

Countability in world Englishes

CHRISTOPHER J. HALL,* DANIEL SCHMIDTKE† AND JAMIE VICKERS‡

ABSTRACT: In this study we explored variation in the countability of nouns in Outer Circle, Expanding Circle and lingua franca Englishes, a phenomenon which is frequently cited as a marker of Inner Circle norms in TESOL and of endonormative and emerging varieties in the Outer and Expanding Circles. We inspected a set of mass nouns like *information* and *equipment* in the VOICE corpus and websites from Outer and Expanding Circle country domains. We also evaluated potential causes of variation, investigating differences between Outer and Expanding Circles and the contribution of substrate influence. Our data show notable and widespread countable use of nouns that are generally non-count in Inner Circle Englishes, but such usage is highly infrequent overall. There appears to be greater variation in the Outer than the Expanding Circle, but little evidence of a determining role for substrate influence. We conclude that the prominence given to countability as a marker of ‘nativeness’ and ‘non-nativeness’ is unhelpful, in both the prescriptive context of TESOL and the descriptive contexts of world Englishes and English as a Lingua Franca. We advocate the use of web-based corpora to investigate lexico-grammatical variation in lingua franca usage and to reveal the ‘plurilithic’ nature of English.

INTRODUCTION

In the speech and writing of some users of Outer and Expanding Circle Englishes, mass nouns like *equipment*, *homework*, *advice*, and *furniture* are attested with countable grammar, including plural morphology (*advices*, *furnitures*) and number-specific determiners (*an advice*, *several furnitures*). This so-called ‘countable usage of non-count nouns’ is a much-cited lexico-grammatical feature in research on the resources of Englishes beyond the Inner Circle and their deployment within and between national varieties (Platt, Weber and Ho 1984; Williams 1987; Ahlu 1998; McArthur 2002; Seidlhofer 2005; Jenkins 2006; Schmied 2006; Björkman 2008; Crystal 2008; Kachru and Smith 2008; Mesthrie and Bhatt 2008; Kirkpatrick and Deterding 2011; Schneider 2011). It also receives a great deal of attention in materials for the learning, teaching, and testing of Inner Circle norms (e.g. MacKay 2002; Schoenberg 2005; Yates 2006; DeCapua 2008; Brook-Hart 2009; cf. also Quirk 1990: 8). In this study we attempt first to quantify variation in the ways that users deploy mass nouns, and then to evaluate potential explanations for this variation, investigating the difference between Outer and Expanding Circle samples and the role of cross-linguistic influence from substrate languages.¹

In the first decade of the twenty-first century, much scholarly debate on Englishes revolved around the pluricentricity of ‘the’ English language on the one hand, and the pragmatics of English ‘linguaging’ on the other, with the World Englishes paradigm

*York St John University, Centre for Languages and Linguistics, Lord Mayor’s Walk, York, YO31 7EX, UK. E-mail: c.hall@yorksj.ac.uk

†McMaster University, Togo Salmon Hall 605, 1280 Main Street West, Hamilton, Ontario L8S 4M2, Canada. E-mail: schmiddf@mcmaster.ca

‡SARA International Language School, 672–12, Jeonha-dong, Dong-gu, Ulsan City, South Korea. E-mail: jamie_v@hotmail.co.uk

concentrating on the former, and the newer field of English as a Lingua Franca (ELF) studies on the latter. The purpose and design of the research reported here are informed by thinking from both orientations. Like Kachruvian world Englishes, we explore the diversity of English around the world and continue to challenge the ‘apotheosis of the native speaker’ (Rajagopalan 1997: 229). And like ELF studies we move beyond national varieties to examine the nature of English in its most common contexts of use outside the Inner Circle: as a variable and dynamic set of linguistic resources exploited as they are needed in communication between speakers with different first languages.

Countability is an interesting feature to explore in this regard, because the countable use of some English nouns is perceived and presented as a salient marker of Outer and Expanding Circle usage, and yet it has no obvious effect on communicative effectiveness (e.g. Björkman 2008). In line with Widdowson’s (1994: 381) insight about the status of grammatical features as ‘social markers’, we can recognize countability as a purely formal shibboleth of the contested native vs. non-native speaker dichotomy (cf. Quirk 1990: 8; McKay 2002: 127; Higgins 2003: 640; Mollin 2007: 180–1). For example, Crystal (2008: 6) invokes countability as social marker in the following passage, part of a brief speculation on some of the future forms of English:

Some people might think [the countable use of ‘mass’ nouns] ‘un-English’, but in fact *informations* was in English once: *an information* and *informations* can be traced back to Middle English, and are found in Chaucer, Shakespeare, Swift, and many other authors. It may only be a matter of time before they are back.

We would argue that they have never gone away. English has never been a monolithic system of fixed forms, and multilingual or L2 users have been a part of its story from the very beginning. The notion of ‘plurilithic’ English (Makoni and Pennycook 2008) encompasses the social practices (and mental resources: Hall, forthcoming) of millions of individual users around the globe, who are all either multidialectal or multilingual, and whose individual versions of the language vary in form and function. A major component of scholarly efforts to problematize the ontological status of fixed varieties of English, and so reveal the plurilithic nature of the language, must, in our estimation, include analyses of the degree to which its forms vary (or conform) across users and uses. For applied linguists, this enterprise also has a practical purpose, informing decisions about the inclusion of worthwhile and attainable ‘targets’ in pedagogical materials and curricula. Given its prominence in the literature and yet its low communicative significance, countability is a suitable candidate for such analysis.

Our focus is on variation in the lexico-grammatical forms deployed and experienced by individuals in lingua franca scenarios, rather than within (more or less institutionalized) national varieties, as represented, for example, in the five Outer Circle corpora of the International Corpus of English (ICE 2010). Users of English in lingua franca scenarios are extremely diverse in terms of the circumstances in which their languages have developed, the breadth and depth of the functional repertoires they control, the extent of their multilingualism, and the kinds of linguistic practices they typically engage in. The use of the binary categories Outer and Expanding Circle Englishes is, therefore, an oversimplification, but a useful one if we wish to characterize the broad distinction between users who have been exposed to English, and have developed their knowledge of it, in primarily second language or foreign language contexts. Although we make use of the distinction

here, what characterizes the phenomenon under study is its concentration in the speech and writing of users beyond the Inner Circle, where lingua franca usage predominates and native-speaker norms are not the inevitable outcomes of acquisition. We used two very different sources of authentic data to tap this reserve of international usage: first, we examined the VOICE corpus of oral interactions constituting English as a Lingua Franca (VOICE 2009); second, we sampled webpages in English in the internet domains of 14 countries from outside the Inner Circle, using the Google search engine.

The paper is structured as follows. In the first section, we sketch the way countability works for speakers of English and other languages, describing some of the basic ground plans which could surface in English through substrate influence. We then address the nature and causes of variability in usage by users from beyond the Inner Circle, emphasizing the disproportionate attention paid to the phenomenon by general and applied linguists, as well as TESOL professionals. The second section reports data, first from the VOICE corpus and more substantially from Internet snapshots; we then discuss potential causes of variability, assessing the evidence for processes of innovation and/or nativization. The article concludes with a discussion of the data framed by our understanding of the plurilithic nature of English and the consequences of this for establishing and teaching language ‘norms’.

COUNTABILITY IN ENGLISHES AND OTHER LANGUAGES

In this section, we describe the phenomenon of countability as it is expressed in English and other languages. The kind of concepts routinely expressed by languages as nouns (henceforth ‘nominalizable concepts’) may be mentally categorized on various dimensions, including whether they are inherently *atomic* (particularized, individuated wholes) or inherently *mass* (non-particularized collectives). Solid, bounded objects (e.g. ‘leaf’ or ‘chair’) tend to be categorized as atomic, whereas nonsolid, dispersed (classes of) objects or ideas are more likely to be categorized as mass. Mass concepts include: (a) fluids and other indefinite collections of atoms (e.g. ‘water’ or ‘gravel’); (b) (kinds of) substance (e.g. ‘meat’ or ‘wood’); (c) generic concepts (e.g. ‘hair’ or ‘lightning’); and (d) qualities (e.g. ‘redness’ or ‘honesty’). The dichotomy is essentially one of ‘individuation’: nominalizable concepts are either taken as *individuals* which can be counted and therefore linguistically expressed as plural, or as *groups* or *masses* which would not normally be counted because they are unindividuated (inherently plural) to begin with.

For the most part, the way our minds categorize the ‘individability’ of nominalizable concepts is common across all human beings, independently of the language(s) we acquire in infancy (cf. Xu 2007; Barner, Inagaki and Li 2009). Furthermore, the conceptual categories we construct internally are externalized by different language systems using a predictable and quite well understood set of grammatical devices (Chierchia 1998; Aikhenvald 2003). But the semantic mapping from concept to form is not uniform across all language systems, and the grammatical devices used for this mapping are not universal. Language systems exploit two basic designs for expressing individuation:

1. *countability* via number morphology (e.g. *page* vs. *pages*); and
2. *non-countability* via classifiers (e.g. *paper* vs. *pieces of paper*).

English, Swahili and Sinhala are languages which (mostly) rely on countable grammar; Mandarin Chinese, Japanese, and Filipino use non-countable grammar. No language

system uses the countability option exclusively, whereas many language systems employ noncountability, and in fact it appears to be the unmarked option cross-linguistically (Allan 1980; Chierchia 1998).

For users of Inner Circle (henceforth IC) Englishes, most nominalizable concepts (independently of individuability) are expressed as count nouns and are morphologically marked as singular or plural, with determiner modification required in some contexts. There is, however, a group of nouns which have limitations on their accessibility to count grammar. The group is a tiny subset of the set of all English nouns, but its membership is larger than similar groups in other non-classifier languages. Spanish, for example, has singular and plural forms for *advice* and *furniture*, and Sinhala expresses many of the same nominalizable concepts, including *advice* and *furniture*, *only* in plural form (so-called *pluralia tantum*, like *entrails* in English). Although countability undeniably plays a significant role in determiner selection in English (e.g. *this information* vs. *these informations*), the small class of nouns which are non-count in IC Englishes actually constitutes a very negligible element in the lexico-grammatical paraphernalia of the language. Indeed, the sections dedicated to countability in Quirk, Greenbaum, Leech and Svartvik's (1985) painstaking description of IC ('Standard' British) grammar take up fewer than seven out of over 1,600 pages (less than half a per cent). Moreover, the group has leaky borders. According to Allan's seminal (1980) study, countability in (IC) English(es) is best understood as a continuum on which nouns may be placed according to their 'countability preferences', rather than in terms of a binary [+count] or [-count] feature marked on each noun.

Mass nouns in the Englishes of multilingual learners and users

Despite its knotty and negligible status, the count/non-count distinction has a strikingly high profile in descriptions of the Englishes taught to and used by native speakers of other languages. It figures especially prominently in mainstream English teaching and testing materials and practices. Among textbooks, for example, it features very early on in Pearson's *Focus on Grammar (Basic)* (Schoenberg 2005). It is listed as the locus of one of the top ten EFL 'mistakes' in Brook-Hart's (2009) *Learning from Common Mistakes* booklet, based on the Cambridge Learner Corpus of exam papers (and billed as part of its 'Real English Guarantee'). The popular www.onestopenglish.com website provided by the publisher Macmillan supplies five intermediate level lesson plans for it (out of 74 overall) – more than for word order or modals – and it is listed as the most popular grammar item in a sidebar. Yates's (2006) *English Vocabulary for Beginning ESL Learners* actually starts out with a very challenging introduction to count and mass nouns (see below). For the professional development of teachers, too, it is seen as worthy of considerable attention: DeCapua's (2008) *Grammar for Teachers*, for example, dedicates 12 pages to it, and Jacobs' (1995) *English Syntax. A Grammar for English Language Professionals* has almost a whole chapter on the subject.

In Quirk's (1990: 8) impassioned defence of native-speakerism, his notorious reference to 'half-baked quackery' was levelled against an English teacher who questioned why his students' use of *several informations* should necessarily be corrected. The need to vigorously teach IC norms for such nouns is taken for granted in most ESL and EFL materials and curricula, despite the marginality of the distinction and its considerable intractability. The complexity revealed by Allan (1980) and apparent too in Quirk et al.'s (1985) publicly authoritative treatment does not appear to have daunted practitioners, nor

prevented many of them from assuming that learners need early mastery of it. An indication of the heightened importance associated with it in resources for learners is the attention it is allocated in grammars for teachers: Carter and McCarthy, in their (2006) grammar for learners and teachers, dedicate over twice as much space to countability as Quirk et al. (1985). But the pedagogical response in EFL and ESL teaching materials on the topic is not distinguished by its clarity and accessibility to learners. Consider, for example, the following passage from the very first page of Yates' *English Vocabulary for Beginning ESL Learners* (2006: 1), which follows an initial introduction to the count/non-count distinction.

Other nouns cannot be counted – *air*, *wind*, and *pollution*, for example. They have no plural forms, are used with singular verbs, and are called “noncount” nouns. But noncount nouns can also be things that we can count! First, there are those that it would take a lifetime to count, so we call them by a more general noncount noun, such as *hair*, *sugar*, or *flour*. And then there are those that we categorize into general groups that are named by noncount nouns, such as *furniture*, *mail*, *silverware*, and *china*. Of course we can count *chairs*, *tables* or *beds*, but the general category *furniture* is never made plural. The noncount noun *mail* includes the *letters* and *cards* that we can count. English has a lot of these words.

The explanation mixes statements of ‘fact’ (‘cannot be counted’, ‘is never made plural’), with more homely descriptions of what ‘we’ (as IC owners of English) do (‘we call them by a more general noncount noun’; ‘we can count *chairs*’; etc.). Students beginning the book must feel some trepidation about the learning journey ahead when confronted so early in the process with this unequivocal, but somewhat random, characterization of IC English (cf. Kachru and Smith 2008: 90).

Other TESOL professionals simplify the challenge, but still recognize that it is a significant one for learners. The following is taken from a live chat transcript of a ‘grammar surgery’ offered by the BBC World Service Learning English website (BBC World Service 2004):

[Learner]: How can we know which noun is countable or uncountable? Is there a rule?

[‘Grammar Masterclass’ Expert]: Most nouns are either countable or uncountable. It is logical which is which. For example, clearly we can count books so the word ‘book’ is countable. On the other hand, we can’t count ‘water’ so the word ‘water’ is uncountable. Unfortunately, though, there are a few words where this distinction isn’t so clear. ‘Furniture’ is a good example of this. We can’t say ‘I bought some furnitures’ because ‘furniture’ is uncountable. However, lots of learners of English think (rightly!) that this is rather illogical.

But whether it is logical or not, it is part of the monolithic system that most learners are expected to internalize, even when there is no evidence that it contributes to communicative effectiveness. Jenkins (2006: 44) signals the injustice of assuming a monolithic account of IC countability for learners:

[A] candidate in an ELT speaking exam would be rewarded for their knowledge of ‘real’ English if they were to say ‘three teas’ or ‘two coffees’ instead of ‘three cups of tea’ or ‘two cups of coffee.’ On the other hand, if they extended this use of uncountable nouns to ‘wine’ and referred to ‘two wines’ instead of ‘two glasses of wine’ they could be penalized for lack of competence with the countable/uncountable distinction.

In confirmation of this, Lowenberg (1992: 116) discusses items in the Test of English for International Communication in which test-takers must indicate counted 'non-count' nouns (like *resistances* and *equipments*) as ungrammatical, although they 'may well be acceptable to educated speakers of Malaysian or other non-native varieties of Standard English'. A similar point is made by Bamgboṣe (1998: 4) for West African English and by Shim (1999) for codified Korean English.

In studies of the Englishes of users in the Outer Circle (henceforth OC), the so-called 'countable usage of non-count nouns' is widely cited as a distinctive feature, especially in African and Asian varieties (e.g. McArthur 2002; Schmied 2006; Mesthrie and Bhatt 2008; Schneider 2011). Platt et al. (1984: 50) observed that 'nouns are sometimes marked for plural in the New Englishes where they would not be marked in the established varieties of English', and they attribute this to the 'reclassification' of uncountable nouns as countable. Williams (1987: 171) states that 'this [count/non-count] distinction frequently does not follow the rules of NS varieties'. For Schneider (2011: 204), the 'pluralization of mass nouns' is listed as a 'fairly common' characteristic of New Englishes, which is 'widespread in Africa and Asia'. Mesthrie and Bhatt (2008: 53) state that '[a]lmost every study of individual WE varieties in Africa and Asia reports frequent examples like *furnitures*, *equipments*, *staffs*, *fruits*, *accommodations*, and less common ones like *offsprings*, *underwears*, *paraphenalias* [sic], etc.' Kirkpatrick and Deterding (2011: 378) also claim that the 'occurrence of *furnitures* and similar words' is 'widespread' in OC Englishes.

Some scholars go slightly further. Ahulu (1998: 19–20), in a paper focusing on number marking on nouns and in verbal concord, highlights countability as 'the most notable and regularly cited' among 'major area[s] of divergence between British Standard English and the English written in postcolonial countries'. Schmied (2006: 198) claims that 'East African usage basically ignores the grammatical distinction of count vs. non-count nouns [...]'. Kachru and Smith (2008: 106) refer to 'the extensive use of collective nouns as countable in almost all OC varieties'.

For Expanding Circle (henceforth EC) Englishes and English as a Lingua Franca (ELF), in which Englishes from all three circles might be represented, the phenomenon has also been noted as a salient characteristic. Seidlhofer's (2005: R92) earlier code-oriented work on ELF includes the 'plural use of mass nouns' as one of eight lexico-grammatical tendencies. Jenkins (2006: 44) states that forms such as *a staff* and *four furnitures* 'are used by many speakers of the Expanding Circle' and that processes of regularization in ELF are likely to lead to features such as 'the countable use of nouns that in ENL are considered uncountable (e.g. *informations*, *advices*)' (Jenkins 2009: 201). Shim (1999) cites the use of 'non-count nouns as count nouns' as one of a dozen morpho-syntactic features which distinguish Korean English from US English. Mollin (2007) includes items such as *informations* and *equipments* in an acceptability test for users of Euro-English. And Björkman's (2008) study of ELF usage in Swedish tertiary education identified 'incorrect plural forms/countability' as one of three morphological features.

In sum, the so-called 'countable usage of non-count nouns' is portrayed as one of the most salient characteristics of Englishes beyond the IC. On the one hand it is viewed as an error to be prevented or corrected as early as possible in the learning and teaching of IC English, and on the other is presented as a common feature of new endonormative varieties of English in the OC, of emerging varieties in the EC, and of the formal resources deployed

in ELF interaction. Before examining samples of English from beyond the IC to assess the *prevalence* of the phenomenon, we address the question of its *causes*, and in so doing begin to question its status as non-IC variety marker.

Variation in countability and its possible causes

As we have noted, scholars (notably Allan 1980) have demonstrated that countability does not work in a binary fashion in the grammars of even monolingual IC users. This fact has been acknowledged and highlighted by researchers on world Englishes. Sey (1973), who documented the phenomenon in Ghanaian English, observed:

There appears to be (a) no consistent semantic relationship between countable and uncountable uses of nouns [in IC Englishes], nor (b) any clearly discernible motivation for using some normally uncountable nouns in countable functions but not others. (cited in Bokamba 1992: 131)

Schmied (2006: 198) echoes this observation, stating that the grammatical distinction ‘does not always correspond to the semantic distinction’ and that ‘in [IC Englishes] some non-countables may occur in the plural in special meanings (e.g. *works*)’, concluding that ‘thus differences are often a question of interpretation and frequency’. This lack of transparent semantic mapping and the absence of a simple binary count/non-count distinction in IC Englishes are at the heart of Allan’s (1980) conclusion that countability is a continuum upon which each noun’s place must be assessed by grammaticality judgements on an item-by-item basis.

If the countability features that IC users construct for nouns in their mental lexicons and grammars is, to repeat Schmied’s words, ‘a question of interpretation and frequency’, then we should not be surprised to find variability also *between* users and learners in the OC and EC, given that their mental lexicons and grammars are ultimately rooted (either historically or pedagogically) in IC usage. Mesthrie and Bhatt (2008, 53–4), signal the interesting possibility that some OC users might maintain countable forms that were present in the IC Englishes to which speakers were exposed in the colonial past, but that are now non-count according to IC norms. In a corpus of Settler English from early nineteenth century South Africa, Mesthrie and West (1995) found, among other variable structures, ‘plural endings for non-count nouns like *progresses, evidences, sufferings*’ (cited in Mesthrie and Bhatt 2008: 191). Further to Crystal’s (2008) observation, cited earlier, that *informations* has a long history in IC English, Toyota (2009) suggests that the count/non-count distinction is a relatively recent development in the history of English, and indeed that Old English and Early Middle English were essentially classifier languages. In the light of this intrinsic variability, we recommend dispensing with the ‘native-speakerist’ nomenclature of *count noun* vs. *non-count noun* and their implication of a monolithic binary typology. Here we use the more neutral term *mass noun* for that subset of nouns which for IC users are (normally) non-count in the relevant grammatical contexts but which for OC and EC users may not be – the group of nouns that Allan (1980: 560) calls ‘true uncountables’.

The extent to which the idiosyncrasy of IC usage has influenced the usage of mass nouns in the Englishes of other regions remains to be seen. But it is almost certainly only one factor of several at play. Platt et al. (1984: 51–2) include it along with three other factors influencing OC users’ ‘reclassification’ of non-count nouns as countable (cf. also Mesthrie and Bhatt 2008: 53–4):

- Countable semantics: Some mass nouns can be taken to refer to ‘separate items, so that *furniture* means *an item of furniture*’.
- Morphological resemblance: Some mass nouns are morphologically related to count nouns (e.g. *jewellery* – *jewels*).
- Substrate influence: Some mass nouns are treated by ‘background languages’ (substrates) as countable (although they note that the tendency is not absolute).

A fifth factor is one that applies to a large number of lexico-grammatical (and other) processes in English: the tendency towards simplification, including regularization, which is attested in both first and second language acquisition (cf. Slobin 1973; Williams 1987). This factor would lead learners to map all nominalizable concepts onto countable nouns, without exception. Of these factors, those related to semantics and substrate are of particular interest, because they will vary in their effects across nouns and speaker groups, and so are more likely to reveal the processes of innovation and nativization which are hypothesized to underlie the emergence of new norms in specific contexts (Bamgboṣe 1998). We explore their role in our own data at the end of the next section.

ASSESSING THE SCOPE OF THE PHENOMENON

There are almost no data on the extent of variability in the grammatical deployment of mass nouns. Given the importance attached to countability as an indicator of normative use, both in English teaching guides and scholarly descriptions of World Englishes, we decided to assess just how extensive the countable usage of mass nouns is in the authentic practice of users in lingua franca contexts. We began our search with the Vienna Oxford International Corpus of English (VOICE 2009) but, as reported below, the data scarcity of VOICE led us to conduct an expanded search using snapshots of the world wide web, and it is this second source that provides most of the data reported in the following sections.

Countability in ELF

Mollin (2007) is the only study we are aware of which attempts to quantify the countable usage of mass nouns outside the IC. As part of an investigation into the status of Euro-English, representing mostly EC speakers, she constructed a 400,000-word corpus of English usage in non-Anglophone Europe, composed of about 60 per cent transcriptions of spoken data from European Commission public discussions, speeches, etc., and 40 per cent written data from chat forums and online discussion groups, with over 900 speakers represented overall. She found little evidence for widespread countable usage of mass nouns. She calculated proportions of countable uses for around 40 mass nouns, including 27 mentioned in Swan (1995) and 13 *pluralia tantum* (Mollin, personal communication). The noun with the highest proportion was *bread*, at 18.75 per cent, followed by *advice* at 11.86 per cent and *evidence* at 9.68 per cent. Mollin (2007) points out that not only are these proportions low, but also that countably used mass nouns are unrepresentative of the full range of speakers and texts in the corpus. She notes, for example, ‘[a]ll three instances of *bread* used as a countable noun stem from the same Finnish speaker in the same text, and two of the three countable uses of *evidence* come even from the same sentence’ (Mollin 2007: 179). This pattern is repeated for other nouns, and overall, the average rate of countable usage of the mass nouns examined was 2.54 per cent.

Mollin's study has been criticized on both methodological and conceptual grounds by Seidlhofer (2009). One problem she identifies regards the representativity of the genres sampled, for example, the emphasis on public debate, which is 'very unlikely to support a natural vernacular way of speaking' (Seidlhofer 2009: 46). For a more 'natural vernacular' sample we chose Seidlhofer's VOICE corpus to begin our quantitative assessment of mass nouns in non-IC usage. Conceptually, Mollin is criticized by Seidlhofer for adopting an overly 'formalist' orientation, which prevents her from problematizing the traditional ENL/EFL/ESL typology. Like Mollin, we are unapologetically 'feature spotting' here, but this does not represent for us a blinkered 'fixation on form', as we hope to show in our concluding discussion.

VOICE contains just over one million words from speech events involving 694 L2 speakers (and 57 L1 speakers) in mostly professional contexts. Fifty one languages are represented, although 87 per cent of them are Indo-European and just over a third of the words in the corpus are spoken by L1 speakers of German (25%) and Dutch (10.5%) (VOICE 2009). To generate a sample set of mass nouns to search for, we compiled a (not exhaustive) master list of 183 potential items on the basis of listings of nouns labelled as 'mass', 'uncountable' or 'non-count' in grammar books, websites, and scholarly articles. For each master list item, we conducted a search for singular and plural stems and recorded countable usage when the noun was situated in one of the following contexts (examples are from VOICE):

1. occurrence with plural *-s*, for example:

- (a) [...] *and help them for their homeworks* [...]
- (b) [...] *the way of reaching informations* [...]

2. preceding indefinite article or numeral *one*, for example:

- (a) [...] *just one luggage* [...]
- (b) [...] *shall I just buy a milk quickly* [...]

3. preceding ordinal number, for example:

- (a) [...] *the third advice again comes from our xx* [...]
- (b) [...] *our fo[u]rth advice will be to monitor international currency flows* [...]

Of the 111 types occurring in VOICE, only 19 occurred with instances of countable grammar in a total of 52 tokens used by 45 speakers. Table 1 shows a summary of countable usage of mass nouns in VOICE.

Among mass nouns frequently cited in the literature that had zero countable usage in VOICE are *staff*, *baggage*, *furniture*, *evidence* and *equipment*. Inspection of contexts of use indicate that of the 111 master list noun forms attested, 31 appeared in contexts in which non-countable grammar would have been inappropriate (because of polysemy in which a non-mass sense was being expressed), resulting in 80 (111 minus 31) plausible mass noun tokens to count. This means that, in effect, fewer than 25 per cent of the types investigated were used countably (19 of 80), and that of these types, countable tokens accounted for around 2.9 per cent of all occurrences. Users of mass nouns with countable grammar were L1 speakers of 15 different languages, most of them European.

We may conclude from this that countable usage of mass nouns is uncommon in lingua franca interaction between mostly European EC speakers. Our finding (2.9%) is of the

Table 1. Mass nouns (n = 19) used countably at least once in VOICE, ordered by number of countable uses

Mass noun	Tokens n	Noncountable uses n	Countable uses		Speakers n
			n	(%)	
<i>advice</i>	30	23	7	23.3	5
<i>information</i>	407	400	7	1.7	6
<i>money</i>	473	467	6	1.3	4
<i>knowledge</i>	201	196	5	2.5	4
<i>employment</i>	88	84	4	4.5	4
<i>research</i>	136	132	4	2.9	4
<i>paper</i>	27	24	3	11.1	3
<i>traffic</i>	18	15	3	16.7	3
<i>luggage</i>	6	4	2	33.3	2
<i>violence</i>	24	22	2	8.3	1
<i>harm</i>	2	1	1	50.0	1
<i>homework</i>	14	13	1	7.1	1
<i>lack</i>	78	77	1	1.3	1
<i>luck</i>	30	29	1	3.3	1
<i>milk</i>	35	34	1	2.9	1
<i>permission</i>	10	9	1	10.0	1
<i>public</i>	70	69	1	1.4	1
<i>reading</i>	9	8	1	11.1	1
<i>stuff</i>	152	151	1	0.7	1
Total	1810	1758	52	2.9	45

same order as the 2.5 per cent found in Mollin's study. But despite the low rate of countable usage, it is still notably above the norm for IC usage, which, according to descriptive linguists, is for non-count grammar across all mass nouns in the contexts inspected. To confirm this from actual user data, we obtained from the British National Corpus (BNC) the frequencies of plural usage for 16 of the 19 mass nouns occurring in VOICE (we omitted *paper* and *reading* because of high plural usage in non-mass semantic contexts, and *permission*, because of its specialized plural usage in publishing). The overall rate of countable usage for this sample in the BNC was just above zero, at 0.3 per cent.² A *t*-test revealed that the difference between the VOICE and BNC rates was highly significant ($t(15) = 2.779, p < 0.01$).

It would, of course, be reckless to make generalizations about countability in non-IC users on the basis of these data alone. Because of the heightened degree of mobility and interaction with native speakers that European standards of living and location bring, one might expect the kind of professional discourse used by the educated Europeans of Mollin's corpus and much of the VOICE corpus to be more likely to conform to IC norms. This means that the extent of countable usage of mass nouns in lingua franca English may be seriously under-represented by these results. Although corpora like VOICE are rich in contextual detail, 'data scarcity' is one of the inevitable consequences, and this appears to limit their usefulness for frequency studies of lexical particularities like countability. We therefore sought a more data-rich sample of Englishes in lingua franca usage, in order to maximize the pool of types and tokens we could analyse and so obtain a wider-angled snapshot of the prevalence of countable mass nouns.

Counting nouns in the world wide web

The world wide web is immensely fertile terrain in which to encounter locally-determined but globally-directed Englishes produced by speakers of international varieties and in ELF contexts (cf. Gupta 2006). Although only some English language text on OC and EC websites is produced in online interaction (in chat forums, for example), a great deal of it is written for and/or read by users who do not share the native language of the writer. To this extent, it can be claimed to constitute a kind of lingua franca English, even if most of it is not necessarily ‘crafted ELF’, displaying the kinds of deliberate accommodation or interactive negotiation of meaning represented in VOICE. Moreover, the web as corpus is staggering in its size, dwarfing even the biggest purpose-built language corpus: according to current best estimates, there are over half a billion primarily English-speaking users of the Web, 42 per cent of all English users world-wide, responsible between them for at least 28 trillion English words (Norvig 2007; Internet World Stats 2010). Indeed, these statistics hugely underestimate English usage because they make the grossly simplifying assumption of global monolingualism, that is, attributing only one ‘official’ language to each nationality.

Like VOICE, the web contains much ‘natural vernacular’ usage; but unlike VOICE, it is not a structured, sampled collection of text. This, we argue, makes it a particularly attractive site to search for variability in the use of mass nouns. Much of the early work using the web as a collection of English text was conducted by researchers in the natural language processing community, who have consistently sought ways to eliminate the ‘noise’ in the system that ELF and world Englishes scholarship is keen to reveal. For example, in order to avoid ‘random noise caused by misspelled words, non-native speakers, pages in other languages, etc.’, Villavicencio (2003: 9) searched the web for verb-particle constructions in contexts constrained by descriptive accounts of IC norms, rather than simply searching for the verb and particle forms alone. And from an even more unequivocally prescriptivist standpoint, Fletcher writes:

As the lingua franca of the digital frontier, English is both the target and source of contamination: non-Anglophones often translate their web pages into Info-Age pidgin English while fusing creolized web English into texts in their native tongue. Similarly, searches for the linguistic examples [sic] can lead to work by learners with imperfect mastery of the language or to baffling machine translations. In many online forums, careless or cryptic language and sloppy spelling prevail. With its frenetic pace of development, the web typically values content creation above perfection and tolerates ill-formed language. (Fletcher 2007: 36).

We chose the web as a database precisely to hear some of this so-called ‘noise’ and ‘ill-formed’ language: just like ELF, it is transnational, un-normed and dynamic, a major site of Jacquemet’s (2005) ‘transidiomatic practices’. As such it provides an ideal search space for an assessment of the scope of countable use of mass nouns in individual Englishes outside the IC.

We cannot offer a very refined analysis here and, given the numbers of pages sampled, make no attempt at qualitative analysis. A country code on a url is no guarantee of provenance, let alone authorship, so provides no information on the nationality of, or language(s) known and/or used by, the author; nor whether the text on the page is written by one or many authors. Indeed, just as many pages in the domains of IC countries will contain text written by OC and EC speakers, so too will there be significant amounts of text

authored by IC speakers on OC and EC country domains. And while the web may be a key locus of global Englishes in lingua franca usage, in its unmediated and unfiltered state it presents other serious problems for the investigator, including questionable text authenticity (e.g. ‘spamdexing’ and ‘keyword stuffing’ for commercial purposes), volatile data shifts (webpages added, deleted, modified, and copied), and very limited options for setting search criteria (e.g. there is no part of speech tagging). Despite this intractability, snapshots of the web have been demonstrated to be very effective for assessing the frequency of language strings across millions of speakers in multiple genres: research shows that the number of ‘hits’ for pages containing particular strings correlates very well with BNC frequencies for those strings (Keller and Lapata 2003) and indeed that the web outperforms corpora for natural language processing simulation purposes (Lapata and Keller 2005). Although specialized tools like WebAsCorpus.org and WebCorp can now significantly refine the searches performed by commercial search engines like Google, Bing and Altavista, at the time of data collection we determined that Google would be adequate for our purposes, given that what we wanted was a wide-angled, quantitative snapshot of the presence of pluralized mass nouns on OC and EC Internet pages.

METHOD

We used the advanced settings of the Google search engine to find pluralized mass nouns in pages tagged as English in sites hosted by 14 different polities, using their country codes. Bergh (2005: 43) has highlighted the recognition by corpus linguists that ‘while it is generally difficult to use the Web in its entirety for frequency studies [. . .], slices of it in terms of domain-specific searches are more rewarding as they lead to a higher level of precision[.]’ Renouf (2003: 49–50), for example, suggests (for opposite purposes to ours) that using the *.uk* country domain ‘might limit the retrieval of non (English)-native-speaking text and word use’. We limited our searches to specific country domains from the Outer and Expanding Circles, with the language domain set to English, precisely to tap into the Englishes of multilingual users in lingua franca scenarios. Searches were performed in the English-tagged internet domains of 13 nations plus the Hong Kong Special Administrative Region. We also performed a search on the *.uk* domain as an IC baseline. The 14 non-IC domains searched were: Outer Circle – Bangladesh (*.bd*), Belize (*.bz*), Hong Kong (*.hk*), Kenya (*.ke*), Malta (*.mt*), Philippines (*.ph*), Sri Lanka (*.lk*); Expanding Circle – Angola (*.ao*), China (*.cn*), Iceland (*.is*), Mexico (*.mx*), South Korea (*.kr*), Thailand (*.th*), Yemen (*.ye*).

The selection was intended to represent a broad sample of English-using domains from the OC and EC (seven from each). Given our interest in potentially tracing cross-linguistic influence from speakers’ L1(s) and other national languages, we strove to select countries in which there was a numerically-dominant first language (or language family) with which IC non-count grammar norms and/or preferences could be contrasted. Most domains had a clearly dominant L1, at over 70 per cent of speakers according to *Ethnologue* (Lewis 2009). The three major exceptions were: Belize, with Spanish at 35 per cent; the Philippines, with Filipino (including Tagalog) at 46.5 per cent; and Thailand, with Thai and related Tai languages at 46.6 per cent. Some domains (e.g. Kenya and Angola, Belize and Mexico, Sri Lanka, and the Philippines) were included because their principal L1s used grammatical number in ways which contrast significantly with IC Englishes. Others (Bangladesh, Hong

Table 2. Country data. Language data from *Ethnologue* (Lewis, 2009); internet data from Internet World Stats (2010)^a.

	Country	Principal L1				Internet penetration (%)
		Language	n	(%)	Type ^a	
Outer Circle	Bangladesh	Bengali	110m	77	Class	0.4
	Belize	Spanish	0.1m	35	Num	19.5
	Hong Kong	Cantonese ^b	6m	91	Class	69.9
	Kenya	Bantu	16.8m	60	Num	8.6
	Malta	Maltese	0.3m	71	Num	59.7
	Philippines	Filipino ^c	46.5m	46	Neither	35.1
	Sri Lanka	Sinhala	15.5m	80	Num	5.5
Expanding Circle	Angola	Bantu ^d	11.9m	99	Num	3.4
	China	Mandarin	840m	69	Class	28.7
	Iceland	Icelandic	0.2m	100	Num	100
	Mexico	Spanish	86m	93	Num	24.8
	South Korea	Korean	42m	100	Class	77.3
	Thailand	Thai ^e	46.6m	90	Class	24.4
	Yemen	Arabic	15.5m	99	Num	1.8

Notes:

^aClass = classifier language; Num = number morphology.

^bFigures from Government of Hong Kong SAR (2007).

^cIncludes Tagalog.

^dVarious Bantu languages.

^eIncludes related Tai languages.

Kong, China, South Korea, the Philippines and Thailand) were included because their principal L1s are classifier languages. We also strove to include countries from all global regions, with East and South-East Asia particularly well represented because of their reported widespread countable use of mass nouns in English, as well as their high Internet penetration and/or large number of webpages. Some domains (Bangladesh, Sri Lanka, Yemen, Angola, Kenya) have very low internet use, and so of course the number of tokens in our database is low for them compared with other domains, and is less representative of the nation as a whole than, say, the figures for Iceland or South Korea, where internet use is an element of most or all citizens' cultural capital. Table 2 summarizes the country data.

The master list of mass nouns was expanded to 215 from the 183 used in the VOICE search, and although still inevitably incomplete, provided a sufficiently representative pool of words from which to identify a subset of Allan's (1980) 'true uncountables' to search for in Internet domains. Using Allan's rankings of countability preferences, a subset of 41 potential candidates for search was isolated from the expanded master list and each item was scrutinized for syntactic (e.g. word class) and semantic (e.g. homographic or polysemic) ambiguities. Numerous candidate nouns had to be eliminated on these grounds because of the likelihood that related forms which were not mass nouns would be included in the count. For example, syntactically, *research* is a common verb as well as a noun, and so instances with a suffixed *-s* may be third person singular verb forms as well as plural noun forms. Semantically, *paper* is a mass noun when it refers to the material or a blank sheet, but is a count noun when it refers to written documents (including newspapers).

The candidate list was thus narrowed down to 25 mass nouns:

- | | | |
|----------------------|------------------------|----------------------|
| 1. <i>advice</i> | 10. <i>feedback</i> | 19. <i>luggage</i> |
| 2. <i>applause</i> | 11. <i>fun</i> | 20. <i>magic</i> |
| 3. <i>baggage</i> | 12. <i>furniture</i> | 21. <i>slang</i> |
| 4. <i>cash</i> | 13. <i>hardware</i> | 22. <i>software</i> |
| 5. <i>corruption</i> | 14. <i>homework</i> | 23. <i>traffic</i> |
| 6. <i>dew</i> | 15. <i>information</i> | 24. <i>underwear</i> |
| 7. <i>employment</i> | 16. <i>jewellery</i> | 25. <i>violence</i> |
| 8. <i>equipment</i> | 17. <i>knowledge</i> | |
| 9. <i>evidence</i> | 18. <i>luck</i> | |

For each mass noun, separate searches were conducted in each country domain for the bare stem and the stem plus plural *-s*, using the ‘exact wording or phrase’ option in the advanced search pane of the Google search engine.³ The total number of hits per mass noun per domain was recorded and, to ensure representativity, figures lower than 100 tokens per domain were not included in subsequent analysis.⁴ The proportion of plural tokens of each noun type was calculated, yielding an approximate indicator of countable usage of the mass noun in the Englishes of the domain searched. Averages were then calculated across the sample set of the domain, as an approximate indicator of overall countable usage of mass nouns for the domain.

FINDINGS

The average percentage of countable usage across all nouns and country domains was 2.45 per cent, a figure which is remarkably consistent with the 2.54 per cent reported in Mollin (2007) for Euro-English and the 2.9 per cent we found in the analysis of VOICE data. This figure confirms the low estimates of the earlier studies, suggesting that countable usage of mass nouns is an infrequent characteristic of Englishes in OC or EC domains. It is, however, significantly higher in frequency than IC usage, as demonstrated by the very low rate for the .uk domain and BNC. For the .uk domain, the rate for the sample was just over 0.01 per cent and for the BNC, 0.02 per cent. The difference between these rates and those for our OC and EC sample are extremely significant according to *t*-tests performed on the means ($t(24) = 5.166, p < 0.00001$ for .uk; $t(24) = 4.669, p < 0.0001$ for BNC). Individual nouns and individual domains (reflecting national varieties) yield some modest but marked departures from the relatively low rate of OC and EC usage, as can be observed in Figures 1 and 2.

The high rate for the Philippines is due in large part to the behaviour of two nouns: *luggage* at a solid 87.92 per cent of countable usage (270,000 instances out of 307,100 in total) and *homework* at a substantial 41.61 per cent (13,400 instances out of 32,200). In order to avoid skewing any estimations of central tendency for the whole class of mass nouns in subsequent discussions, we replace the counts for these two words with the mean count for the Philippines domain. Aside from these two words, *equipment* shows consistently higher rates of plural usage across all domains except Iceland and Malta, with an overall mean of just over 9 per cent. Interestingly, *equipments* was judged ‘acceptable’ by just over 50 per cent of over 400 European academics responding to an acceptability test used as part of Mollin’s study of European ELF (Mollin 2007: 180–1). In sum, our

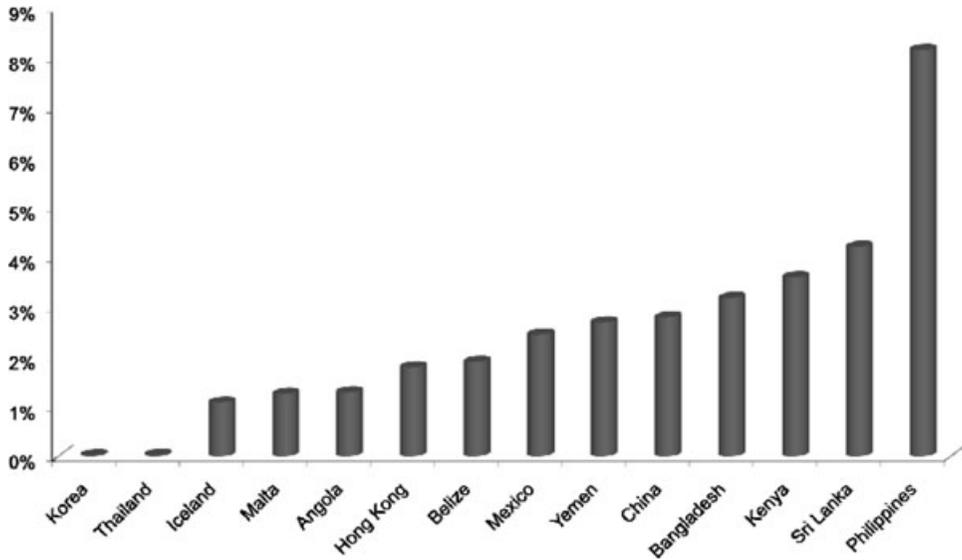


Figure 1. Countable usage by country domain, lowest to highest.

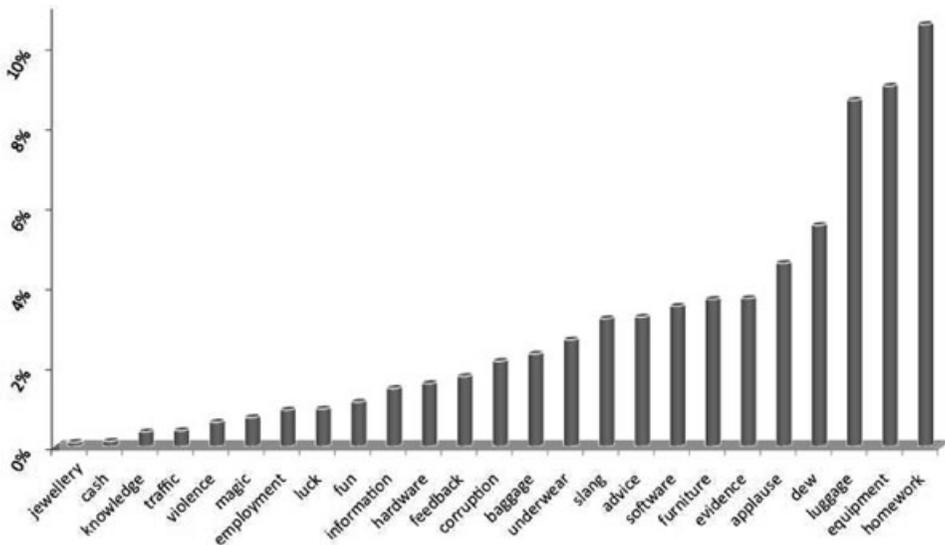


Figure 2. Countable usage by mass noun, lowest to highest.

figures suggest that, in the speech of (mostly) European ELF users and the written English of a diverse range of OC and EC Internet domains, the countable use of mass nouns is: (a) *infrequent*, even for the most variable nouns; but (b) *recurrent* compared with IC usage, where it is almost entirely absent; and (c) *widespread*, attestable across numerous L1 backgrounds and geographical regions.

Investigating causes

In an attempt to identify potential causes of the patterns of occurrence identified, we tested a couple of specific hypotheses regarding language change and language acquisition. These factors have been identified by previous scholars as potentially accounting for the countable use of mass nouns by OC speakers (e.g. Platt et al., 1984: 51–2; Gramley 2001: 119–20) and play a major role in processes of standardization/institutionalization or norm-fixing (cf. Kachru 1992, on ‘deviations’ and Bamgboṣe 1998, on ‘innovations’), but hitherto have not, as far as we know, been specifically tested.

First, we considered the status of English in the domains sampled, using the simplifying binary distinction *second language* in the OC vs. *foreign language* in the EC (Kachru 1985). We hypothesized that there would be higher rates of countable usage in the former than the latter, given the assumed ‘post-learner’ status and relative norm-independence of many Englishes in the OC. As is widely supposed, ‘innovations’ (i.e. non-IC forms and usages) are more likely to occur and be tolerated in endonormative OC contexts than in the exonormative EC, where IC norms still dominate the TESOL agenda and mindsets of users. Second, we explored the role of cross-linguistic influence. Given the lack of plural morphemes in classifier languages, we hypothesized that the overall rate of pluralization for mass nouns would be higher in classifier language domains than in non-classifier language domains. Within the latter, we hypothesized that nouns translated by *pluralia tantum* in the substrate language would yield higher rates of pluralization than those with countable translations, and that these in turn would yield higher rates of pluralization than those with non-count translations.

Outer vs. expanding circles

The rate of countable usage of the mass nouns studied in the seven OC country domains was 3.43 per cent, compared with only 1.01 per cent for the EC. The numbers for the OC are perhaps exaggerated because of the behaviour of the words *homework* and *luggage*, especially in the Philippine data. If we replace the percentages for those items with the mean, the OC rate falls from 3.43 per cent to 2.48 per cent. But as the graph in Figure 3 shows, the tendency for greater countable usage in the OC is maintained, with *equipment* the one major exception, and *advice* and *violence* occurring slightly more in EC domains.

This tendency is not an artefact of one or two words or countries. If we concentrate on the 51 relatively high countability rates of above 5 per cent (representing over 14 per cent of all 350 rates calculated), we find that they are distributed across 18 of the 25 mass nouns and occur in all country domains but one (Iceland, in the EC). The OC has 34 individual country counts above 5 per cent, compared with only 19 in the EC. Of these ‘high’ rates, 12 in the OC are above 10 per cent, compared with only nine in the EC. Overall, the difference between OC and EC rates is extremely significant statistically ($t(24) = 5.204$, $p < 0.00001$).

Cross-linguistic influence

Of the fourteen country domains investigated, six were associated with dominant classifier languages as L1s, and eight with countable grammar similar to that employed by IC Englishes. There is some support in the data for the hypothesis that the rate of pluralization would be lower in classifier language domains than in non-classifier language domains:

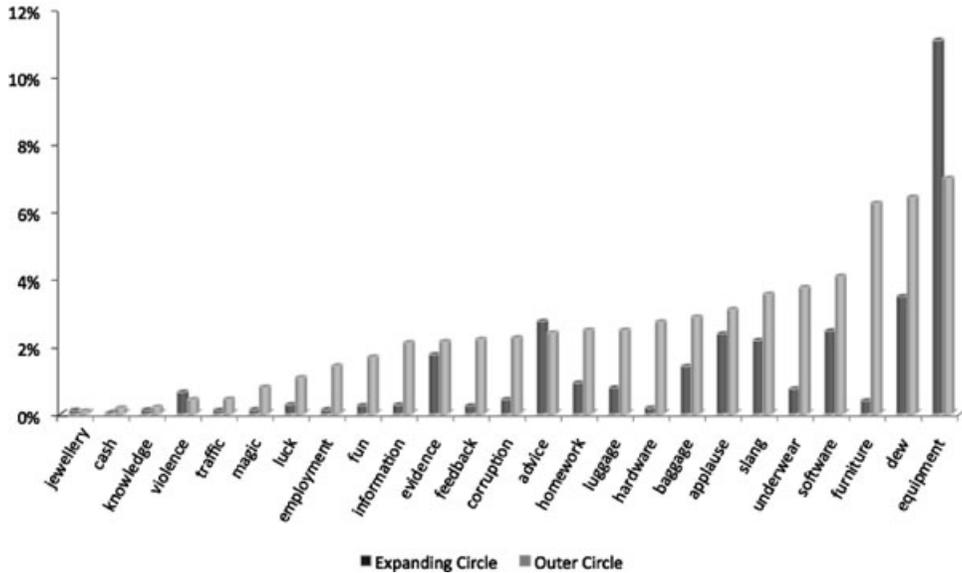


Figure 3. Countable usage by Kachruvian Circle, ordered by OC, lowest to highest.

between them, the classifier language domains (Korea, Thailand, Bangladesh, Hong Kong, the Philippines and China) exhibit only 1.83 per cent of plural usage, compared with 2.45 per cent over all 14 domains (when we exclude the Philippine outliers). Domains with dominant count grammar languages show an average of 2.3 per cent pluralized (therefore countable) usage. These differences are suggestive, but are not significant statistically and are probably too small to warrant claims for a generalized substrate role.

In order to detect the potential role of cross-linguistic influence from particular nouns in domains where dominant L1s use countable grammar, we obtained translations of the mass noun sample set from linguists who were native speakers of, respectively, Sinhala (the dominant language of Sri Lanka), Spanish (dominant in both Mexico and Belize), and Swahili (a Bantu language serving as a national language and lingua franca in Kenya). Informants were asked to provide the first form that occurred to them for each noun in the mass-oriented frame 'A lot of ____'. They were asked also to record any alternative form that occurred to them as a natural translation.

There is some evidence of a role for cross-linguistic influence, but it is not generalized and does not appear to be a powerful predictor of countability patterns. In Sinhala, 18 of the 25 nouns were expressed by our informant using *pluralia tantum*; all second preferences for translation equivalents, where given, were also plurally inflected nouns. Six nouns were expressed by singular nouns in both first and (where given) second options. One noun (*traffic*) was expressed as first preference by a plural noun phrase or optionally by a singular noun. The average rate of countable use for the nouns pluralized in Sinhala was 4.98 per cent (4.75% if we include *traffic*), whereas for the singular nouns it was half as much, at 2.48 per cent. But the difference is not statistically significant. For the primarily Spanish-speaking domains, Belize and Mexico, the pattern is repeated, although in attenuated form, as we had hypothesized, since unlike the case of *pluralia tantum* of

Sinhala, Spanish does have singular forms available for all the nouns given as translations. Excluding cases where there is no Spanish translation and the English form has been borrowed (*software* and *slang*), we find a mean of 2.22 per cent countable usage for the seven nouns translated as plural in Spanish, compared with 1.62 per cent for the 17 translated as singular. The difference is not statistically significant.

In the case of domains where Bantu languages are predominantly spoken, namely, Kenya and Angola, the pattern is murkier. There is little consistency across the figures for the two domains. In line with our observation regarding OC vs. EC trends, Kenya has over twice the rate of countable usage than Angola. But it is not consistently the same nouns which receive proportionately greater pluralization than IC norms. For example, *evidence* was translated by our Swahili-speaking informant as *ushahidi*, a noun marked with the singular prefix *u-* (prefix class 11), used more for non-count nouns (Contini-Morava 2000), and although only 0.77 per cent of tokens are pluralized in the Kenya domain, this figure rises to 8.92 per cent in the Angola domain. Conversely, the noun with the highest rate of pluralization for Kenya is *furniture*, at 16.96 per cent, compared with only 0.91 per cent for Angola. The Swahili translation provided by our informant, *samani*, belongs to the unprefix class 9/10, used mostly with countable nouns. However, a highly pluralized mass noun in the Kenya sample is *equipment*, at 11.29 per cent, expressed with the Swahili noun *vifaa*, exhibiting the plural prefix *vi-* (class 8). *Equipment* is the second highest pluralized form also for Angola, at 6.5 per cent. The evidence from Swahili noun classes is potentially unreliable, however, since the extent to which number-marking patterns are generalizable across Bantu languages is unclear (Contini-Morava, personal communication). So while cross-linguistic influence may have a potential role in Englishes with Bantu substrates, it is patently not a defining role.

On the whole, our data suggest that although cross-linguistic influence may explain part of the variability in usage (especially in the case of classifier languages and L1 *pluralia tantum*), the scepticism of Platt et al. (1984) regarding the importance of substrate influence is warranted.

Summary

In sum, although these data do not yield much support for the systematic nativization of mass nouns where the substrate might lead one to expect this, there is some compelling evidence for the greater independence of OC users from IC norms, with over twice the rate of plural use attested in ‘second language’ (OC) users than in ‘foreign language’ (EC) users. This difference suggests that it is exposure to, and attitudes to, different models of English that determine differences in the amount of countable usage of mass nouns, rather than the initial L1 state of users’ linguistic knowledge: EC learners tend to converge on IC norms more than OC learners do, independently of the way their first languages express the individuation of nominalizable concepts.

CONCLUSION

The picture that emerges from our data and analysis is one of notable and widespread use of mass nouns used countably, although the phenomenon is highly infrequent in comparison with usage that coincides with IC norms. There is little evidence from any domain that countable usage is emerging as a new norm across or within OC or EC varieties, or is the more common of the two grammatical options for these nouns in ELF interaction. In general, our data do not confirm the suggestion in much of the World Englishes literature

that countable usage of mass nouns is a frequently encountered feature of non-IC Englishes, at least inasmuch as these are represented in the speech of (mostly European) ELF and text on OC and EC websites. Of course, higher concentrations of countable usage may be revealed in some contexts through inspection of samples of localized, informal spoken interaction, especially in African and Asian OC and EC contexts, or in non-European regional ELF (e.g. Kirkpatrick 2010). And particular nouns certainly seem to be more prone than others to appear countably.

Although our analysis of potential causes yields evidence for a heightened degree of innovation in OC over EC Englishes, it does not point to a major role for substrate influence on the basis of lexico-grammatical differences in speakers' first languages. Claims for 'nativization' through cross-linguistic influence from L1s are therefore not substantiated by our data, although there are signs of the greater 'freedom' that OC users inherit to construct their own mental grammars on the basis of the local and distal models available. What is most surprising about our data is that mass nouns used countably were so hard to find, given their prominence in scholarly and professional work in world Englishes, ELF, and TESOL. It seems that the disproportionate invocation of the count/non-count distinction in prescriptive works for learners is not because of any major significance it has within the overall scheme of English forms and their functions, but rather because of its status as a conspicuous social marker of the border between IC and non-IC Englishes: of 'nativeness'. That is, it is being viewed as part of what Widdowson (1994: 381) identified as 'the symbolic possession of a particular community, expressive of its identity, its conventions, and values'. This is recognized in some pedagogically-oriented discussions of the phenomenon. In his grammar for English language professionals, for example, Jacobs (1995) observes that 'errors' such as *many waters*

do not hinder communication and are not associated with problems in learners' comprehension. Nevertheless, frequent determiner errors do pose questions of English language mastery that may disturb employers, admissions personnel, and teachers of other courses. (Jacobs 1995: 107)

Similarly, McKay (2002) acknowledges that

the pluralization of nouns like *equipments* and *evidences* will not cause problems of intelligibility; however, some may contend that such differences reflect a lack of competence on the part of the speaker and be an indication of the deterioration of the language. (McKay 2002: 127)

Clearly, then, monolithic conceptions of English underlie the IC gate-keeping functions of much TESOL practice, leading to the unquestioned belief that there is a fixed class of uncountable mass nouns that must be mastered in order to 'join the club'. Less obviously, however, the highlighting of the phenomenon in descriptions of OC and EC Englishes, often apparently on the basis of casual observation, also invites (neo)-monolithic thinking about English (cf. Pennycook 2009: 200). Our low counts in VOICE and the web lead to the conclusion that it has been given prominence because it is typically absent in IC Englishes rather than because it is present to any great extent outside the IC. Overall, we do not detect rates of usage that would justify the designation of countable mass nouns as a 'norm' for any non-IC national, international or supranational variety.

We suggest that linguistic descriptions which highlight such peripheral and communicatively inconsequential formal elements on the basis of their contrastiveness with IC English rather than quantitative data about actual usage, obscure the plurilithic reality of English

grammars and exaggerate unhelpful monolithic notions of nativeness and non-nativeness. Furthermore, we argue that pedagogical models which forefront such elements in teaching and testing may be doing a disservice to many learners: not because the countability of mass nouns may be deemed 'legitimate' in English models beyond the IC, but because: (a) there is no evidence that it affects communication, especially in ELF contexts; and (b) it perpetuates deficit models of learning by setting up monolithic, purely formal, indicators of accomplishment (Hall, forthcoming).

We believe that corpus studies which mine authentic lingua franca usage across a range of genres can provide an informative empirical resource with which to assess the ways in which individual mental grammars can coalesce or diverge, and therefore the extent to which patterns of usage can be alleged to characterize the linguistic resources of some groups and not others, and so, in turn, be identified as localized or globalized norms. The case discussed here, a lexico-grammatical feature homogenized as a norm for IC Englishes in the relatively recent past (Crystal 2008; Toyota 2009), has been revealed as a marginal phenomenon in OC and EC Englishes: a shibboleth of the native/non-native dichotomy rather than a significant feature of specific groups of speakers or situations of use. Our analysis demonstrates an approach to the lexico-grammatical forms of English which stresses their variable and contingent nature, as part of plurilithic systems which are fashioned and refashioned through the communicative practices of individuals and groups (Makoni and Pennycook 2008) and in the minds of millions of individual learners and users (Hall, forthcoming). Such an approach is consistent with a descriptive linguistics which is blind to 'nativeness' and an applied linguistics which recognizes the subservience of form to function.

NOTES

1. The research for this paper was supported by grants to the first author by York St John University Research Office and HEFCE (TESS funding). We are grateful to Rachel Wicaksono, Nathan Page and anonymous reviewers for their very helpful comments on earlier drafts.
2. BNC cannot *a priori* identify and exclude non-IC Englishes used in British samples, so not all usage in BNC can be assumed to be produced by IC users.
3. Due to the high number of tokens involved, we were unable to inspect surrounding context. A consequence of this was that searches for a prior indefinite article were not conducted, as they were with VOICE, because an article + noun sequence can always be the beginning of a noun phrase headed by a compound noun, e.g. *a corruption trial* or *a luggage tag*.
4. This filtering process still resulted in data for all 25 mass nouns in all domains, with only two exceptions: Yemen, with 10 nouns eliminated, and Angola, with two.

REFERENCES

- Ahulu, Samuel (1998) Grammatical variation in international English. *English Today* 14, 19–25.
- Aikhenvald, Alexandra Y. (2003) *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Allan, Keith (1980) Nouns and countability. *Language* 56, 541–67.
- Bamgbose, Ayo (1998) Torn between the norms: Innovations in world Englishes. *World Englishes* 17, 1–14.
- Barner, David, Inagaki, Shunji, and Li, Peggy (2009) Language, thought, and real nouns. *Cognition* 111, 329–44.
- BBC World Service (2004) Learning English. Virtual English masterclass. Retrieved 6 June 2011 from http://www.bbc.co.uk/worldservice/learningenglish/events/grammar_transcript.shtml.
- Bergh, Gunnar (2005) Min(d)ing English language data on the web: What can Google tell us? *ICAME Journal* 29, 25–46.
- Björkman, Beyza (2008) English as the lingua franca of engineering: The morphosyntax of academic speech events. *Nordic Journal of English Studies* 7, 103–22.
- Bokamba, Eyamba G. (1992) The Africanization of English. In Braj B. Kachru (ed.), *The Other Tongue: English across Cultures* (2nd. edn., pp. 125–47). Urbana, IL: University of Illinois Press.

- Brook-Hart, Guy (2009) *Learning from Common Mistakes. Intermediate*. Cambridge: Cambridge University Press.
- Carter, Ronald, and McCarthy, Michael (2006) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Chierchia, Gennaro (1998) Plurality of mass nouns and the notion of semantic parameter. In Susan Rothstein (ed.), *Events and Grammar*. Dordrecht: Kluwer.
- Contini-Morava, Ellen (2000) Noun class as number in Swahili. In Ellen Contini-Morava and Yishai Tobin (eds.), *Between Grammar and Lexicon* (pp. 3–30). Amsterdam: John Benjamins.
- Crystal, David (2008) Two thousand million? Updates on the statistics of English. *English Today* 93, 3–6.
- DeCapua, Andrea (2008) *Grammar for Teachers: A Guide to American English for Native and Non-native Speakers*. New York: Springer.
- Fletcher, William H. (2007) Concordancing the web: Promise and problems, tools and techniques. In Marianne Hundt, Nadja Nesselhauf and Carolin Biewer (eds), *Corpus Linguistics and the Web* (pp. 25–46). Amsterdam: Rodopi.
- Government of Hong Kong SAR (2007) Table 140: Population aged 5 and over by usual language, 1996, 2001 and 2006. Retrieved 19 August 2010 from http://www.censtatd.gov.hk/hong_kong_statistics/statistical_tables/index.jsp?tableID=140&ID=&subjectID=1.
- Gramley, Stephan (2001) *The Vocabulary of World English*. London: Arnold.
- Gupta, Anthea Fraser (2006) Standard English in the world. In Rani Rubdy and Mario Saraceni (eds), *English in the World: Global Rules, Global Roles* (pp. 95–109). London: Continuum.
- Hall, Christopher J. (forthcoming) Cognitive contributions to plurilithic views of English and other languages. *Applied Linguistics*. doi: 10.1093/applin/ams042.
- Higgins, Christina (2003) “Ownership” of English in the Outer Circle: An alternative to the NS-NNS dichotomy. *TESOL Quarterly* 37, 615–44.
- ICE (2010) International Corpus of English. Retrieved 10 June 2011 from <http://ice-corpora.net/ice/index.htm>
- Internet World Stats (2010) Internet world users by language. Retrieved 18 May 2010 from <http://www.internetworldstats.com/stats7.htm>.
- Jacobs, Roderick A. (1995) *English Syntax. A Grammar for English Language Professionals*. Oxford: Oxford University Press.
- Jacquemet, Marco (2005) Transidiomatic practices, language and power in the age of globalization. *Language and Communication* 25, 257–77.
- Jenkins, Jennifer (2006) The spread of EIL: A testing time for testers. *ELT Journal* 60, 42–50.
- Jenkins, Jennifer (2009) English as a lingua franca: Interpretations and attitudes. *World Englishes* 28, 200–7.
- Kachru, Braj B. (1985) Standards, codification and sociolinguistic realism: The English language in the outer circle. In Randolph Quirk and Henry G. Widdowson (eds), *English in the World: Teaching and Learning the Language and Literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kachru, Braj B. (1992) Models for non-native Englishes. In Braj B. Kachru (ed.), *The Other Tongue: English across Cultures* (2nd edn., pp. 48–74). Urbana, IL: University of Illinois Press.
- Kachru, Yamuna, and Smith, Larry (2008) *Cultures, Contexts, and World Englishes*. New York: Routledge.
- Keller, Frank, and Lapata, Mirella (2003) Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics* 29, 459–84.
- Kirkpatrick, Andy (2010) Researching English as a lingua franca in Asia: The Asian Corpus of English (ACE) project. *Asian Englishes* 13, 4–18.
- Kirkpatrick, Andy, and Deterding, David (2011) World Englishes. In James Simpson (ed.), *The Routledge Handbook of Applied Linguistics* (pp. 373–87). London: Routledge.
- Lapata, Mirella, and Keller, Frank (2005) Web-based models for natural language processing. *ACM Transactions on Speech and Language Processing* 2, 1–30.
- Lewis, M. Paul (ed.) (2009) *Ethnologue: Languages of the World* (16th edn.). Dallas, TX: SIL International.
- Lowenberg, Peter H. (1992) Testing English as a world language: Issues in assessing non-native proficiency. In Braj B. Kachru (ed.), *The Other Tongue. English across Cultures* (2nd edn., pp. 108–21). Urbana, IL: University of Illinois Press.
- Makoni, Sinfree, and Pennycook, Alistair (2007) Disinventing and reconstituting languages. In Sinfree Makoni and Alistair Pennycook (eds.), *Disinventing and Reconstituting Languages* (pp. 1–41). Clevedon: Multilingual Matters.
- McArthur, Tom (2002) *The Oxford Guide to World English*. Oxford: Oxford University Press.
- McKay, Sandra L. (2002) *Teaching English as an International Language: Rethinking Goals and Approaches*. Oxford: Oxford University Press.
- Mesthrie, Rajend, and Bhatt, Rakesh M. (2008) *World Englishes. The Study of New Linguistic Varieties*. Cambridge: Cambridge University Press.
- Mesthrie, Rajend, and West, Paula (1995) Towards a grammar of proto South African English. *English World-Wide* 16, 105–34.
- Mollin, Sandra (2007) New variety or learner English? Criteria for variety status and the case of Euro-English. *English World-Wide* 28, 167–85.
- Norvig, Peter (2007) Theorizing from data: Avoiding the capital mistake. Video lecture. Retrieved 18 May 2010 from <http://www.youtube.com/watch?v=nU8DcBF-qq4>.

- Pennycook, Alistair (2009) Plurilithic Englishes: Towards a 3D model. In Kumiko Murata and Jennifer Jenkins (eds), *Global Englishes in Asian Contexts. Current and Future Debates* (pp. 194–207). Basingstoke: Palgrave Macmillan.
- Platt, John T., Weber, Heidi and Ho, Mian Lian (1984) *The New Englishes*. London: Routledge and Kegan Paul.
- Quirk, Randolph (1990) Language varieties and standard language. *English Today* 6, 3–10.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Rajagopalan, Kanavillil (1997) Linguistics and the myth of nativity: Comments on the controversy over ‘new/non-native’ Englishes, *Journal of Pragmatics* 27, 225–31.
- Renouf, Antoinette (2003) WebCorp: Providing a renewable data source for corpus linguists. In Sylviane Granger and Stephanie Petch-Tyson (eds.), *Extending the Scope of Corpus-based Research: New Applications, New Challenges*. Amsterdam: Rodopi.
- Schmied, Josef (2006) East African Englishes. In Braj B. Kachru, Yamuna Kachru and Cecil L. Nelson (eds), *The Handbook of World Englishes* (pp. 188–202). Oxford: Blackwell.
- Schneider, Edgar W. (2011) *English Around the World*. Cambridge: Cambridge University Press.
- Schoenberg, Irene E. (2005) *Focus on Grammar: A Basic Course for Reference and Practice*. London: Longman.
- Seidlhofer, Barbara (2005) English as a lingua franca. In Albert S. Hornby (ed.), *Oxford Advanced Learner's Dictionary of Current English* (7th edn.), Oxford: Oxford University Press.
- Seidlhofer, Barbara (2009) Orientations in ELF research: Form and function. In Anna Mauranen and Elina Ranta (eds), *English as a Lingua Franca: Studies and Findings* (pp. 37–59). Newcastle upon Tyne: Cambridge Scholars.
- Sey, Kofi A. (1973) *Ghanaian English: An Exploratory Survey*. London: Macmillan.
- Shim, Rosa J. (1999) Codified Korean English: Process, characteristics and consequence. *World Englishes* 18, 247–58.
- Slobin, Dan (1973) Cognitive prerequisites for the development of grammar. In Charles Ferguson and Dan Slobin (eds), *Studies in Child Language Development* (pp. 175–208). New York: Holt, Rinehart and Winston.
- Swan, Michael (1995) *Practical English Usage*. Oxford: Oxford University Press.
- Toyota, Junichi (2009) When the mass was counted: English as classifier and non-classifier language. *SKASE Journal of Theoretical Linguistics* 6, 118–30.
- Villavicencio, Aline (2003) Verb-particle constructions in the world wide web. Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications. Toulouse, France. Retrieved 18 May 2010 from <http://www.cl.cam.ac.uk/~av208/vpc-www.final.ps.gz>.
- VOICE (2009) The Vienna-Oxford International Corpus of English. Retrieved 11 October 2009 from <http://voice.univie.ac.at>.
- Widdowson, Henry G. (1994) The ownership of English. *TESOL Quarterly* 28, 377–89.
- Williams, J. (1987) Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8, 161–99.
- Xu, Fei (2007) Concept formation and language development: Count nouns and object kinds. In Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics* (pp. 627–34). Oxford: Oxford University Press.
- Yates, Jean (2006) *English Vocabulary for Beginning ESL Learners*. New York: McGraw-Hill.

(Received 13 June 2012)