

Chapter 3

Strongly non-countable nouns: Strategies against individuality

Scott Grimm, Ellise Moon & Adam Richman

University of Rochester

Studies in countability have uncovered a range of ontological entities which permit counting, including natural concrete individuals, discrete events, and taxonomic subkinds. Identifying the reasons why nominal referents may *not* be counted has been less successful, however, and remains controversial. This paper examines nouns that are “strongly non-countable”, those nouns for which combination with the plural marker, quantifiers, and nearly all other forms of determination is a vanishingly rare event. This paper develops a data set of nearly 500 such nouns, adducing their strongly non-countable status from usage over a 350 million word corpus (Davies 2009). Through further internet searches, we attest rare, but possible, patterns of coercion available to these nouns. We then develop a classification of the different notional categories that these nouns belong to. Finally, we examine broad distributional patterns and argue that these strongly non-countable nouns contrast with countable nouns as to their patterns of usage, in particular, being less discourse-salient and less referential than their count noun counterparts.

Keywords: countability, non-countable nouns, coercion, abstract nouns

1 Introduction: Assessing the varieties of non-countable nouns

When a noun has a countable interpretation, it is often intuitively clear why the countable interpretation comes about: The noun references some sort of unit which permits counting. The nature of this unit may be different depending on



the noun, whole objects of the natural kind sort (*dogs*) or measurement units (*kilos*) to give just two examples among many, but it appears reasonably straightforward to identify that there is a unit and that that is what is being counted.

When a noun fails to permit a countable interpretation, the situation is usually far less clear. Much research over the last two decades has gone into distinguishing two types of non-countable nouns: SUBSTANCES, those nouns traditionally considered to be “mass” such as *water* or *clay*, and AGGREGATES, including *furniture*, the most famous example, along with other nouns such as *jewelry* or *mail*. The non-countability status of substances has traditionally been supported by the strong intuition that neither *water* nor *clay* in their primary uses make reference to individual units, more technically speaking ‘atoms’, which would serve as a basis for quantification. In contrast, *furniture* and other nouns of the aggregate type *do* refer to individuals, despite their grammatical non-countability status. Theoretical models of countability have mostly been content to account for these three types of nouns: individuals, substances and aggregates (see, for instance, Bale & Barner 2009, Chierchia 2010, or Deal 2017). Most agree that the grammatical contrasts among these noun types reflect an ontological contrast although it is a matter of controversy as to how tight the relation is.

This paper contends that the challenge of accounting for non-countable nouns is far greater than typically assumed in the literature and establishes some basic results on the diversity of non-countable nouns in English.¹ We will have little to say about the different virtues or short-comings of any particular theoretical account of non-countable nouns in this paper, instead we limit ourselves to establishing empirical baselines as to what types of non-countable nouns there are and how they behave contextually and grammatically.

The structure of the paper is as follows. In §2, we establish our methodology and, through corpus work, isolate close to 500 nouns that are rigidly non-countable or nearly always so. §3 asks if these nouns *ever* are counted and examines the different patterns of coercion observed through further internet-based searches and categorizes them. In §4, we elaborate a classification of the different notional categories that these nouns belong to, which themselves fall into four super-categories: Entities, Eventualities, Phenomena, and Abstract. We then examine the correlation between the different notional categories and the different types of coercion observed in §3. §5 examines broad distributional patterns of these nouns at the level of clauses and nominal phrases, demonstrating that, on

¹See also Allan (1980) and Kiss et al. (2016) for other larger-scale studies which help establish the diversity of countable and non-countable nouns, as well as Sutton & Filip (2019) and Sutton & Filip (2020), which provide recent empirical work on certain domains of abstract nouns.

average, the non-countable nouns of our data set show behavior consistent with less discourse-salient and less referential uses. We conclude in §6.

Our aim is that this study will facilitate the investigation of non-countability in several directions. First, establishing, initially for English, what the lexical variation is among non-countable nouns, viz. what types of nouns have non-countable readings? Answering this question with a systematic approach will hopefully open up avenues for cross-linguistic comparison: Do the countability statuses of different notional categories co-vary across languages? Clearly, answers to these questions will help test the predictive power of current theories: What would a theory look like that not only explains the non-countability of *water* and *furniture* but also of *coriander*, *parenthood*, *fun*, or *sportsmanship*? Ultimately, this effort contributes to understanding the causal foundations of non-countability.

2 Methodology: Discovering strongly non-countable nouns

To assess the spectrum of non-countable nouns, we extracted a large set of nouns which, based on several measures, showed the lowest degree of countability. We chose those with the lowest degree of countability in part to exclude polysemous nouns, also known as “dual-life” or “flexible” nouns, and to minimize interference from nouns lending themselves to secondary interpretations through coercion. In all, we assess nearly 500 nouns, a sufficient quantity to deliver insight into potential classes of non-countable nouns while remaining of a manageable size.

The non-countable nouns were selected from the database described in Grimm & Wahlang (2021), derived from a 350 million word portion of the Corpus of Contemporary English (Davies 2009). This was subsequently processed via a natural language processing (NLP) pipeline, parsing and annotating each occurrence of each noun with all relevant dependencies in which the noun stood (using Universal Dependencies from De Marneffe et al. 2014). This process captured a vast amount of distributional information about each noun, permitting further analytical investigation of nouns’ behaviors. (See Grimm & Wahlang 2021 for further details on the corpus processing and database development.)

We filtered this database to extract strongly non-countable nouns. Occurrence in bare plural was found in Grimm & Wahlang (2021) to be the strongest predictor of countable nouns in the database, so we filtered the data most tightly on this feature, requiring a noun’s percentage of occurrences in the bare plural to be lower than 2% of all occurrences and, additionally, occurrence with numeric modifiers to be lower than 20%. We allowed for some amount of flexibility in these

constraints to account for possible noise in the corpus data, due to parsing or other processing errors, as well as to not exclude rare coerced occurrences of the noun. We allowed for more flexibility in the occurrence with numeric modifiers since, for our purposes, there is a larger amount of noise due to how De Marneffe et al. (2014) treat numeric modifiers, since they include under numeric modifiers not only cardinal numbers and the like but also measure terms such as *2 kilos* (which do not discriminate between countable and non-countable nouns).²

To select the best candidates, the data was sorted first by the lowest bare plural noun percentage (giving preference to nouns with the least noise in that category), then by lowest proper noun percentage (that is, those nouns which were very rarely, if at all, tagged as proper nouns, thereby excluding proper nouns, like *William* or *Cincinnati*, which would have almost no occurrences in the plural), and then highest value of overall occurrences (to preference nouns that we had the most data for). The resulting data was further filtered to only include nouns coded as uncountable in the CELEX database (Baayen et al. 1996) as another measure to narrow our scope. Finally, we selected only nouns for which there were greater than 200 example sentences in our data, giving us sufficient data from which to generalize. From this sorted list, we selected the top 550 nouns as the starting point for our research, assuming that around 50 of these would ultimately be excluded due to noisiness in the data or ambiguity between multiple senses.

This list of 550 nouns was then further pared down by hand during the process of analyzing nouns for rare and contextual count examples using Google searches (see §3). A number of these searches returned established countable uses of the noun (e.g. *prospects*, *writings*) which led us to remove that noun from our list. In total, 26 nouns had enough count examples to be excluded from the data and 42 nouns had multiple distinct senses (some of which were highly countable), gerund uses, or appeared almost exclusively in fixed phrases (*in spite of*) and so were also excluded. With the final list of 482 nouns, we built a dataset containing distributional information with the data from Grimm & Wahlang (2021) for each noun, as well as additional data compiled from COCA example sentences. This provided us with not only summary statistics for the behavior of each noun (e.g., the percentage of occurrences with the definite article or as the subject of the verb phrase) but also lists of the unique modifiers (e.g., adjectives, case modifiers, possessive constructions) compiled from every example in our data pulled from

²The settings for these filters are not the only ones possible, and are proposed based on our (subjective) experiments with different percentages for both of the filters and examining the resulting sets of nouns. These settings were felt to be optimal for permitting some level of noise or ambiguity while also narrowing down the set to truly non-countable nouns.

COCA. For the comparison between the grammatical distribution of strongly non-countable nouns and countable nouns in §5, this data set was also extended with a set of core countable nouns, grouped separately, from data in Grimm & Wahlang (2021) (clusters 7 and 8, containing 799 count nouns).

In addition to this distributional information, our final dataset also contains data on each noun’s countability as well as derivational morphology from CELEX (Baayen et al. 1996), hand-annotation of each noun’s notional category (see §4), and the possible count coercion contexts that noun was found to appear in (as discussed in §3). A separate file contains examples demonstrating each type of coercion found for each noun. We have made the final dataset and accompanying files publicly available at <https://quantitativesemanticslab.github.io/>.

3 The contexts of coercion

While our data set contains a large number of occurrences for each noun considered (at least 200), this is not sufficient to determine if a noun which is normally non-countable *ever* gets counted, and if so, upon which basis that counting is carried out. To examine valid, albeit rare, countable examples of these nouns, we performed a battery of Google searches for each of the 482 nouns. For each noun, we searched for occurrences with the definite article *the*, plural demonstratives *these* and *those*, numerals *two* and *three*, as well as quantifiers *some*, *many*, and *multiple*. We limited ourselves to inspecting the first five pages of results per search (~50 results per search), which in practice was sufficient to turn up any countable uses.³

We collected a number of example sentences demonstrating each type of countability coercion observed with a given noun. Table 1 lists the different countable uses, which we will refer to as COERCION TYPES, observed of the 482 nouns in the data set and provides the number of nouns observed for each coercion type. While no countable examples were found for 262 of the nouns, the remaining nouns had examples that could be attributed to one or more coercion types.

The coercion types were determined by the authors and a research assistant who separately annotated the collected examples.⁴ They discussed the annotations and agreed upon a final set of labels on a small training portion of the

³We ignored a range of occurrences with plural forms that arose in uses with proper nouns, in typos, translations, non-native uses, or misuses.

⁴This was carried out on a portion of the data for Jargon and Archaic had already been excluded.

Table 1: Types of countable uses of non-countable nouns. *Note:* As some nouns were found to have more than one type from multiple example sentences, the “Number of nouns observed” column does not sum to 482.

Coercion type	Number of nouns observed
Entity Type	96
Event	67
Possessor	35
Relational	29
Event Type	24
Packaging	8
Value	6
Modificational	4
Countable only in specific contexts ^a	126
Archaic ^b	59
No countable uses observed	137

^aI.e. jargon.

^bAll count uses predate 1880.

data (150 example sentences). Then two of this group served as annotators independently annotated the remaining 377 example sentences and compared their annotations. Inter-annotator agreement was ‘moderate’ (Cohen’s $\kappa = 0.56$) when calculated on the entire test dataset (527 sentences). Agreement was even higher on two subsets of the data. One subset excluded even more archaic or jargon uses and the inter-annotator agreement was ‘substantial’ (Cohen’s $\kappa = 0.65$), and similarly for a different subset which excluded a specific error pattern from one of the annotators who over-labeled with the Packaging coercion type (Cohen’s $\kappa = 0.66$).⁵

Since the theoretical understanding of different types of coercions possible – beyond the familiar contexts discussed in the literature under “packaging” and “grinder” – is still limited, despite a growing literature which describes some of the lesser-studied countability shifts (Payne & Huddleston 2002, Grimm 2014, Husić 2020, Zamparelli 2020), we now detail with examples the different coercion types we observed for these nouns.

⁵See full data set at <https://quantitativesemanticslab.github.io/>.

3 Strongly non-countable nouns: Strategies against individuality

Entity Type: These count uses refer to multiple classifications, compositions, severities, etc. of the entity designated by the lexicon. As a heuristic diagnostic, *type of* or a similar phrase may be felicitously added to clarify the contrast evoked.

- (1) **Twines** can differ by their material and strength, which changes how they should be used. Some **twines** are ideal for cooking since they can withstand heat and don't impart flavor onto your food, while other **twines** are perfect for decoration or more heavy-duty use.⁶
- (2) We now face two **agricultures**. The long-term model is exploitive and degenerative, while the new model is regenerative and more profitable.⁷

Event: These count uses refer to multiple occurrences or iterations of the event designated by the noun. If the events *are not* simultaneous, ordinal numerals or lexical items denoting temporal location may stand in to distinguish the events, as in (3). If the events *are* simultaneous, other modifiers such as locations may be used to distinguish the events, as in (4).

- (3) The **automations** are not necessarily run at the top of the hour, and it may not be exactly one hour between executions of an automation.
(Google Books)
- (4) Most important of the **minings** were those of the Gotthard and Sim-plon tunnels.
(Google Books)

Possessor: These count uses make reference to distinct agents displaying the property, often implicitly.

- (5) The management team understands how individualized the recovery process is and that no two **sobrieties** look the same.⁸

Relational: These count uses arise from distinguishing multiple types in terms of their relation to, e.g., other event participants. In example (6), different types of *contentment* are established with respect to the different things with which one may be content, i.e., different stimuli.

⁶<https://www.webrestaurantstore.com/guide/880/types-of-twine.html>

⁷<https://www.farmprogress.com/management/we-now-face-two-agricultures>

⁸<https://m.yelp.ca/biz/the-district-recovery-community-huntington-beach>

- (6) Those **contentments** have come to include housing, healthcare, schooling and employment as well as freedom from intimidation.⁹

Event Type: While countable Event uses refer to multiple, specific occurrences of the event designated by the noun, Event Type uses do not refer to specific events, but more abstractly, contrasting different types of the event in question.

- (7) Again, this is not to hold equivalence between either the types of violence or particular **violences** in each category. (Google Books)

Packaging: These count uses evoke a bundling or containment of the noun's referent as a single unit, often assuming a standard measure or container.

- (8) Six quarts of milk, two **buttermilks**, two chocolates, and three pints of cream. (Google Books)

Value: These count uses refer to varying levels or numerical values of a scale associated, perhaps implicitly, with the noun. This use differs from Entity Type coercions as this relies on a value or degree. Explicit values may be added to distinguish between the singular units.

- (9) Low **latitudes** are those locations found between the Equator (0 degrees N/S) and 30 degrees N/S. The middle **latitudes** are found between 30 degrees N/S and 60 degrees N/S. And the high **latitudes** are found between 60 degrees N/S and the poles (90 degrees N/S).¹⁰
- (10) Barley was germinated in soils of two **moistures** (40 and 50 per cent). (Google Books)

Modificational: These count uses are of (typically) adjectives, where the head noun is absent and the modifier or distinguishing property is actually what bears the plural morphology.

- (11) If there really were 6 vanilla and 6 peanut butter candies in the box, what is the probability that you would have picked three **vanillas** in a row?¹¹

⁹<https://reader.exacteditions.com/issues/59737/page/10>

¹⁰https://www.shsu.edu/~dl_www/bkonline/131online/f02latitude/02index.htm

¹¹<https://www.slader.com/discussion/question/someone-hands-you-a-box-of-a-dozen-chocolate-covered-candies-telling-you-that-half-are-vanilla-cre-2/>

3 Strongly non-countable nouns: Strategies against individuality

Jargon: These are count uses that occur only within specific contexts, primarily technical jargon. The example in (12) is a commonly found example of jargon occurring in chemistry and physics contexts that describe atoms and molecules.

- (12) However, at the oxygens bridging two **aluminums**, oxygens were swapped only about once every 13 hours.¹²

Archaic: These count uses occur only in poetic uses or examples predating 1880, and current countable uses are not found outside of these contexts.

- (13) The capytyle doth shew of the **fortitudes** of the planetes.
(Google Books)

No countable uses observed: These nouns had no occurrences of count uses.

In summary, this data set leads us to observe a wide range of possible shifts from non-countable to countable interpretations, many of which have been little explored at this point. For Type coercions, while there is some discussion and even controversy about (the lack of) subtype coercions (see Grimm & Levin 2017 and Sutton & Filip 2016 and references therein), it has primarily revolved around nouns describing liquids or substances (*wines*) and artifactual aggregates lacking subtype readings (*furniture*), yet there are many other domains to check to see how type coercion is effected, as exemplified in (2). The interpretational shifts we list under Event and Possessor have to date only received brief treatments (Grimm 2014, Zamparelli 2020, Husić 2020) and similarly for Relational (Grimm 2014) (although a more sophisticated treatment has begun to be developed for informational nouns in the line of work of Sutton & Filip 2019 and Sutton & Filip 2020), while the observation of Value-based and Modificational count shifts is novel to the best of our knowledge. Again, it is possible that this classification stands in need of revision and, for instance, Relation or Value could be grouped under Type if understood more broadly, but we have erred on the side of being more explicit to bring out some of more unusual cases of coercion observed. A related issue is if all of the examples examined are truly cases of coercion as opposed to polysemy – again we have erred on the side of inclusion as coercion when a plausible case can be made.

¹²<https://www.ucdavis.edu/news/oxygen-swapping-offers-clues-toxics-management/>

4 The notional varieties of strongly non-countable nouns

A major theme of countability research is the relation between contrastive grammatical countability classes and corresponding contrasts, or lack thereof, of notional, or ontological, types of the corresponding referents of the grammatical classes. As mentioned, several authors propose that the referential types of individuals, aggregates, and substances are those that are responsible for countability contrasts (Bale & Barner 2009, Chierchia 2010, Deal 2017). It is therefore critical to examine the relation between the strongly non-countable nouns and their corresponding notional types. The different notional types brought forth by the strongly non-countable data demonstrate that those referential types may be necessary to account for the grammatical behavior related to countability, but those three types are far from sufficient. Instead, we observed rich variation in the notional types that correspond to strongly non-countable nouns, transcending the contrasts typically posited to explain grammatical countability patterns, as in those between, e.g., substances vs. individuals vs. aggregates or events vs. states.

4.1 Notional categories of strongly non-countable nouns

This section puts forth a classification of the 482 nouns into 27 separate “notional” categories, such as LIQUIDS or DISEASE. While the categorization presented here no doubt reflects some core aspects of the nouns’ meaning, we hasten to emphasize that this classification is preliminary – nearly all of these nouns have never been systematically analyzed and we do not pretend to have been able to fully analyze them here. That said, even this initial categorization establishes that the range of notional noun types which show strongly non-countable behavior is far greater than one would suppose from the discussions in the literature.

Table 2 (page 68) displays the categorization. The 27 categories are broadly grouped into four super-categories: Entities, Eventualities, Phenomena, and Abstract. These are organized in terms of the apparent ontological commitments of the nominal descriptions falling under each category: Entities includes nouns describing entities rooted in physical existence (“concrete entities”); Eventualities includes those entities rooted in a temporal dimension, here using the term “eventualities” in the sense of Bach (1986) for both events and states; Phenomena – such as diseases or natural forces – while having a connection to the physical world are more abstract than the concrete objects found in Entities; and Abstract contains nouns that are, at least on their primary reading, detached from the physical world, comprised of nouns describing, e.g., atemporal, non-physical qualities (*cleanliness*) or domains of knowledge (*geology*). In the following, we

3 Strongly non-countable nouns: Strategies against individuality

discuss the different categories and their nouns and note their particularities against the background of the expectations from the countability literature.

These proposed categories can also often be distinguished via contrasting grammatical properties, often those related to the argument structure considerations. For instance, unlike pure substances, the BY-PRODUCTS category contains nouns that allow a *from* argument which specifies from where the substance originated (*refuse from the facility*). Similarly, MENTAL STATES differ from GENERAL STATES in that the former, such as *awe* require a participant who is mentally engaged in the event. While these grammatical contrasts have informed our categorization, we only discuss them in passing as they do not directly map onto countability contrasts.

4.1.1 Entities

The Entities super-category includes some representatives of “classic” non-countable noun types, such as SUBSTANCES (*dirt*), MATERIALS (*asphalt, hemp, latex*), GRAINS AND FLOURS (*bran, cornstarch, flax*), and LIQUIDS (*booze, kerosene, oil*). Although these notional categories are the most typical ones used to exemplify non-countable nouns (e.g. *water*, a liquid) in our data, these categories are somewhat sparsely populated compared to the number of other categories (e.g., mental state nouns). No doubt this results from the high number of nouns in these categories which are “dual-life” nouns, that is, nouns which also manifest a countable use and thus were excluded from our set of strongly non-countable nouns. At the same time, other instances of liquids and substances do arise, namely those that have been processed or manufactured, falling under the categories of CHEMICALS & ELEMENTS and DRUGS.

Better represented are AGGREGATE nouns, for which nearly all the examples from the literature are found in our data set (*footwear, furniture, luggage, silverware*) along with nouns which have some claim to “aggregate” status, even if most likely possessing some different characteristics than *furniture*, such as *bedding, homework, merchandise, paperwork, parking, traffic, weaponry, and wildlife*. Thus, our methodology is able to replicate the observation made at several points in the literature that aggregate nouns like *furniture* are less flexible and therefore more strongly non-countable than typical substance or liquid nouns.

The category BY-PRODUCTS collects nouns that either designate materials which result from some prior activity (*rubble, sawdust, sewage, smoke, soot*) or designate collections of entities or materials deemed worthless (*garbage, refuse, trash, filth*). While the cause for the first group’s non-countability status may

Table 2: Notional classes of non-countable nouns

Category	Examples
Entities (108)	
aggregates (27)	footwear, furniture, glitter, traffic
by-products (10)	garbage, rubble, sawdust, soot
chemicals & elements (20)	ammonia, glucose titanium, uranium
drugs (7)	cocaine, morphine, nicotine
meat (3)	pork, poultry, venison
grains/flours (4)	bran, flax, oatmeal
herbs and spices (11)	cumin, nutmeg, paprika, parsley
materials (11)	carpeting, denim, plywood
liquids (11)	bile, buttermilk, oil, rainwater
natural substances (4)	dirt, driftwood, flesh, quartz
Eventualities (109)	
events (8)	atonement, bribery, legalization
multi-participant events (6)	acclaim, applause, bloodshed, gunfire
coming-into-/going-out-of-exist. (13)	abolition, emergence, eradication
mental states (28)	awe, bewilderment, remorse, unease
general states (17)	illiteracy, prosperity, puberty
activities (25)	banking, espionage, gardening
gradual/repeated processes (12)	conservation, enforcement
Phenomena (21)	
diseases (6)	arthritis, flu, hepatitis, herpes
disorders (7)	alcoholism, amnesia, anorexia
natural force (8)	antimatter, electricity, momentum
Abstract (212)	
domains (16)	agriculture, geology, journalism
social ideas (27)	communism, conservatism
general quality (52)	cleanliness, permanence, resiliency
human quality (55)	cynicism, sportsmanship, stardom
asymmetric relations (25)	abstinence, paucity, precedence
symmetric relations (11)	coexistence, companionship, peace
sports (16)	archery, golf, soccer
location/time (10)	airspace, dawn, latitude
unclassified (32)	fun, haste, parenthood

3 Strongly non-countable nouns: Strategies against individuality

be similar to that of materials or aggregates, for nouns such as *trash* the seeming cause of non-countability is more indirect: Even if a use of *trash* designates entities that would otherwise be countable individuals, designating (and evaluating) them with the nominal description *trash* avoids identifying or individuating elements.

Food terms such as *chicken* are well-known as “dual-life” nouns, but the nouns of the MEATS category here are those that describe classes of meat (*poultry*, *pork*, *venison*) for which reference to the animal is named separately. While *chicken* is often used as an example of a noun with both a count and non-count use to exemplify the claim that many nouns in the lexicon are “flexible” nouns (e.g. Bale & Barner 2009: 241), this is not to be taken for granted, since, for instance, *pork* and *pig* (or *mutton* and *sheep*) are not “flexible”, that is, do not, in typical circumstances, display both a count and non-count use. This is clearly due to the fact that the reference to the animal and the meat are accomplished by two distinct nouns, whereas in the case of *chicken*, a single noun lexicalizes both types of referents.

HERBS AND SPICES, such as *coriander*, *cumin*, *fennel*, *incense*, and *nutmeg* provide another interesting puzzle. In their physical form, many members of this class (e.g. a parsley plant or sprig, or a fennel bulb) are just as easy to individuate as many other small plants or bulbs which are described by countable nouns in English (*dandelion*, *onion*), as well as countable nouns which are similarly able to divide their reference, such as *twig* or *branch*. Yet, it is presumably their use, typically as processed bits or powders, that accounts for their strongly non-countable behavior (Wierzbicka 1988).

In sum, the now-common notional contrast between individual, aggregate and substance nouns is not sufficient to explain the variety of types of non-countable nouns observed even in the domain of physical entities: Evaluativity, interaction/use, and lexical contrast all may play a role in why a given noun may be (non-)countable.

4.1.2 Eventualities

The Eventualities super-category contains nominal forms designating various events, activities, processes or states. As one might expect from previous work linking countability and aktionsart (see Mourelatos 1978, Grimm 2014 and references therein), the non-countable nouns in this category are imbalanced among types of eventualities. More nouns refer to activities, processes or states than to events and, further, the strongly non-countable nouns that do refer to events have very particular semantics.

MULTI-PARTICIPANT EVENTS enforce reference to multiple individuals or events, thus *applause* normally comprises clapping from more than one member of an audience, and *bloodshed* is used to describe the killing or wounding of multiple people.¹³ Similarly, *centralization* requires bringing multiple elements together while *dissemination* requires distributing multiple elements in multiple locations. The intrinsic plurality in these nominal descriptions most likely inhibits the use of a plural form.¹⁴

The category of COMING-INTO-/GOING-OUT-OF-EXISTENCE contains nouns which describe the beginning or the end or demise of an entity, which typically is an argument of the noun, such as *abolition*, *emergence*, *eradication*, *incineration*, or *regeneration*. Thus, *eradication* designates the end of some entity's or set of entities' existence, as in *the eradication of smallpox*, while *emergence* is the beginning of the existence of some entity or the appearance at a location. While these eventualities designate precise points in time where the entity in question passes into or out of existence, the grounds for canonical non-countability would appear to stem from the uniqueness of the events, as entities do not typically pass into or out of existence more than once.

The category of EVENTS contains a rather miscellaneous set of eventive nouns which do not fit into the categories discussed above. Those such as *atonement* or *reclamation* would also appear to be rather unique occurrences and as such resist pluralization.

The remaining categories in the Eventuality super-category are the more expected non-countable eventualities: ACTIVITIES, GRADUAL/REPEATED PROCESSES and STATES. We distinguish two types of states. In addition to MENTAL STATES, which are often cited as non-countable nouns, we include GENERAL STATES (*mayhem*, *poverty*, *unemployment*), by which we indicate nouns that refer to a general situation, equally able to be predicated of individuals and groups, and unlike the category of GENERAL QUALITY, are straightforwardly compatible with temporal localization. Many of these nouns manifest what has been termed in Grimm (2016) a "non-particularized use," that is, the nouns refer to instances of, e.g., *poverty*, but without making any claims to these instances being spatio-temporally located or being of a particular number.

¹³Some lexicographical resources note the multiple-participant facet of *bloodshed*'s meaning, as in the definition from the Oxford lexicography website lexico.com: "The killing or wounding of people, typically on a large scale during a conflict."

¹⁴An anonymous reviewer suggests that this class could constitute a morphologically singular counterpart to pluralia tantum nouns like *scissors* or *entrails*, which have been argued to be lexically plural (Acquaviva 2008), differing in that the lexical plurality is not overtly marked.

4.1.3 Phenomena

These nouns lack reference to any specific temporal or spatial location, the vagueness and unbounded nature of which is most likely the cause of their non-countability. For instance, DISEASES includes nouns which have meanings charged with physical and temporal aspects, e.g., *smallpox* or *tuberculosis* have physical causes and manifestations, but these are not the same as nouns which describe a (potentially) bounded physical entity, like *table*. Similar observations apply to DISORDERS, such as *autism* or *vertigo*, which are related to events, but cannot be reduced to particular events or states, as well as to NATURAL FORCES, such as *magnetism* or *sunshine*.

4.1.4 Abstract

The nouns in ABSTRACT are those which are not necessarily interpreted as connected to spatial or temporal dimensions. DOMAINS OF KNOWLEDGE (*forestry, psychoanalysis, voodoo*) or SOCIAL IDEAS (*federalism, materialism*) describe bodies of knowledge, ideas or cultural practices which are not embodied by one particular act or event. Qualities, both HUMAN QUALITIES (*chastity, foolishness*) and GENERAL QUALITIES (*health, toughness*), may be exemplified by acts or events, but are not co-extensional with those events. That is, the meaning of *chastity* or *foolishness* is not equivalent to the set of chaste or foolish acts. Nouns which designate relations are found in this class, too. These are distinct from nouns most often discussed under “relational nouns” such as *brother* or *neighbor*, which designate an entity in terms of the relation it stands in with respect to another entity. The nouns, whether in the SYMMETRIC RELATIONS (*accordance, relatedness*) or ASYMMETRIC RELATIONS (*governance, subordination*) category, designate the relation itself.

The nouns in the category of LOCATION/TIME describe or reference some aspect of spatial or temporal experience, as in *horseback, midair, or sundown*, but again cannot be reduced to a specific location or event. The category of SPORTS too shares the aspect of at once having physical and temporal aspects while also transcending them.

4.1.5 Unclassified

The inclusion of this category reinforces a point made at the beginning of this section, that this classification is incomplete and many unresolved issues remain. This varied group includes nouns such as *postage, slang, eyesight, and firepower*, which fit poorly in any of the categories discussed so far. No doubt a larger sample

would help to establish even more fine-grained categories in which these nouns could be located. Some interesting cases are still worth pointing out.

Manslaughter appears to be rigidly non-countable, which is odd if one takes it to be analogous to, for instance, *murder*; however, the observed uses of *manslaughter* do not appear to be directly referencing acts or events, but rather offer a classification of acts or events as falling under manslaughter or not – that is, the noun provides a second-order property, a property of properties. In a similar vein, the nouns *conduct* or *haste* do not refer an event itself, but serve as a secondary predication over an event, referring to the manner in which an event or set of events was carried out.

Another interesting case is the small group of nouns derived by *-hood*, including *fatherhood*, *motherhood*, and *parenthood*.¹⁵ Here *-hood* combines with a relational noun to derive some more abstract quality or property associated with participating in that relation. These nouns do not appear to be stative, as evidenced by their infelicitous combination with temporal modifiers (*his homelessness/?fatherhood lasted two years*), nor do they straightforwardly fit with human qualities (*composure*), which depict a quality that humans can possess or not, nor with general qualities (*cleanliness*), which characterize a situation.

In sum, the wide variation in different notional categories of non-countable nouns vividly demonstrates the challenge awaiting theories of (non-)countability. It is unlikely that there is a single, monolithic source of non-countability for which the semantics of *glitter*, *homelessness* and *archery* interact in the same way. To the contrary, it appears that many of the principles by which something is deemed non-countable, in English and across languages, have yet to be fully understood.

4.2 Notional types and coercion types

We now turn to examine if correspondences can be found between the notional categories of nouns laid out in this section and the coercion types discussed in §3. Figure 1 presents a heatmap that maps the number of nouns in each notional category manifesting each type of coercion shift. Several trends are visible upon inspecting this visualization of the data. First, as one would expect, Packaging and Event coercions are effectively in complementary distribution, with Packaging being found among nouns of the Entities super-category and Event being

¹⁵ *Womanhood* is also included in this group, although it differs semantically from those derived from a relational noun. Derivations with *-hood* are not semantically transparent, as the countable nouns *childhood*, which is temporally grounded, or *neighborhood*, which is spatially grounded, attest.

3 Strongly non-countable nouns: Strategies against individuality

found across the rest of the notional super-categories. Second, Type coercions are robustly found across the different super-categories, although are unobserved for some of the categories, such as AGGREGATES, ELEMENTS, or LOCATION/TIME. The Jargon and Archaic coercions are primarily found with the more eventive and abstract nouns. The None column, which tracks the number of nouns for which no coercions were observed, shows that across the different categories there are almost always some nouns which are rigidly non-countable, while certain notional categories, such as SPORTS or NATURAL SUBSTANCES, appear to be mostly comprised of rigidly non-countable nouns.

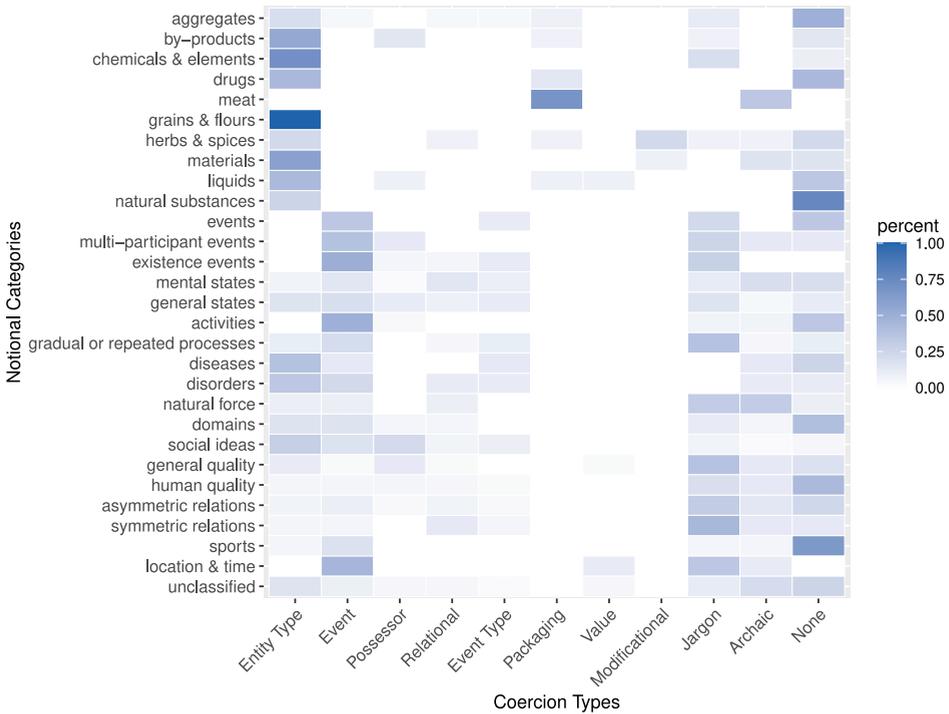


Figure 1: Heatmap showing the proportion of observed coercions in each coercion type for each notional category

A major effort for future research is to understand which types of nominals allow which types of coercions. We expect that contributing this explicit data set of coercions will help systematize this effort.

5 Variation in grammatical behavior of strongly non-countable nouns

This section investigates the general distributional characteristics of these nouns, beyond those solely concerned with countability. We ask if it is possible to detect any broad scale contrasts in grammatical environments between these strongly non-countable nouns and a group of “standard” countable nouns and hypothesize that these sets of nouns which already differ in countability status will also differ in two other aspects of their grammatical distribution. First, we expect them to differ in their propensity for occurrence in different grammatical positions, i.e. if they are more frequently governed by verbs or prepositions and what position they have in those structures, e.g. verbal subject or object. Measuring the nouns’ distribution in clausal position, e.g., use as subject, serves as a proxy for understanding their typical discourse salience (see Kaiser 2006 and references therein): Verbal subjects tend to be more salient in the discourse as a whole than nouns occurring in the object position, and similarly for nouns occurring as a nominal head modified by a prepositional construction (*the ire of parents*) as opposed to being in the complement of a preposition (*the ire of parents*). Second, we measure the “referential weight” of the nouns’ uses, tracking the amount of determination, especially definite determiner usage, the noun manifests across its occurrences. We expect countable nouns to have a higher proportion of referential (definite) uses and we use the occurrence of the definite determiner as a proxy for referential uses (while noting that this is clearly a simplification, given the complexity of the uses of the definite determiner, see Lyons 1999 i.a.). For countable nouns, on the whole, we expect more occurrences with the definite determiner and in salient argument positions (*The vase is on the table.*) while strongly non-countable nouns will occur less often with definite determiners and in non-argument positions (*The stoppages of work could not be justified by the standards of arbitral jurisprudence.*)

Together, if validated, these hypotheses would indicate that countable nouns tend toward greater discourse salient and referential uses while strongly non-countable nouns, and perhaps non-countable nouns more generally, have fewer discourse salient and referential uses. This is intuitively plausible inasmuch as countable nouns describe entities for which it is useful to regularly pick out, or individuate, the referents. To explore these hypotheses, we expanded our data set to include countable nouns with which we could contrast the 482 non-countable nouns. We selected the Core Countable nouns of Grimm & Wahlang (2021), a set

3 Strongly non-countable nouns: Strategies against individuality

of 799 nouns identified through a clustering experiment based on distributional properties shown to be predictive of countability status.¹⁶

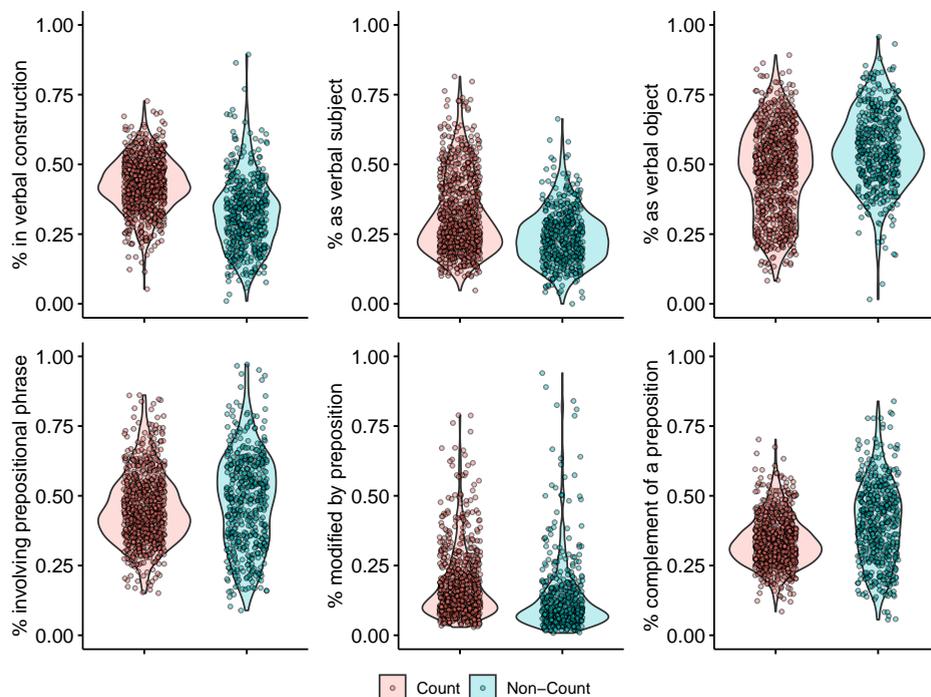


Figure 2: Comparison of distributional properties of non-count and count nouns: Percentage of occurrence of each noun in each environment

Figure 2 presents plots displaying the distribution of the grammatical positions of the countable and non-countable nouns examined. The violin plots include each noun as an individual point and the probability density of the distribution of the sample showing the general distributional trends. The upper half of Figure 2 shows in the leftmost panel the nouns' occurrence in verbal constructions generally and then the proportion of a noun's verbal occurrences as verbal subject and as object. The lower half shows their occurrence with prepositions generally, and then, relative to the total number of prepositional occurrences, the proportion as nominal head modified by a preposition and the proportion as complement of a preposition. As can be seen, countable nouns have a greater

¹⁶These properties were occurrence in the bare Plural, the bare singular, and with "unit", "fuzzy" and "other" denominators. See Grimm & Wahlang (2021) for further discussion.

propensity to be in verbal constructions and to be the subject of those constructions more often than non-countable nouns do, and conversely, non-countable nouns have a greater tendency to be in object position.¹⁷

The behaviors of the different types of nouns in prepositional phrases is more variable, especially for non-countable nouns: Non-countable nouns have a greater propensity to occur generally in prepositional phrases and to occur in the complement of prepositional phrases than countable nouns do, but what is most striking is the far greater variability among non-countable nouns than among countable nouns. Countable nouns can be seen to vary from approximately 25%–75% of occurrence in prepositional phrases with a mean tendency of 45.4%. Non-countable nouns range from hardly ever occurring in prepositional phrases (*parking, bowling*) to nearly always (*entirety, lack, emergence*), and the central tendency, at 47.7%, is far less pronounced. The same contrast occurs in measuring occurrence in prepositional complement positions, with some non-countable nouns hardly ever occurring as a complement to a preposition (*shopping, gripe*) and some nearly always doing so (*manslaughter, colonialism, disgust*). The rate of occurrence as the head of the prepositional phrase is similar for countable and non-countable nouns, although less frequent for non-countable nouns.

Figure 3 presents violin plots which display the distributional traits hypothesized to correspond to the different degree of determination and referential uses among countable and non-countable nouns. For this study, we consider the singular and plural occurrences of nouns separately, since their ability to occur without determiners differs: Plural nouns, like non-countable nouns may be bare (that is, have a “null determiner”), while this is disallowed for countable nouns.

The plots in the left panels display coarse-grained information about determination patterns. The upper-left panel shows the percentage of nouns’ occurrences *not* as bare nouns, that is, occurrences that lack any sort of quantifiers, determiners or modifiers. The lower-left panel displays the proportion of determiners found with a given noun. Here we observe a trend that holds across all the plots. There is an ordering among the mean proportion of determination for the different groups: Singular count nouns have the highest proportion of determiner or non-bare use, plural count nouns next highest and non-count nouns lowest. In the upper-left panel, non-countable nouns display a high degree of variation as to whether they occur bare, with some exclusively occurring bare (*peacetime, photosynthesis*) and some most always occurring with some sort of determination or modification (*fondness, nakedness, woodwork*). In contrast, countable nouns

¹⁷All significance tests were carried out using simple *t*-tests, and all result reported as “significant” are of $p < 0.001$. For comparisons between the distributions of singular and plural occurrences of nouns, paired *t*-tests were used. See further details in the data and code repository.

3 Strongly non-countable nouns: Strategies against individuality

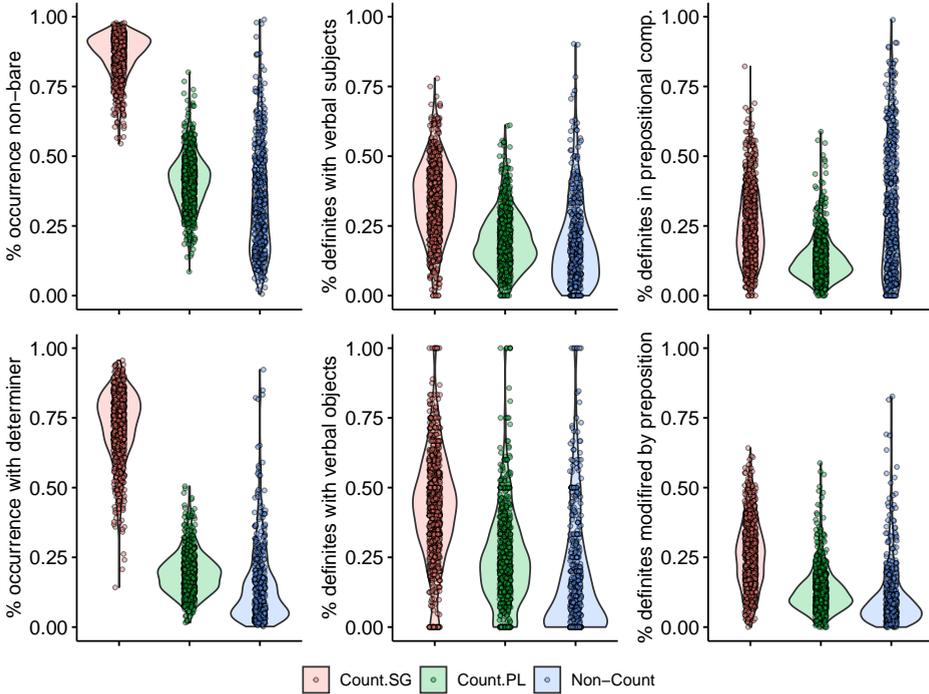


Figure 3: Comparison of determiner distributions of non-count and count nouns: Percentage of occurrence of each noun in each environment

are more tightly grouped for singular and plural occurrences, with a substantial proportion of plural uses occurring bare, no doubt largely due to generic uses.

The four middle and right-hand side panels track the occurrence of definite determiners in different syntactic positions. We calculate the proportion of definite uses among all uses of a given noun. For the mid-upper panel, the proportion of definite uses of count nouns, both in singular and plural uses, and non-count nouns are given for all occurrences in subject position. Count singular uses have the greatest proportion of definite uses (ranging from 0%–77% of their occurrences, mean tendency of 34.6%), while count plural uses and non-count nouns have a lower proportion of definite occurrences (0%–61%, mean 19.7%, and 0%–90%, mean 18.6%, respectively). While non-count nouns have the lowest proportion of definites in subject position, the distribution of plural uses of count nouns does not differ significantly in subject position from that of non-count nouns, although both differ significantly from the distribution of the singular uses of the count nouns.

Turning to nouns in the verbal object position (lower-mid panel) and those in the complement of a prepositional phrase (upper-right panel), the occurrences of definites among singular and plural uses of count nouns and non-count nouns all do differ significantly. While each type has nouns that have no or all occurrences as definites, making the ranges of proportions from 0% to 100% for all three, their central tendencies differ: count singular 34.6%, count plural 24.2% and non-count 17.6%.¹⁸

The general trend holds for the distribution of definite determiners with prepositions as well, with count singular nouns having a higher proportion than count plural nouns which is itself higher than non-count nouns. The definite uses of non-countable nouns in the complement of prepositions, as would be expected from Figure 2, show a large range of variation, although the central tendencies of count singular, count plural and non-count nouns differ significantly in the expected directions. The lower-right panel shows that many non-countable nouns show a high proportion of their definite uses when the noun is modified by a preposition, which appears to primarily occur when the non-countable noun is related to another referent, e.g. *the acidity of the soil*, i.e., the non-countable noun has a particular referent, here an acidity value, in relation to another referent (*soil*).

Overall, we are able to observe that the strongly non-countable nouns have a greater tendency to occur in syntactic positions which correspond to lesser discourse salience – in particular as verbal objects and complements of prepositions. Further, on average, they occur more often bare, that is, with less determination overall and, in particular, fewer definite uses, especially in argument positions. This is to be expected if countable nouns are more individuated, easily identified, and referred to, while non-countable nouns are those that are less individuated and less easy to establish as referents (see Grimm 2018 and references therein).

6 Outlook

This paper has presented a systematic study of a large number of non-countable nouns, tracking various aspects relevant for the ongoing discussions in the countability literature, including notional categories, as well as contextual and grammatical behavior. While this data set is to date far larger than any collected for

¹⁸These figures exclude copular constructions, although there too we found similar (statistically significant) trends. Count singulars have a higher proportion of definites in subject position than count plurals which in turn have a higher proportion than non-count nouns. However, for copular objects, while count singulars had a greater proportion of definite uses overall, this only contrasted significantly with count plurals, but not with non-count nouns.

this purpose, we must again emphasize the preliminary nature of the results here. Within the confines of this paper, we have only been able to bring forth a number of contrasts present in this data set, but certainly not all of them, nor have we *explained* these contrasts in detail beyond contributing some informal remarks.

It remains to be seen how current models of the count/non-count contrast need to be extended or revised to account for the various non-countable nouns examined here. Most of the countability literature has delivered analyses from the perspective of part-structures, such as mereology, a natural enough approach for nouns falling under entities or eventualities. Yet, for many of the nouns observed in the data set, such as *fatherhood*, *eyesight*, or *eloquence*, pressing them into the mould of a part-structure analysis seems far less convincing, pointing to the need for a more general theory of countability contrasts.

Acknowledgments

The first author would like to thank the organizers of SinFonJA 12 in Brno at the Faculty of Arts of Masaryk University for an invigorating special session on topics related to countability, and especially Mojmir Dočekal and Marcin Wągiel. The authors jointly would like to thank the current and former members of the Quantitative Semantics Lab at the University of Rochester, especially Rebecca Friedman, Kai Schenck, Matthew Sundberg, Katherine Trice, and Aeshaan Wahlang. We further would like to acknowledge two anonymous reviewers whose comments resulted in a number of improvements to this paper.

References

- Acquaviva, Paolo. 2008. *Lexical plurals: A morphosemantic approach*. Oxford: Oxford University Press.
- Allan, Keith. 1980. Nouns and countability. *Language* 56(3). 41–67. DOI: 10.2307/414449.
- Baayen, R. Harald, Richard Piepenbrock & Hedderik van Rijn. 1996. *Celex2*. Philadelphia, PA: Linguistic Data Consortium.
- Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 15(1). 5–16. DOI: 10.1002/9780470758335.ch13.
- Bale, Alan C. & David Barner. 2009. The interpretation of functional heads: Using comparatives to explore the mass/count distinction. *Journal of Semantics* 26(3). 217–252. DOI: 10.1093/jos/ffp003.

- Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174(1). 99–149. DOI: 10.1007/s11229-009-9686-6.
- Davies, Mark. 2009. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14(2). 159–190. DOI: 10.1075/ijcl.14.2.02dav.
- De Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *LREC*, vol. 14, 4585–4592. Reykjavik, Iceland: European Language Resources Association (ELRA).
- Deal, Amy Rose. 2017. Countability distinctions and semantic variation. *Natural Language Semantics* 25(2). 125–171. DOI: 10.1007/s11050-017-9132-0.
- Grimm, Scott. 2014. Individuating the abstract. In Urtzi Etxeberria, Anamaria Fălăuș, Aritz Irurtzun & Bryan Leferman (eds.), *Proceedings of Sinn und Bedeutung 18*, 182–200. Vitoria-Gasteiz: University of the Basque Country. <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/312>.
- Grimm, Scott. 2016. *Crime investigations: The countability profile of a delinquent noun*. *Baltic International Yearbook of Cognition, Logic and Communication* 11(1).
- Grimm, Scott. 2018. Grammatical number and the scale of individuation. *Language* 94(3). 527–574. DOI: 10.1353/lan.0.0230.
- Grimm, Scott & Beth Levin. 2017. Artifact nouns: Reference and countability. In *Proceedings of the 42nd Annual Meeting of the Northeastern Linguistics Society*, vol. 2, 45–64. University of Massachusetts, Amherst, MA: GLSA.
- Grimm, Scott & Aeshaan Wahlang. 2021. Determining countability classes. In Tibor Kiss, Francis J. Pelletier & Halima Husić (eds.), *The semantics of the count/mass distinction*, 785–794. Cambridge: Cambridge University Press.
- Husić, Halima. 2020. *On abstract nouns and countability*. Bochum: Ruhr-Universität Bochum. (Doctoral dissertation). DOI: 10.13154/294-7203.
- Kaiser, Elsi. 2006. Effects of topic and focus on salience. In Christian Ebert & Cornelia Endriss (eds.), *Proceedings of Sinn und Bedeutung 10*, vol. 1, 139–154. Berlin, Germany: ZAS Papers in Linguistics. <https://ojs.ub.uni-konstanz.de/sub/index.php/sub>.
- Kiss, Tibor, Francis Jeffrey Pelletier, Halima Husić, Roman Nino Simunic & Johanna Marie Poppek. 2016. A sense-based lexicon of count and mass expressions: The Bochum English countability lexicon. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik,

3 Strongly non-countable nouns: Strategies against individuality

- Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *LREC 2016*, vol. 10, 2810–2814. Paris, France: European Language Resources Association (ELRA).
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511605789.
- Mourelatos, Alexander. 1978. Events, processes, and states. *Linguistics and Philosophy* 2(3). 415–434. DOI: 10.1007/bf00149015.
- Payne, John & Rodney Huddleston. 2002. Nouns and noun phrases. In Rodney Huddleston & Geoff Pullum (eds.), *Cambridge grammar of the English language*, 323–524. Cambridge: Cambridge University Press. DOI: 10.1017/9781316423530.006.
- Sutton, Peter R. & Hana Filip. 2016. Counting in context: Count/mass variation and restrictions on coercion in collective artifact nouns. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Semantics and Linguistic Theory*, vol. 26, 350–370. University of Texas: Linguistic Society of America & Cornell Linguistics Circle. DOI: 10.3765/salt.v26i0.3796.
- Sutton, Peter R. & Hana Filip. 2019. Singular/plural contrasts: The case of informational object nouns. In Julian J. Schlöder, Dean McHugh & Floris Roelofsen (eds.), *Proceedings of the 22nd Amsterdam Colloquium*, 367–376. Amsterdam: ILLC.
- Sutton, Peter R. & Hana Filip. 2020. Informational object nouns and the mass/count distinction. In Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab (eds.), *Proceedings of Sinn und Bedeutung 24*, vol. 2, 319–335. Osnabrück: Osnabrück University. DOI: 10.18148/sub/2020.v24i2.900.
- Wierzbicka, Anna. 1988. *The semantics of grammar*. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/slcs.18.
- Zamparelli, Roberto. 2020. Countability shifts and abstract nouns. In Friederike Moltmann (ed.), *Mass and count in linguistics, philosophy, and cognitive science*, 191–224. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/lfab.16.09zam.

