional.

Joosten's critique of the conceptual-semantic approach is its inability to account for the uncountability of for example 'rice':

> It is highly improbable that all count-mass alternations can be explained in terms of conceptualisation. Why, for instance, has the English language chosen counthood for *pea* (*a pea, many*

> *peas*) and masshood for *rice* (*rice, much rice*)? Postulating a difference in conceptualisation looks very much like an ad-hoc solution. (Joosten, 223)

As an example, Joosten refers to Wierzbicka's (1988) claiming that boiled rice (the most often encountered state) is a continuous mass as opposed to uncooked rice. While some of Wierzbicka's interpretations of (un)countability of English nouns may trigger certain doubts,[1] they are not sufficient to undermine her general claim that conceptualization can account for the countable/uncountable opposition. I would like to support the conceptual-semantic approach to countability by applying a mereological part-whole perspective to answer Joosten's question of why the noun *pea* is countable and for example *maize* (or *rice*) is not.

The key notion of the method can be found in Berezowski (1999: 166): "One quality that seems to be at stake here is the size of the referent a n d t h e w a y i t i s p e r c e i v e d b y t h e s p e a k e r" (emphasis — J.W.). The motivation for (un)countability should be looked for not only in but around the referent because the size perception depends on the physical context. A similar idea was expressed by Radden and Dirven: "When we look at a group of objects or people from a distance they tend to shade into each other and appear as a mass" (2007: 68). The equivalent meaning seems to be expressed by noun phrases like 'a handful of coins' or 'a river of cars.' The perceived size of the single referent-particles seems to be inversely proportional to their number. For example: the more coins there are, the smaller the perceived size of a single coin-particle. The only problem that remains to be solved before we can put the above reflection to practical use is how to measure the perceived size of the referent. The solution to this problem and the proof that the (un)countability of nouns can be predicted and calculated from the linguistic corpus data is presented in the next section.

2. The method of 'calculating' the perceived referent size

Berezowski (1999) gives the following three series of nouns to exemplify the change of countability as a result of the 'perceived referent size' (further on referred to as PRS).

(a) boulder	(a) tree	(a) nut
(a) rock	(a) bush	(an) acorn
(a) stone	(a) plant	(a) bean
(a) pebble	(a) flower	(a) pea
------------------	-------------------	--------------------

[1] For example, I disagree with Wierzbicka's explanation of the uncountability of 'jewellery,' 'equipment' and 'furniture' (Wierzbicka 1983, qtd. in Berezowski 1999: 165). In my opinion, those nouns are examples of what Jackendoff (1990: 84) calls cross-categorical uncountability — in this case reflecting the continuity and monotony of the activities the three nouns refer to (making jewellery, equipping and furnishing).

gravel	grass	corn (maize — J.W.)
sand	clover	wheat
dirt	hay	pepper
dust	moss	poppy seed

(Berezowski 1999: 166)

I decided to account for the PRS by estimating the quantity in which the above nouns appear in the British National Corpus and assuming that the perceived relative size (PRS) of a referent (*r*) is inversely proportional to its quantity:

$$PRS\ (r) \sim 1/quantity(r) \qquad (1)$$

In mereological terms, the count/mass conceptualization depends on the quantitative part-whole relation. The frequency with which the nouns appear in different size contexts in the BNC speaks directly of an average quantity a given noun is associated with — and hence of its PRS. Let us define the average perceived referent size as the reverse of the average quantity of the referent in the BNC:

$$APRS(r) = 1/\ aquantity(r) \qquad (2)$$

The proportionality sign '~' in equation (1) was replaced by equality '=' in equation (2) for the sake of simplicity, since we are not going to use any particular units, but just order of magnitude. Given an *N*-element (*N* texts) sample of the BNC, the average quantity of a referent is defined by the following formula:

$$aquantity\ (r) = \frac{1}{N}\sum_{i=1}^{N} q_i(r) \qquad (3)$$

where $q_i(r)$ is the quantity of the referent in text$_i$ of the BNC sample. Of course, the BNC does not provide information on the exact amount of the substance in question, but it does not mean that it cannot be estimated and compared.[2] To this end I decided to divide the quantities of the referents into three orders of magnitude, because it would be difficult to divide them unambiguously into more quantity groups due to the imprecision of the information concerning the quantity in question. It is, after all, a rare case that a BNC sample contains the exact weight measurement of the referent. For example, for the first group of nouns (stones, pebble, gravel, sand) the following orders of quantity were used:

quantity 1 — a couple, a small number like ten or twenty, less than a kilogram
quantity 2 — a trailer, a ton, garden quantity
quantity 3 — beach, larger area, many tons

Each of the BNC texts in the random sample was then classified into one of those three quantity groups (cf. App. H, which contains all the BNC texts together

[2] In physics, especially in astronomy, scientists very often have to rely not on the quantity itself, but its order of magnitude only. For example, if a star is classified as a *red dwarf* it can weigh anything between 0.5 and 10 masses of the sun, which means that its mass is estimated to the nearest 1,988,920,000,000,000,000,000,000,000,000 kilograms.

with their quantity assessment). For example, the following three texts from a random BNC sample of a hundred texts were classified as *quantity 2*, *quantity 3* and *quantity 1*, respectively.

1. As we turn, discussion falls on the possibility of getting rid of the stones. They could, in theory, be collected in another trailer and taken away, instead of being put back on the field. A3A [quantity 2]
2. The infill of the cave is made up of a wide variety of materials, from silts to large stones, and the radioactive content of these materials is similarly variable. AC9 [quantity 3]
3. It is perfectly clear that long before the procession came into sight, long before the procession had formed, these people in the Markets in their desire to be offended had come down from the side streets and had taken great trouble to be offended, and not only were prepared to be offended but were prepared to throw missiles, stones and other weapons. AD2 [quantity 1][3]

After considering all 100 texts of the BNC sample in this manner, it was possible to establish the number of texts in each of the quantity classes. For example, for *stones* it was 60 texts in *quantity 1*, 25 in *quantity 2* and 15 in *quantity 3* class. It was then possible to approximate the average quantity from formula (3) in the following way:

$$aquantity(stones) \approx (1/100)\ (60 \times 1 + 25 \times 2 + 15 \times 3) = 1.55$$

For simplicity, since our considerations do not require considering the exact physical measurements of mass, I decided to replace the quantities q_1 , q_2 and q_3 simply with 1, 2 and 3. From Formula (2) we conclude that the average perceived size of a stone is

$$APRS(stone) = 1/1.55 = 0.65$$

which in itself is not informative at all, but very interesting if we can compare it with average perceived sizes of other referents like, for example, a pebble or a piece of gravel, which we will do in the next section.

The method of dividing a quantity into orders of magnitude is well known in mathematics as the *rectangle method of numerical integration.*[4] It is used for numerical approximation of the integral. For example, let us imagine that we have a function *f(x)* (Fig. 1) for which we want to calculate an integral, which in case of a 1-dimensional function is equal to the area between the graph of the function and the *x*-axis (it is the so called *Lebesgue integration*).

[3] Classifying corpus samples into one of the three quantity slots is not always straightforward. In this case we can imagine that an individual thrower would have a kilogram or a couple of kilograms of stones to throw rather than a ton or a trailer.

[4] An integral of a function can be represented as an infinite sum of the values of that function, so the rectangle method of numerical integration is also typically applied to estimate values of sums like the one in formula (3).

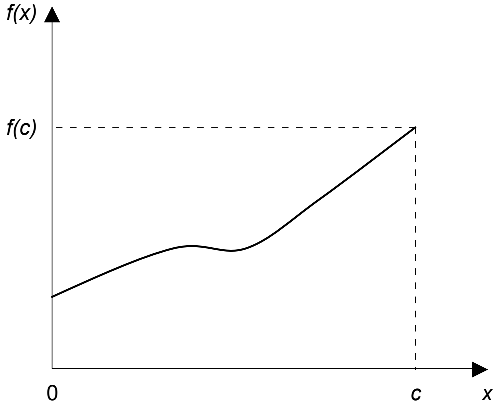

Figure 1. The integral of function *f(x)* equals the area under the graph

The simplest but least accurate method of estimating the area under the graph of function *f(x)* is to treat it as equal to the area of the rectangle partly marked with the dashed lines in Figure 1.

$$\int_0^c f(x)\,dx \approx c \cdot f(c) \qquad (4)$$

A much better approximation of the integral, however, would be to divide *x* into several orders of magnitude and estimate the integral as the area of several rectangles, as in Figure 2.

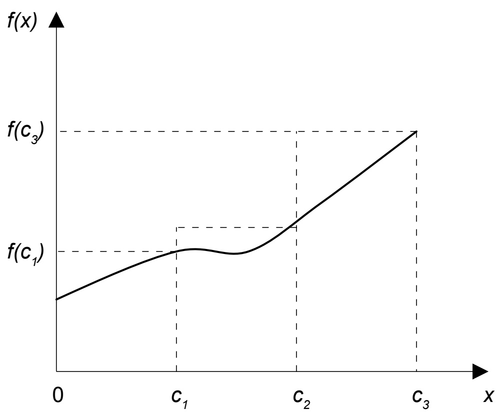

Figure 2. The rectangle method of estimating an integral

$$\int_0^c f(x)\,dx \approx \sum_{i=1}^{3} c_i \cdot f(c_i) = c_1 \cdot f(c_1) + c_2 \cdot f(c_2) + c_3 \cdot f(c_3) \qquad (5)$$

By applying (5) to (3) we obtain

$$aquantity(r) \approx (1/N)\ (n_1 q_1(r) + n_2 q_2(r) + n_3 q_3(r)) \qquad (6)$$

where N is the number of texts in a BNC sample, n_1 is the number of texts with quantity q_1, n_2 is the number of texts with quantity q_2 and n_3 is the number of texts with quantity q_3 of referent r. For example, as we have already observed above, for *stones* n_1=60, n_2=25, n_3=15 (cf. Table 1 in the next section), and for all referents considered below N=100, because in each case such was the size of the BNC sample used.

We are now ready to calculate the average quantity (*aquantity*) and *APRS* (average perceived referent size) for the three series of nouns mentioned at the beginning of this section.

3. The results of the corpus research

Table 1 below contains the results of the corpus based research for the first series of nouns (stones, pebbles, gravel, sand). Tables 2 and 3, respectively, contain the results for the second and the third series of nouns mentioned in the previous section. The last two columns of the three tables (average quantity and average PRS) are represented graphically in Figures 3–8. The significance of the numerical results presented below will be discussed in the final section.

Series 1: stones, pebbles, gravel, sand

Table 1. The percentage of occurrences in the BNC of series-1 nouns in different quantity contexts

noun	quantity 1 [%]	quantity 2 [%]	quantity 3 [%]	average quantity (*aquantity*)	APRS (average perceived referent size)
stones	60	25	15	1.55	0.65
pebbles	30	15	55	2.25	0.44
gravel	0	30	70	2.70	0.37
sand	10	10	80	2.70	0.3

quantity 1 — a couple, a small number like ten or twenty, less than a kilogram; quantity 2 — a trailer, a ton, garden quantity; quantity 3 — beach, larger area, many tons

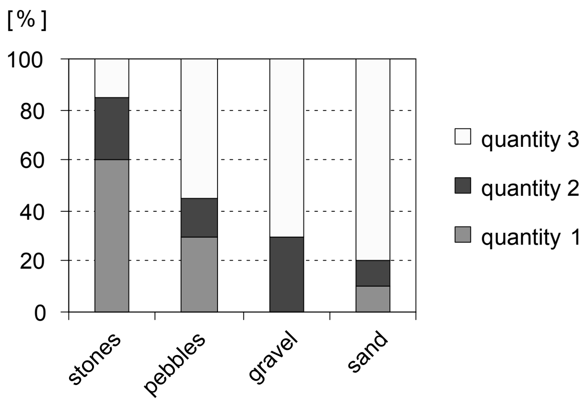

Figure 3. The percentage of occurrences in the BNC of series-1 nouns in different quantity contexts

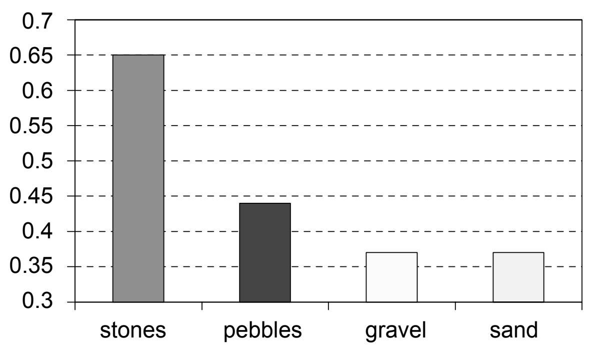

Figure 4. APRS (average perceived referent size) values for series-1 nouns

Series 2: plants, flowers, grass, clover

Table 2. The percentage of occurrences in the BNC of series-2 nouns in different quantity contexts

noun	quantity 1 [%]	quantity 2 [%]	quantity 3 [%]	average quantity	APRS (average perceived referent size)
plants	35	45	20	1.85	0.54
flowers	60	30	10	1.50	0.67
grass	5	50	45	2.40	0.42
clover	5	10	85	2.80	0.36

quantity 1 — around or fewer than ten specimens, house/flat decoration; quantity 2 — garden context; quantity 3 — plantation, country flora

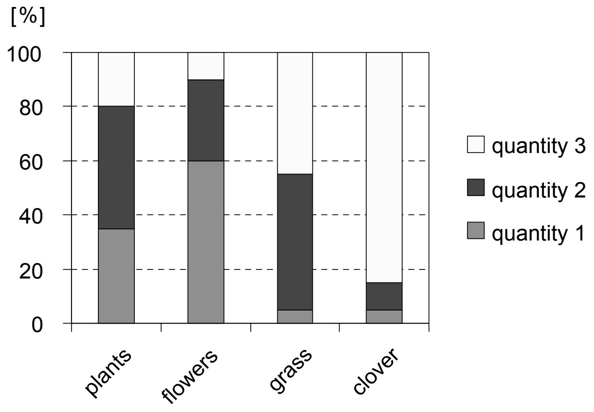

Figure 5. The percentage of occurrences in the BNC of series-2 nouns in different quantity contexts

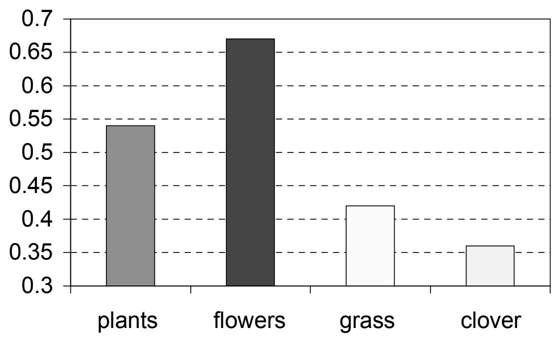

Figure 6. APRS (average perceived referent size) values for series-2 nouns

Series 3: beans, peas, maize, wheat

Table 3. The percentage of occurrences in the BNC of series-3 nouns in different quantity contexts

noun	quantity 1 [%]	quantity 2 [%]	quantity 3 [%]	average quantity	APRS (average perceived referent size)
beans	80	15	5	1.25	0.80
peas	100	0	0	1.00	1.00
maize	25	50	25	2.00	0.50
wheat	10	65	25	2.25	0.44

quantity 1 — kitchen, cooking context, a few packets, less than a kilogram; quantity 2 — farm context, field crops, a trailer, a couple of tons; quantity 3 — industrial and trade context, hundreds of tons, national or regional crops

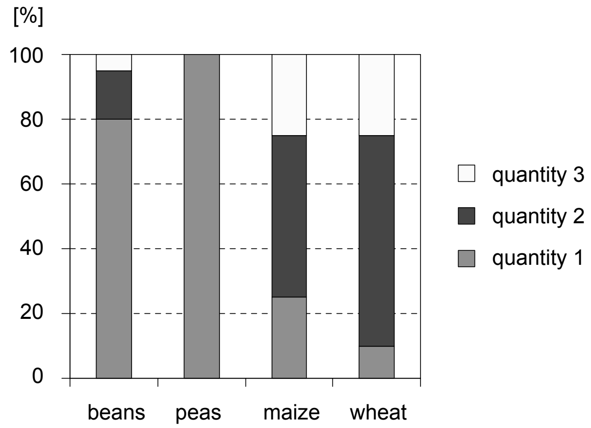

Figure 7. The percentage of occurrences in the BNC of series-3 nouns in different quantity contexts

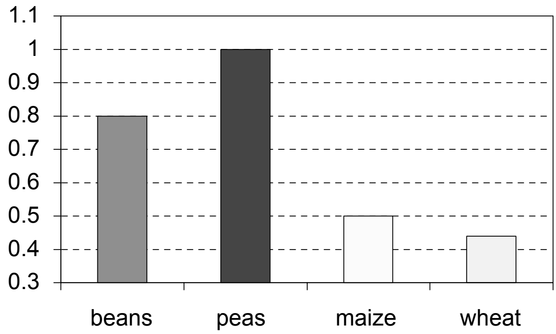

Figure 8. APRS (average perceived referent size) values for series-3 nouns

4. Summary and conclusion

Figures 3, 5 and 7 clearly show the same tendency — if we look at the bars representing only quantity 1 first, we will see a considerable drop in the middle of each diagram where the countability border is situated. And thus, looking at Figure 3, we notice that while pebbles in 30% of the BNC samples appear in quantity 1 context, gravel does not appear in quantity 1 at all, which means that in 100% of the samples it appears in much lager quantity 2 and quantity 3 contexts. The 'jump' at the pebbles-gravel border then measures 30%. The analogous drops for flowers-grass and peas-maize pairs are 55% and a staggering 75%, respectively. The decrease of quantity 1 bars is accompanied by an increase of quantity 2 and quantity 3 bars in Figures 1, 3 and 5, meaning that the quantity of the referents increases. In accordance with formula (1), the increase of the quantity of the referent causes the decrease of the PRS for each of the noun series,[5] which is confirmed by the APRS values shown by Figures 4, 6 and 8. The APRS drops at the countability border by 0.07 (16%), 0.25 (37%) and 0.50 (50%), respectively, which confirms that noun countability status can be explained by its perceived referent size (PRS), a mereological part-whole quotient and a quantity which lends itself to a corpus-based assessment.

Let us come back to Joosten's (223) example concerning the countability of *peas* and the uncountability of *rice*. As we can see in Table 3, all of the 100 random BNC texts refer to the minimum *quantity 1* (kitchen, cooking context, a few packets, less than a kilogram) of peas. In BNC peas appear in small numbers, therefore each individual pea seems big in comparison to the whole quantity referred to in a given text. Conversely, *maize* appears in *quantity 1* only 25 times, 50 times in *quantity 2* (farm context, field crops, a trailer, a couple of tons) and again 25 times in *quantity 3* (industrial and trade context, hundreds of tons, national or regional crops). An individual grain of maize, wheat or rice, although of similar size to a pea, is conceptualised as smaller because it is typically encountered in larger quantities.[6] The previous sentence may appear to be a rather imprecise intuition, but hopefully the method of calculating and comparing the APRS presented above will allow us (after extending the number of nouns investigated) to state it with some degree of certainty; however, at present, we should bear in mind that calculating and comparing the APRS for a small group of nouns leaves us still very far from establishing whether countability of nouns is actually motivated by how the relative sizes of their grain-referents are perceived.

References

Berezowski, L. 1999. "To Count or Not to Count? A Fresh Look at Countability". In: Lewandowska-Tomaszczyk, B. (ed.) *Cognitive Perspectives on Language (Polish Studies in English Language and Literature 1)*. Frankfurt: Peter Lang, 163–175.

Bloomfield, L. 1933. *Language*. New York: Henry Holt.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge MA: MIT Press.

Joosten, F. 2003. "Accounts of the Count-Mass Distinction: A Critical Survey". In: Conenna, M. and É. Laporte (eds.) *Grammaires et lexiques comparés*. Amsterdam: Benjamins, 159–173.

Langacker, R.W. 1987. *Foundations of Cognitive Grammar I: Theoretical Prerequisite*. Stanford: Stanford University Press.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Radden, G. and R. Dirven 2007. *Cognitive English Grammar*. Amsterdam: John Benjamins.

Ware, R. 1979. "Some Bits and Pieces". In: Pelletier, F.J. (ed.) *Mass Terms: Some Philosophical Problems*. Dordrecht: D. Reidel, 15–29.

[5] In simpler terms — the more pebbles, flowers or peas there are, the smaller a single pebble, flower or pea appears to be.

[6] The APRS values are: peas 1.0, maize 0.5, wheat 0.44 (Table 3).

Wierzbicka, A. 1983. “Oats and Wheat: The Fallacy of Arbitrariness”. In: Heiman, J. (ed.) *Iconicity in Syntax*. Amsterdam: John Benjamins.
Wierzbicka, A. 1988. *The Semantics of Grammar*. Amsterdam: John Benjamins.
Wierzbicka, A. 1991. “Semantic rules know no exceptions”. *Studies in Language* 15. 371–398.