Quantifying Count/Mass Elasticity

keywords: mass nouns, abstract nouns, semantics, vector-space models

INTRODUCTION

In this paper we investigate the count/mass distinction from a large-scale distributional perspective. In particular, we address the elasticity of this contrast (Chierchia 2010), the fact that many nouns appear in both mass and count contexts (many hopes, much hope). Elasticity raises important questions about the formal status of the count/mass distinction (a type-based lexical contrast, or a token-based contrast in usage?) and spotlights the role played in this distinction by abstract, proper-name-like nouns.

In the vast literature on mass nouns, the following criteria, listed in Chierchia (1998), have played a key role in distinguishing mass from count uses: (i) determiners like much, less, a little, a bit of, more + N
sing are associated with mass uses; (ii) the determiners every, a, each select for count uses; (iii) ‘bare singulars’ in argument positions (water is wet) are mass uses; (iv) plural uses are count uses (modulo mass ‘pluralia tantum’, Acquaviva 2008); (v) there is a small set of more-or-less felicitous type-shifting operations between these classes: kind-formation: “many wines” ⇒ “many kinds of wine”); canonical portions: “I had two beers” ⇒ “I had two glasses of beer”; grinding: “I put apple in the salad” ⇒ “...apple pulp”.

Given the lexicalist assumption, one would expect that nouns could either satisfy (i) and (iii) and appear primarily, if not only, in the singular (=“mass”), or satisfy (ii) and appear in either numbers (=“count”). Surprisingly, this is not the case. Searches in UKWAC, a nearly 3-billion word corpus of English (Baroni and Ueyama 2006), revealed that nouns that occur in the mass contexts (i)/(iii) have rates of plural uses (0.25) similar to those of nouns appearing in count contexts (ii) (0.37). How can this be? An examination of the most frequent Ns that satisfy (i) reveals that most (arguably, around 70%) are abstract. Examples are:

(1) aggravation, boredom, candour, condensation, cynicism, dependence, disenchantment, doom, honesty, imagery, intricacy, mischief, morbidity, onus, periodicity, pornography, propaganda, recourse, sarcasm, sensitivity, supervision, ...

The fact that so many appear in the plural is unlikely to be due to a massive shift toward a ’subkind’-meaning, but rather to the effect of multiple, yet-uncharted meaning alternations. We searched for general patterns using distributional techniques.

VECTOR-SPACE DISTANCE AS A MEASURE OF NUMBER-INDUCED MEANING SHIFT

Vector Space Models (Turney and Pantel 2010) represent the ‘meaning’ of a content word in terms of the number of times it cooccurs with other content words in a corpus. These cooccurrences can be rendered as vectors; semantically similar words tend to appear in similar context, so their vectors will be ‘closer’ than those of words that appear in unrelated context. To study the effect of grammatical number, we built separate vectors for singular and plural nouns above a certain frequency in UKWAC, then checked to what extent the vectors for the two forms of the same word differed. The hypothesis was that the plural would trigger a detectable meaning/vector shift only with mass nouns. This was indeed the case: while for many nouns the vectors for singular and plural were close neighbours, this was not so for a set of nouns which significantly overlaps with nouns that satisfy (i). Experiments are ongoing to cluster the latter set of nouns into separate types of number-triggered alternations: stuff/kinds,
stuff/container, but also a category which we believe especially important for abstract entities:
name/instanciation:

**Modified indefinites as a probe for the “Abstract Names” Hypothesis**

Parallel pairs like “hope/hopes”, “fear/fears”, “law/laws”, suggest that many abstract nouns might be categorized (with Putnam 1973, 1975) as a type of proper names (e.g. names of psychological attitudes, which might be graded, thus satisfying (i)). The corresponding plurals could then be manifestations, or (event-related) instanciations of the entity named by the singular. Of course, the latter meaning could also surface in the singular, undermining the VSM number-based approach described above. To overcome this problem, we exploited the parallelism between (2) and (3):

(2) A *(smiling/confident/tired) Barak Obama entered the senate.
(3) A ??(solemn/great) peace descended on the land.

Modifiers seem to be crucial in triggering the possibility to extract time-bound manifestations of personal proper names. We propose that this is true for abstract names as well. Using the BNC, we tested the ratio between “a” N / “a” ADJ N, for ‘count’, then for ‘mass’ Ns. The result (1.71 vs. 0.84) confirms that the name/instanciation ambiguity is a plausible source for the high number of “plural masses” in English.

**References**