

Daniel Schmidtke and Victor Kuperman

Mass counts in the World Wide Web:
A corpus linguistic study of noun countability across varieties of English

This research uses a corpus of global Internet English to explore the variation in the pluralisation of mass nouns (e.g. *luggages*, *violences*, and *advices*) across varieties of English. The countable use of mass nouns is a lexico-grammatical phenomenon that is hailed as a discernible proxy of the dividing line between native and non-native varieties of English (McArthur, 2002; Mesthrie and Bhatt, 2008). A recent study (Hall, Schmidtke & Vickers, 2013) confirmed this, using Google searches of the World Wide Web as a corpus linguistic tool. They revealed a significantly higher concentration of countable usage of 25 mass nouns among non-native L2 English users, compared to native L1 users of British English. In the current study we employed a less restricted data-driven methodology, with which we were able to probe the extent of ‘mass noun’ countability across an expansive list of nouns, and among a greater variety of Englishes.

We queried Davies’ (2013) 1.9 billion-word corpus of Global web-based English (GloWbE), for the raw frequencies of 17,757 singular noun lemmas and their plural counterparts. GloWbE represents 20 samples of English from 7 native-English speaking, ‘Inner Circle’ countries (e.g. Canada, Britain and Australia) and 13 ‘Outer Circle’ countries (e.g. India, Hong Kong and Tanzania). We identified nouns that occurred significantly more frequently in plural form in the Outer Circle compared to the Inner Circle, irrespective of the atomic (countable/mass) quality of each noun’s referent. Once these nouns were isolated, we explored the underlying semantic and morpho-syntactic causes of the dissimilarity in countability preferences across Inner and Outer Circle Englishes.

Firstly, we found a significant convergence of the nouns used more countably in the Outer Circle with nouns that are routinely cited in the literature as grammatically ‘mass’. This list represents a widespread countable usage of mass nouns, such as *equipments*, *softwares* and *slangs*, in Outer Circle Englishes. These nouns were also part of Hall et al.’s (2013) pre-defined mass noun word-list. Secondly, we observed a large number of previously unattested cases of pluralisation of mass nouns, more commonly in the Outer than in the Inner Circle. Using Latent Semantic Analysis (Landauer and Dumais, 1997), a computational method of calculating semantic distances between words, we found that the categories of nouns that exhibit countability preferences in Outer Circle English are semantically predictable. For example, plural forms of nouns that denote non-individuated concepts reliably cluster into the semantic category of ‘occupational terminology’, such as: *assistances*, *trainings*, *welfares*, *recruitments*, and *taxations*. In addition, the following nouns represent the semantic category of ‘written language paraphernalia’: *alphabets*, *graphites*, *handwritings*, *mails* and *punctuations*. Furthermore, we note an overrepresentation of abstract nouns with Latinate morphology used countably in the Outer Circle, such as *acclaims*, *ascendants*, *destructions* and *servitudes*.

Taken together, our results provide further confirmation of the heterogeneity of noun countability behaviour across the Inner Circle and Outer Circle varieties of English. We adopt the notion that the observed variance in noun countability supports the conception of English as dynamic and plurilithic entity (cf. Pennycook, 2009). Moreover, we report an original finding that suggests that the semantic properties of the nouns inform the unconscious cognitive processes involved in marking nominalizable concepts with plural morphology. Our method was blind to the theoretically grounded grammatical count/mass distinction, and, given the pattern of our results, we propose that the fixed binary distinction

of count/mass is not an essential component in a theory of English language structure. Instead, it is a phenomenon best viewed as a gradient that is semantically and regionally dependent.

References

- Davies, M. (2013). Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries. Available online at <http://corpus2.byu.edu/glowbe/>.
- Hall, C. J., Schmidtke, D., & Vickers, J. (2013). Countability in world Englishes. *World Englishes*, 32(1), 1-22.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.
- McArthur, T. (2002). *The Oxford guide to world English*. Oxford: Oxford University Press.
- Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties*. Cambridge University Press.
- Pennycook, A. (2009). Plurilithic Englishes: towards a 3D model. *Global Englishes in Asian contexts: Current and future debates*, 194-207.