A great deal of evidence based on a great many instances: A usage-based comparative corpus study of two English nominal constructions

Online Proceedings of UK-CLA Meetings
http://uk-cla.org.uk/proceedings
Vol 4: 249 – 272

KIM EBENSGAARD JENSEN

University of Copenhagen

ebensgaard@hum.ku.dk

Abstract
In a reference grammar of English for Danish students, Hjulmand & Schwarz (2015: 137) state that, when translating from Danish, "'en hel del' is a good/great deal of *in front of uncountable nouns, but* a good/great many *in front of countable nouns in the plural". This claim calls for empirical support. With significant distributions of count nouns vs. non-count nouns, a study of COCA suggests that the claim holds up at least for American English. However, the claim ultimately belongs to what Harder (2015; see also Gregory 1967) calls incomplete accounts. In the perspective of usage-based linguistics, such a claim would leave out information potentially useful to Danish learners of English. Drawing on principles from construction grammar (e.g. Goldberg 1995; Croft 2001) and variationist cognitive sociolinguistics (Pütz et al. 2014), this paper presents a usage-based comparative corpus study of the two constructions. Drawing on data from COCA, a distinctive collexeme analysis (Gries & Stefanowitsch 2004) shows that, not only do the constructions differ in terms of preference for count vs. non-count nouns, they also have different preferences for specific individual nouns and semantic classes of nouns. Moreover, variety-centered multidimensional scaling analyses and heatmaps indicate that the patterns of use of the constructions display register variation. In addition, a lexical richness analysis reveals differences in constructional productivity.*

Key words: cognitive sociolinguistics, corpus linguistics, nominal constructions, register variation, usage-based construction grammar.

1. Introduction

In their grammar of English for Danish learners, Hjulmand & Schwarz (2015: 137) make the following statement:

> When translating from Danish, note than [sic.] 'en hel del' is *a good/great deal of* in front of uncountable nouns, but *a good/great many* in front of countable nouns in the plural: *We spent a great deal of time on the project. I have a great many friends in London.*

Needless to say, this piece of information is useful to learners of English as a foreign language. However, many statements of this type may ultimately be

based on "received wisdom", begging the question: is that really what speakers of English do, and, if yes, is that really all they do? The second part of the question is relevant to the social turn in cognitive linguistics (e.g. Croft 2009; Harder 2010), as it questions the monolithic character that such a statement imposes upon the linguistic phenomenon in question. Note that, in this paper, 'monolithic' is used as an antonym of 'variationist': a monolithic statement about a linguistic phenomenon is one that assumes non-variation in usage or simply ignores variation. The question posed above forms the basis of what we shall investigate in this paper. Drawing on data from the *Corpus of Contemporary American English* (Davies, 2016) (henceforth, COCA), we aim to investigate whether *a great/good deal of N* and *a great/good many N* have countability preferences and, focusing on registers, whether or not they display variation across varieties of language use.

It makes sense to position the study within the theoretical framework of usage-based construction grammar, whose focus is on constructions, their patterns of use, and their communicative and cognitive functions. A *caveat lector* is in place here: since the present study draws on data from COCA, it only addresses the use of the two constructions in American English. Moreover, the study is not to be taken as an exhaustive account of the two constructions.

2. Usage-based construction grammar

In usage-based construction grammar, a construction is defined as "an entrenched routine [...] that is generally used in the speech community [...] and involves a pairing of form and meaning" (Croft, 2005: 274). The definition of a construction as a form-meaning pair is a basic construction-theoretical principle (e.g. Fillmore et al. 1988; Goldberg 1995) while the notion of entrenched routines in speech communities is a general tenet in usage-based linguistics as such (Bybee & Hopper 2001; Kemmer & Barlow 2000). The main principles of usage-based construction grammar have been neatly described by Patten (2014: 91):

> On this model, humans are not innately programmed with grammatical knowledge; instead, all aspects of language are learned from the input (or rather from the speaker's linguistic experiences). Both language learning and language change involve the speaker inductively generalizing over instances to form mental schemas (or constructions) which are represented in the language system. On a usage-based model then, constructions are simply conventionalized chunks of linguistic knowledge … From this, it follows that the storage and organization of grammatical knowledge is dependent on, and can change according to, patterns of use …"

Thus, grammar is a network of constructions as exemplars (Bybee 2013) in the sense that the network includes knowledge of not only form and meaning but also linguistic and non-linguistic association patterns (as defined by Biber et al. 1998). Consequently, constructions and subconstructions may be item-specific, item-class-specific (Croft 2003; Tomasello 2003), or context-specific, and the constructional network may feature redundancies across levels of constructional categorization. Drawing on the above insights, we define a

construction as a conventionalized pairing of form and meaning which is entrenched in a speech community, which is acquired inductively, and which includes information on form, meaning, and contexts of use. Moreover, constructions may be fully lexically filled, partially schematic, or fully schematic. A constructional network may thus feature several levels of abstraction and redundancy, ranging from very specific lower level subconstructions to very schematic superconstructions.

While there has been a tendency within general linguistics to treat syntactic phenomena as monoliths, construction grammarians have recently begun to take an interest in constructional variation, probably as a result of the emergence of cognitive sociolinguistics (e.g. Pütz et al. 2014). The idea that variation is relevant to grammar has been at the core of sociolinguistic research for decades:

> It seems to be implied that grammatical description ... is not infinitely extensive, that a sort of exhaustiveness is possible there. This is perhaps an illusion. Any monolithic grammatical model which aims to be productively explicit will, as a consequence, fail to be explicit about and overtly recognize all the possible relationships in a language which might be called grammatical (Gregory 1967: 179)

Construction grammarians and, more broadly, cognitive linguists have come to the realization that, as Harder (2015) puts it, "[a] so-called structural description that ignores variation is simply incomplete". Variation emerges alongside the language system (or as part of the language system) in speakers from multiple agent interactions in different contexts, and thus constructional variation is part of speakers' knowledge of constructions.

Taking cues from variationist sociolinguistics, construction grammarians have identified patterns of constructional variation across user-based varieties (Halliday et al. 1964; McArthur 1992; Quirk 1989), such as lects (e.g. Levshina 2014), and use-based varieties (Halliday et al. 1964; Quirk 1989; McArthur 1992), such as registers (Schönefeld 2013).

3. Overview of the constructions

The two constructions addressed in this study are nominal constructions whose primary propositional act function is that of reference (Searle 1969: 26-29; Croft 2001: 66). As Hjulmand & Schwarz (2015: 137) suggest, they differ in terms of countability. In cognitive-linguistic terms, the two constructions differ in terms of constitution construals (Croft & Wood 2000: 66-69; Talmy 2000: 48-50) of the N-referent such that, in *a great/good many N*, it is construed as a multiplex of bounded entities, while, in *a great/good deal of N*, it is construed as an unbounded and internally homogenous mass(-like) entity. The former is essentially an example of what Croft (2000: 3) calls a counting construction, as it features a plural trigger in the form of the quantifier *many*, while the latter is an example of a bare singular construction.

What both constructions have in common – and share with the Danish *en hel del N* construction – is that they seem to have a discursive function of referent-intensification such that they both quantify the N-referent and

intensify the quantity as illustrated in the following examples:

(1) That being said, it is the conventional gasoline engine that still rules the American road. It is the power plant of the lion's share of small pickup trucks, sport utility vehicles, and passenger cars. And it can consume *a great deal of fuel*. (COCA 2006 ACAD MechanicalEng)

(2) Remarkably, there is *a good deal of evidence* in support of much of this seemingly fanciful story. (COCA 2015 MAG HistoryToday)

(3) I think *a great many Americans* thought why aren't these two people running for president. (COCA 2002 SPOK KingWknd)

(4) "She doesn't know. She doesn't know *a good many things*. She thinks you're some sort of... Dark Sister (COCA 1992 FIC BkSF:Meri)

(5) Participants described some HRCs as beginning with *a great deal of theory* and little practical application. (COCA 2009 ACAD CommCollegeR)

(6) *A great many theories* and beliefs and revisions had clung to me over the course of the years, and also dropped away from me one by one – but the belief in the sanctity of labor never left me. (COCA 1993 FIC Bk:BonetoBone)

In all examples, other quantifying constructions could have been used, such as *much N*, *many N*, and *a lot of N*. When comparing *it can consume a lot of fuel* to (1) or *many Americans thought* to (3), it should be clear that (1) and (3) intensify the amount of fuel and the quantity of Americans; one might perhaps describe this intensifying function as a type "dramatic effect" function.

4. Data and method

The study is based on data in COCA (Davies 2016), which is a corpus of American English. Consisting of 533,788,932 words, it covers the period 1990-2015 within the following registers: spoken language (SPOK), fiction (FICT), magazines (MAG), newspapers (NEWS) and academic writing (ACAD). All instances of *a great deal of* followed by any noun, *a good deal of* followed by any noun, *a great many* followed by any noun, and *a good many* followed by any noun were retrieved from the corpus. An obvious limitation is that the current study only addresses instances of these constructions in which the N-position is not premodified. This yielded 6,384 instances of *a great/good deal of N* and 1,036 instances of *a great/good many N*.

4.1. Countability

To address the issue of countability, the occurrences of singular and plural forms in the N-positions were quantified. Seeing that, according to Croft (2000: 3), counting constructions are associated with countability while the bare singular construction is associated with non-countability, a majority of plural forms in the N-position indicates countability and a majority of singular forms indicates non-countability. Thus, if *a great/good many N* appears with

a significant majority of plurals (which can be expected due to the presence of *many*) and *a great/good deal of N* appears with a significant majority of singulars, then the statement in Hjulmand & Schwarz (2015: 137) holds up.

4.2. Collostructional analysis

Patterns of construction-lexeme interaction may provide insights into the semantics of a constructional position (Gries & Stefanowitsch 2004; Stefanowitsch & Gries, 2003, 2005), as lexemes tend to be attracted to constructions that they overlap with semantically. Moreover, semantic patterns among lexemes in a position may be indicative of item-class-specific subconstructions within the constructional network. In the context of the present study, collostructional analysis can help shed light on the countablity of *a great/good many N* and *a great/good deal of N*. If the latter attracts primarily nouns with mass(-like) referents, then it supports Hjulmand & Schwarz (2015: 137). It can also shed some light on the constructional networks, as attraction patterns may be indicative of item-specific and item-class-specific subconstructional exemplars.

Simple collexeme analysis (Stefanowitsch & Gries 2003) is a type of collostructional analysis that allows the analyst to gage, via a measure called collostruction strength, the degree of attraction or repulsion between a constructional position and a lexeme (a so-called collexeme). Collostruction strength is calculated on the basis of the following frequencies:

- Lexeme in construction
- Lexeme in other constructions
- Other lexemes in construction
- Other lexemes in other constructions

The total sum of these frequencies in a 2-by-2 table is run through a Fisher-Yates Exact test or a similar statistical test (for alternative tests, see Stefanowitsch & Gries 2003: 238-239), using Gries (2014). The outcome is a *p*-value which serves as the collostruction strength. Seeing that log-likelihood allows for rather fine-grained distinctions among collostruction strengths, log-transformation has been applied in this study. This procedure is applied to all lexemes in the construction, resulting in a list of collexemes in the constructional position which is ranked by collostruction strength.

Seeing that there may be two variants of either construction (a *great*-variant and a *good*-variant), it might be interesting to also address distinctive collexemes (i.e. collexemes that strongly prefer one construction over another) in either pair of variants. To this end, distinctive collexeme analysis (Gries & Stefanowitsch 2004), which allows the analyst to measure the degree of preference of a lexeme for a construction among two or more constructions, was applied to each pair of constructional variants. As with simple collexeme analysis, distinctive collexeme analysis generates a collostruction-strength measure. When measuring the degree of attraction to, or preference for, one construction over another, the following frequencies serve as input:

- Lexeme in construction A
- Other lexemes in construction A
- Lexeme in construction B

- Other lexemes in construction B

As with simple collexeme analysis, the total sum of these in a 2-by-2 table is run through a Fisher-Yates Exact test (which in this case was also log-transformed), using Gries (2014). The procedure is applied to all lexemes that appear in the constructions.

This description of the application of collostructional analyses is, due to space constraints, admittedly rather minimalistic. Readers who want a deeper understanding of collostructional analysis are encouraged to read Gries & Stefanowitsch (2004) and Stefanowitsch & Gries (2003, 2005).

4.3. Register variation, multidimensional scaling and heatmapping

Turning to register variation, our focus remains on construction-noun interaction as we investigate the degree to which the N-positions display variation across the five registers in terms the lexemes that appear in those positions. The main hypothesis here is that the constructions are likely to display cross-register variation the N-position. In addressing this, we apply two quantitative analyses: multidimensional scaling (e.g. Borgatti 1997) and heatmapping (e.g. Perrot et al. 2015).

A similarity measure, multidimensional scaling visualizes degrees of similarity among data objects as proximities and distances in two-dimensional space. Similarity measures can help analysts relatively objectively identify groups of items that behave similarly and separate them from items that behave differently in terms of the factors included in the analysis. In this study, multidimensional scaling is applied for each of the four constructional variants at register-level and is based on Euclidean distance calculations of the frequencies of occurrences of each lexical item in the N-position (henceforth, lexemes in the N-positions will be referred to as N-elements); all N-elements were taken into account. This technique visualizes how alike or different the COCA registers are in terms of the N-elements in each constructional variant.

This gives us a birds' eye view of variation in construction-noun interaction across the five registers with focus on the registers themselves. However, if we want a more fine-grained perspective on the interaction between N-elements and registers, we can apply heatmapping combined with agglomerative cluster analysis, which is a means of visually representing intercategorical associations between data objects. In this case, the categories are N-elements and the five COCA registers. Heatmapping visualizes associations between two datapoint dimensions by having the dimensions – that is, registers and N-elements – as axes in a diagram, and the interaction points between the items on the axes being colored fields such that de degree of association is indicated via a color code: the brighter the color the stronger the association. This way, our heatmaps show the degree of association between each N-element and each register. Agglomerative hierarchical cluster analysis is a similarity measure that generates clusters in a dendrogram based on similarities and differences between the points in question. In heatmaps that include cluster analysis, a heatmap is set up in which the interacting datapoints on either axis are clustered into dendrograms. In this study, normalized frequencies (per

million words) of occurrence of N-elements in each constructional variant were used as input into the generation of a heatmap, and the N-elements on the one hand and the registers on the other were clustered using Euclidean distance calculations and Ward-clustering. The heatmap analyses only take into account the Top 15 attracted N-elements in either construction.

The two techniques complement each other fairly well with one providing a bird's eye view of which varieties are similar to, and different from, each other in terms of N-elements in the constructions, and the latter providing a more fine-grained view on associations between N-elements and registers.

4.4. Lexical growth curves and constructional productivity

Given the usage-based framework of this study, an investigation of the productivity of the two constructions is warranted, as we can assume that they may display differences in productivity. The analyses presented here draw on Bybee's (1995) definition of productivity: the more lexemes that appear in a constructional position, the more productive the position. This means that, if there are few hapax legomena among the occurrences of a construction, it has low productivity, while constructions whose occurrences feature many hapax legomena have high productivity.

In many ways, constructional productivity is parallel to lexical richness. Lexical richness is traditionally considered a textual feature: texts that have a high number of types and a low number of tokens are considered richer than texts that have a high number of tokens and a low number of types. Given the parallels between constructional productivity and lexical richness, Shibuya (2015; see also Zeldes 2013) has argued that measures of lexical richness can also be applied as measures of constructional productivity. Lexical growth curves constitute one such measure. This method is based on LNRE-analysis (large number of rare events) as described in Baayen (2008). In the present study, extra- and intrapolated curves were generated on the basis of actual curves so as to allow for comparison. As Shibuya (2015; see also Zeldes 2013) has argued, this method can be applied at the level of constructional positions to see how lexically rich a position is which then gives us an idea of how productive it is as well. Thus, by comparing the lexical growth curve of one constructional position to that of another, we can get an idea of which is more productive. We will apply lexical growth curve analysis to the N-position of each variation of the two constructions.

Productivity is immensely complicated, both theoretically and empirically (for a detailed discussion, see Zeldes 2013: 17-47). Bybee's (1996) approach to, and definition of, productivity is only one out of many (see Zeldes 2013: 17-21). Moreover, lexical growth curves and based on LNRE and other lexical richness measures are by no means the only, and not necessarily a perfect, means of gaging constructional productivity (for a discussion of productivity measures in morphology and syntax, see Zeldes 2013: 48-137). However, as Zeldes (2013: 70) points out, they have the advantage that they provide a "visual heuristic for productivity judgments". Even so, there are caveats to keep in mind, such as the fact that extra- and intrapolations in growth curves, while based on actual curves, do not correspond to actual curves. With this in mind,

the productivity analyses presented in this paper should be taken with a grain of salt.

5. Findings and discussion

5.1 Overall frequencies and distribution of number in the N-position

Table 1 provides an overview of the overall frequencies of the two constructions:

Table 1: Overall frequencies

Construction		Frequency	
		n	%
a great/good deal of N	a great deal of N	5,455	73.42
	a good deal of N	939	12.64
	Total	6,394	86.06
a great/good many N	a great many N	809	10.89
	a good many N	227	3.06
	Total	1,036	13.94

As seen in the '%' column, *a great/good deal of N* makes up no less than 86.06% (n = 6,394) of the total number of occurrences of the two constructions, while *a great/good many N* makes up only 13.94% (n = 1,036). The very fact that the constructions do not occur equally frequently (with the distribution being statistically significant at $p<0.05$) suggests that they do not have the same status in the language system.

Moreover, given the size of the corpus, we can assume that the constructions are not particularly frequent in American English. A comparative diachronic frequency analysis (Jensen 2015) of *a great deal of N* and *a great many N* in the *Corpus of Historical American English* (Davies 2010) clearly shows that both constructions are decreasing in use in American English with the latter being on the verge of extinction (see Figure 1).

Let us turn to the distribution of plural and singular forms in the N-position. Table 2 accounts for this distribution. The N-position in *a great/good many N* is entirely realized by plurals, thus suggesting that *a great/good many N* appears with countable nouns. 98.01% (n = 6,267) of N-elements in *a great/good deal of N* are singulars, while a minority of 1.99% (n = 127) consists of plurals. The distribution is significant at $p<0.05$ which confirms that *a great/good deal of N* prefers uncountable nouns.

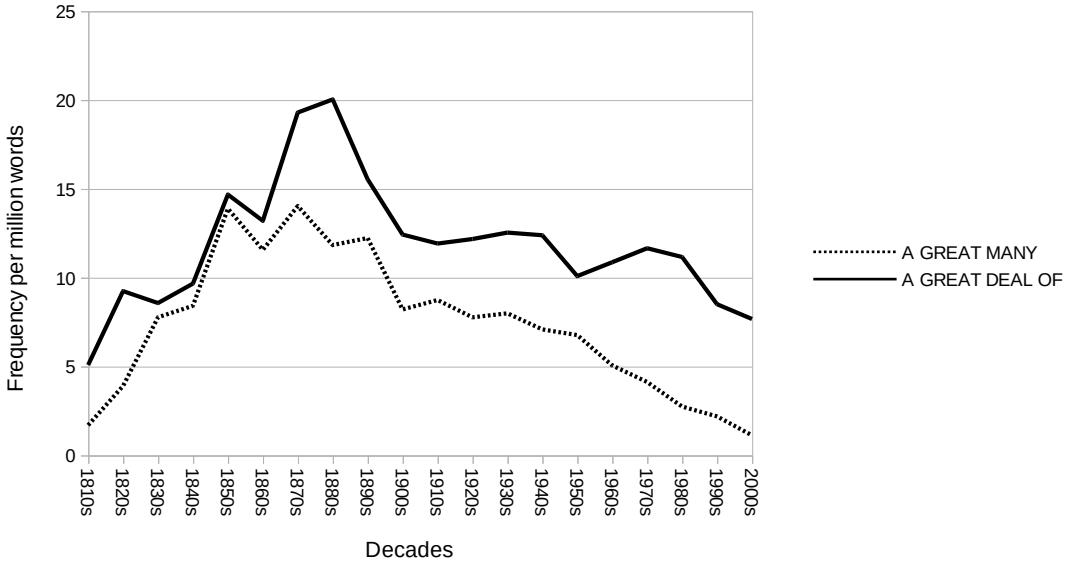

Figure 1: Frequency of use (normalized to per million words) of *a great deal of N* and *a great many N* in COHA (Jensen 2015)

Table 2: Distribution of plurals and singulars in the N-position

Construction		Distribution			
		N		%	
		plural	singular	plural	singular
a great/good deal of N	a great deal of N	117	5,338	2.14	97.86
	a good deal of N	10	929	1.06	98.94
	Total	127	6,267	1.99	98.01
a great/good many N	a great many N	809	0	100.00	0.00
	a good many N	227	0	100.00	0.00
	Total	1,036	0	100.00	0.00

5.2 Construction-noun interaction

Consider Tables 3 and 4 which account for the Top 15 attracted items to *a great/good deal of N* and *a great/good many N* respectively. *A great/good deal of N* attracts lexemes with primarily unbounded referents, while *a great/good many N* attracts lexemes whose referents are clearly bounded and countable.

Table 3: Top 15 attracted items in *a great deal of N* and *a good deal of N*

a great deal of N		a good deal of N	
Lexeme	Collostruction strength	Lexeme	Collostruction strength
time	2783.2088	time	1011.6079
money	2355.0613	attention	269.2940
attention	1769.3395	money	231.2468
respect	1118.7901	thought	96.5864
information	692.1078	criticism	91.9155
effort	634.9072	evidence	89.7134
pain	510.5325	information	88.5792
concern	491.9191	damage	84.3783
thought	491.0499	confusion	80.3052
trouble	491.0081	research	73.1989
work	484.3652	leeway	71.4347
confidence	477.9809	influence	70.9818
research	444.1826	discussion	64.4747
difficulty	441.5605	skepticism	56.6337
interest	398.9796	trouble	56.2705

Table 4: Top 15 attracted items in *a great many N* and *a good many N*

a great many N		a good many N	
Lexeme	Collostruction strength	Lexeme	Collostruction strength
people	1762.2702	year	225.9709
thing	536.3462	people	170.4272
other	179.7814	thing	48.8342
American	178.8864	American	48.4425
year	93.1483	other	36.8898
question	39.4537	woman	29.8476
man	36.7916	picture-taker	24.8469
Republican	31.6274	congresspeople	19.2420
woman	29.8472	Republican	19.2257
similarity	27.6016	incompetent	18.9214
factor	27.5407	whirligig	18.5200
mind	24.6560	Indian	18.3401
asthmatic	24.0281	okapi	17.9907
variation	23.2393	Black	17.5974
hanap	22.9816	politician	16.7513

Overall, this seems to support Hjulmand & Schwarz (2015: 137). However, Tables 3 and 4 also reveal further usage patterns which are arguably relevant to learners who wish to use the two constructions. Firstly, *time*, *money* and *attention* are very strongly attracted to *a great/good deal of N*; it may even be the case *that a great/good deal of time*, *a great/good deal of money*, and *a great/good deal of attention* constitute item-specific subconstructions akin to what is advocated by Croft (2003) and Tomasello (2003). Moreover, as is clear in Table 4, *a good many N* appears to have a preference for a specific semantic class of nouns – namely, nouns that refer to people and categories of people such as *people*, *American*, *Republican*, *man*, *woman*, *picture-taker*, *congresspeople*, *Indian*, *asthmatic* and *politician* as well as nouns that refer to people via ATTRIBUTE-FOR-PERSON metonymy such as *incompetent*, *Black* and *mind*. *Thing* also ranks very highly on the list, suggesting an item-specific subconstruction. Tables 3 and 4 also suggest that there may be differences between the *great*- and the *good*-variants of either construction in terms construction-noun interaction, which calls for a distinctive collexeme analysis. Table 5 accounts for the Top 20 distinctive collexemes in *a good deal of N* vs. *a great deal of N* and *a good many N* vs. *a great many N* respectively. One interesting distinctive pattern seems to apply to *a great many N* vs. *a good many N*, as *a good many N* appears to attract nouns that refer to time units such as *year*, *minute*, *month* and *hour* as well as *time* in the sense of OCCASION. Regarding *a great deal of N* vs. *a good deal of N*, on the face of it, there do not seem to be any semantic patterns among the distinctive collexemes; however, a more in-depth analysis of the entire range of distinctive collexemes may well reveal semantic patterns (for instance, there seem to be more N-elements that express emotion in *a great deal of N* than in *a good deal of N*).

Our collostructional analyses of the construction-noun interaction indicate that the basic countability distinction applies not just at the level of countability as a grammatical category, but also at the conceptual-semantic level. This should not be a major surprise within a cognitive-linguistic framework, in which grammar encodes meaning, such that counting constructions with plural triggers prompt construals of multiplexity and singular constructions prompt construals of uniplexity (see also Talmy 2000: 48-50). More importantly, they also reveal additional patterns in that the constructions seem to have both item-class-specific and item-specific preferences. These preferences can, in a usage-based perspective, be argued to be specific lower-level exemplars in the constructional networks.

5.3 Register variation

A good starting point when addressing constructional register variation in a corpus-linguistic perspective is the overall distribution constructions across the five COCA registers. Figure 2 offers an overview of the frequencies, normalized to per million words, of the constructional variants in the five registers of COCA. To enable a relatively fine-grained view, the figure distinguishes between the *great*- and *good*-variants of either construction.

Table 5: Top 20 distinctive collexemes in a good deal of N vs. a great deal of N and a good many N vs. a good deal of N

A good deal of N		A great deal of N		A good many N		A great many N	
Lexeme	Collostruction strength	Lexeme	Collostruction strength	Lexeme	Collostruction strength	Lexeme	Collostruction strength
time	366.500	difference	151.900	year	350.636	people	252.631
body	115.613	respect	136.613	minute	60.339	thing	81.079
shade	115.613	effort	91.800	month	60.339	child	30.225
criticism	111.705	pride	85.293	senator	60.339	factor	25.173
chatter	111.705	difficulty	72.690	soldier	60.339	mind	25.173
grunting	77.057	control	69.468	friend	50.590	change	20.127
steam	77.057	money	65.469	time	50.590	instance	15.087
text	77.057	interest	55.899	hour	50.590	letter	15.087
wine	77.057	pain	55.899	administration	33.246	mistake	15.087
luck	73.769	work	54.417	ally	30.135	problem	15.087
leeway	60.024	power	47.386	analyst	30.135	reader	15.087
ambivalence	54.597	affection	44.177	artist	30.135	similarity	15.087
space	51.047	detail	44.177	bather	30.135	skill	15.087
water	45.142	anxiety	42.368	box	30.135	species	15.087
material	44.654	latitude	41.018	boy	30.135	study	15.087
anti-Semitism	42.012	compassion	37.860	century	30.135	variation	15.087
behavior	42.012	disagreement	37.860	college	30.135	administrator	10.053
building	42.012	love	37.860	congresspeople	30.135	adult	10.053
competition	42.012	opposition	37.860	Conservative	30.135	arm	10.053
distress	42.012	risk	37.860	cut	30.135	Canadian	10.053

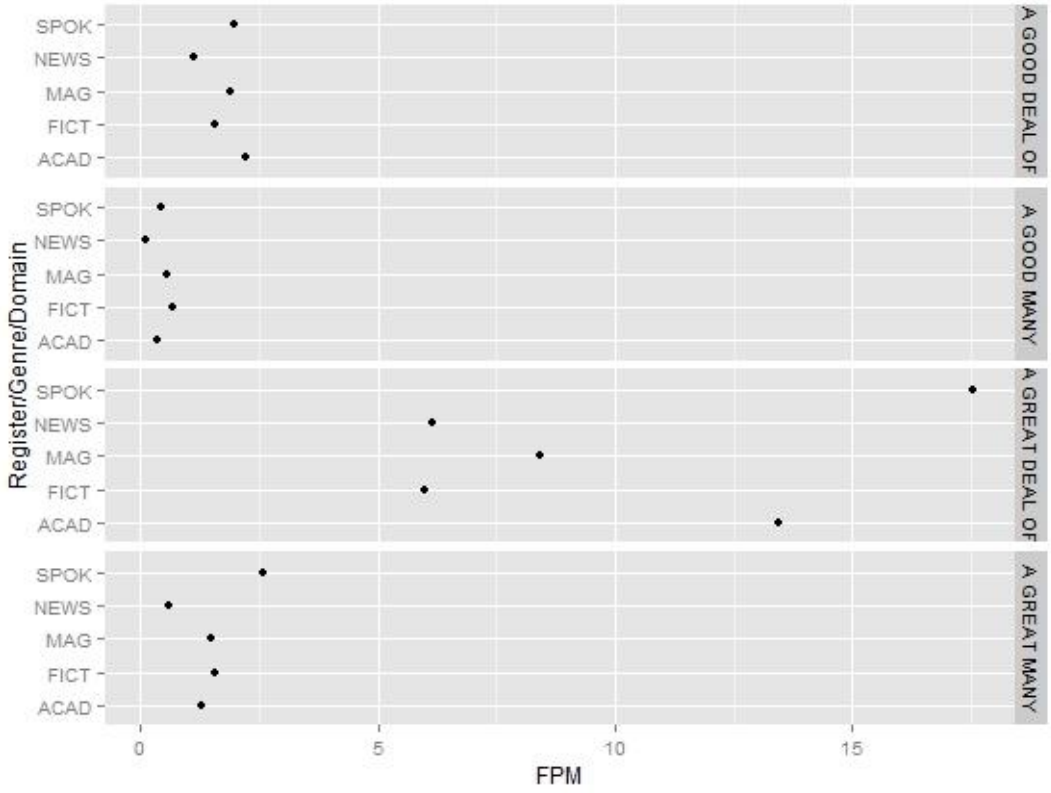

Figure 2: Distributions across the five registers of COCA

As is quite clear already here, the constructional variants do not display monolithic behavior, as they all differ in frequency across the registers. *A great deal of N* in particular displays considerable degrees of variation, being much more frequent in the spoken and academic registers than in the three other registers. *A good deal of N* is much more frequent in the spoken register than in the other registers. This could suggest that the *great*-variants of the construction are perhaps more strongly associated with spoken American English than with written varieties.

If we want to address how similar or different the registers are in terms of the frequency the N-elements in each constructional variant, we can apply multidimensional scaling. This type of analysis, while not able to give us details of this aspect of the behavior of the constructions, enables us to take a bird's eye view on how similarly or differently the constructions behave within the five registers of COCA. Figures 3-6 present the results of the multidimensional scaling analyses.

The newspaper, fiction and magazine registers appear near each other in Figure 3, while the academic and spoken registers appear isolated. This suggests overlaps in terms of N-elements in *a great deal of N* in the three registers in the top left corner of the figure, while, in the spoken and academic registers, the N-position behaves more idiosyncratically in this respect.

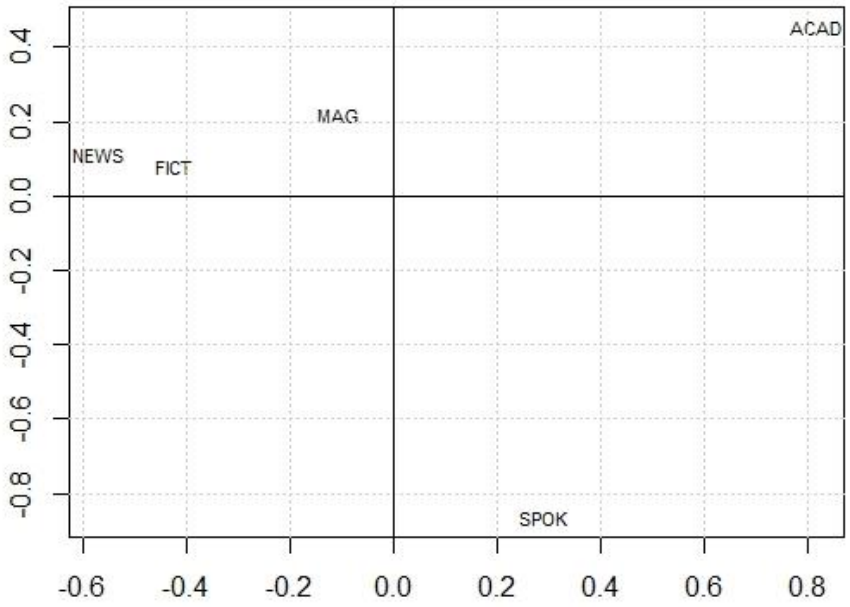

Figure 3: Multidimensional scaling – registers and *a great deal of N*

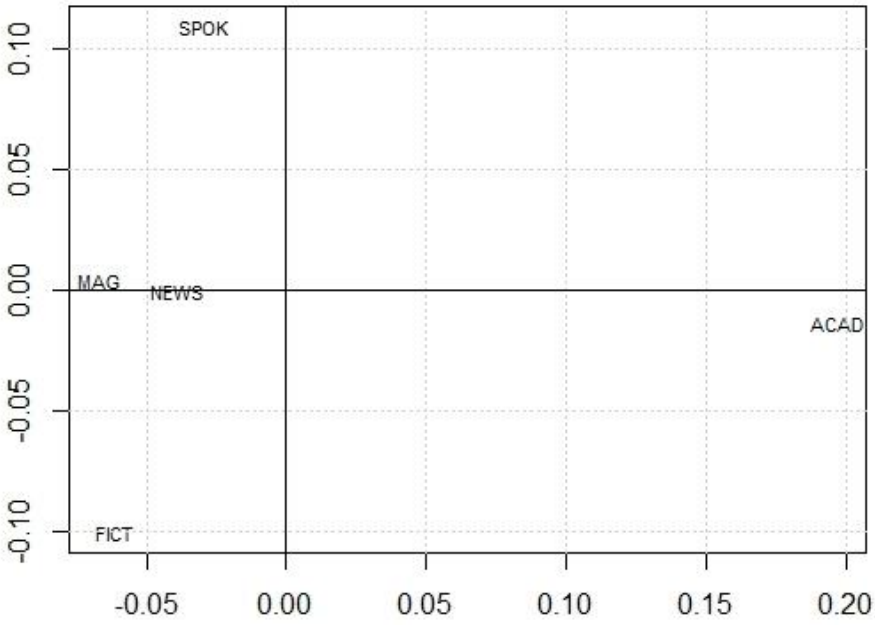

Figure 4: Multidimensional scaling – registers and *a good deal of N*

As Figure 4 indicates, the *good*-variant of the construction behaves differently in terms of cross-register variation. Here the magazine and newspaper registers group together, while the academic, spoken and fiction registers appear in isolation, suggesting that the N-position is idiosyncratic in each of those registers while there are overlaps between the magazine and newspaper registers. Overall, Figure 3 suggests that there are more cross-register differences in terms of the discursive behavior of the constructions when it comes to N-elements in the *good*-variant of *a great/good deal of N* than in the *great*-variant. Turning to *a great/good many N*, Figure 5 shows that, in the *great*-variant, the newspaper, magazine and academic registers group together, while the fiction and spoken registers appear in isolation, suggesting that this variant is used with different N-elements in the spoken register and the fictional register than in the three other registers. As with the *good*- and *great*-variants of *a great/good deal of N*, the *good*- and *great*-variants of *a great/good deal of N* seem to behave slightly differently from each other in terms of N-elements. As seen in Figure 6, the newspaper and academic registers group together – which is similar to the *great*-variant. The fiction and spoken registers appear in isolation which also mirrors the *great*-variant. However, unlike in the *great*-variant, the magazine register appears in isolation in Figure 6, suggesting in this register, the *good*-variant, is considerably idiosyncratic in the magazine register in terms in terms of N-elements.

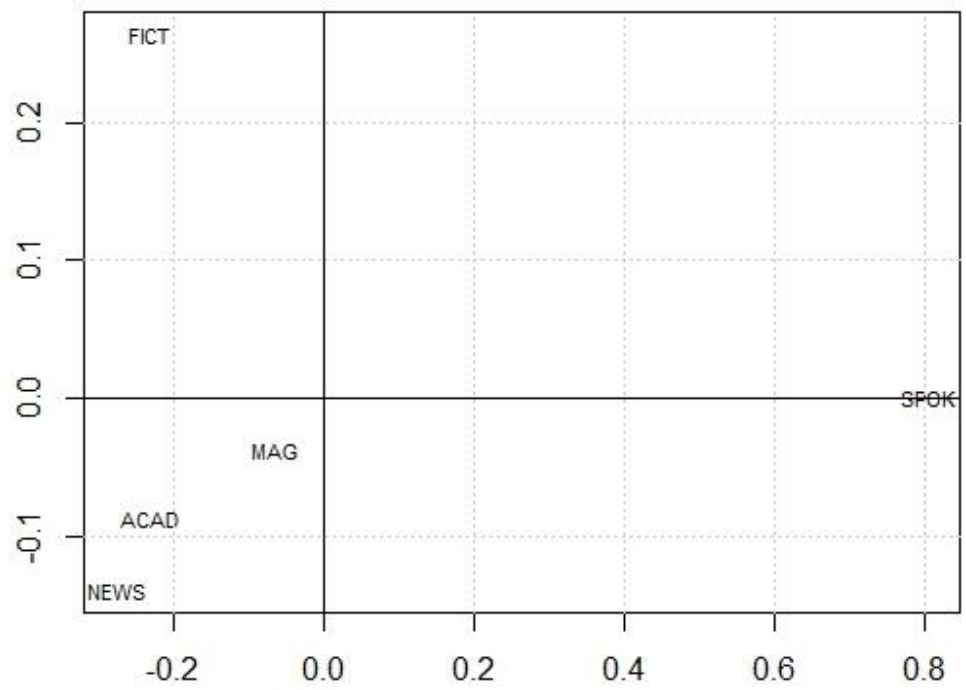

Figure 5: Multidimensional scaling – registers and *a great many N*

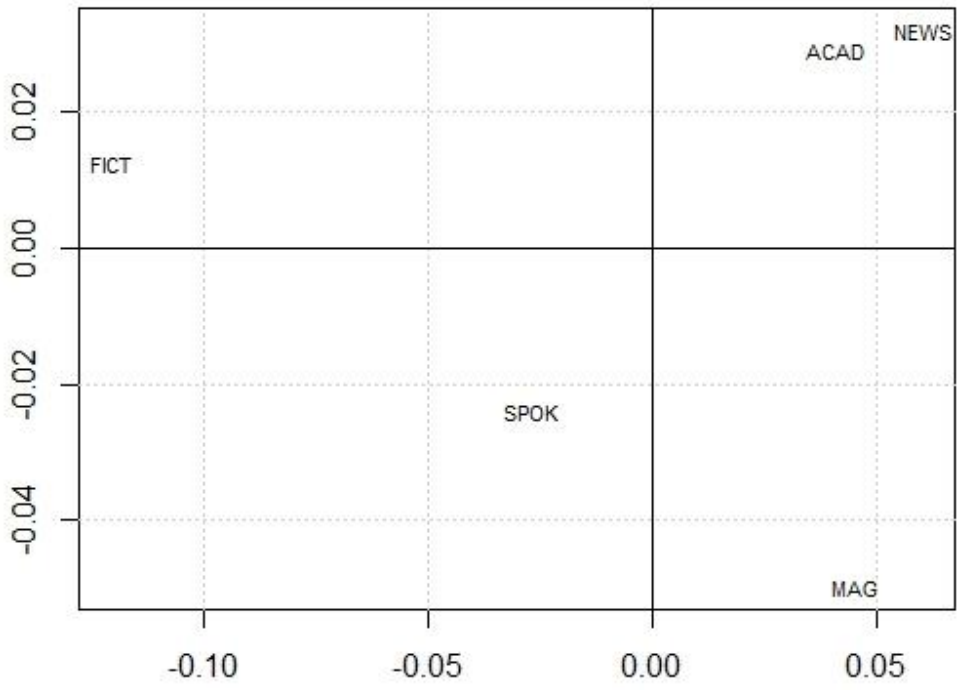

Figure 6: Multidimensional scaling – registers and *a good many N*

As mentioned earlier, multidimensional scaling is useful in providing a bird's eye perspective on register variation, but it does not enable us to address the interaction between individual lexemes in a construction and the registers that the construction occurs in.

One technique that enables analysts to take this perspective is heatmapping. Figures 7-10 present heatmaps of the interaction between the fifteen most strongly attracted verbs in each of the variants if the two constructions (see Tables 3 and 4). The horizontal axis lists the lexemes, and the vertical axis lists the registers. The degree of association is represented via color coding such that, the brighter the color, the stronger the association of an item on the horizontal axis to an item on the vertical axis. This means that white indicates a very strong association while dark orange indicates a weak association.

Figures 7 and 8 account for the patterns of lexeme-register association among the 15 most attracted nouns to the N-positions of *a great deal of N* and *a good deal of N*. As is clear, *time* is strongly associated will all registers except the academic register in *a good deal of N*, while, in *a great deal of N*, it is the most strongly associated with the magazine register and more strongly associated with the academic and fiction registers than the newspaper and the spoken registers. Also, *money* is generally more strongly associated with all registers in *a great deal of N* than in *a good deal of N*, but it is not as strongly associated with the academic register as it is with the other registers. Also, in

both variants, *attention* is the most strongly associated with the academic register. Note also that *money*, *attention*, and *time* are separated off from the other nouns in a cluster in a *great deal of N* while *time* appears in its own cluster in a *good deal of N*. This difference in clustering may be indicative of the *good*- and *great*-variants being different constructions.

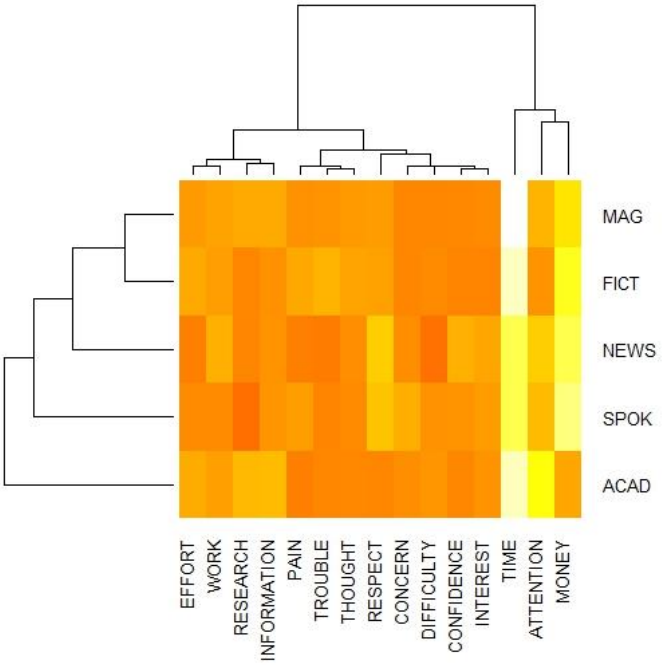

Figure 7: Heatmapping of lexeme-register association in *a great deal of N*

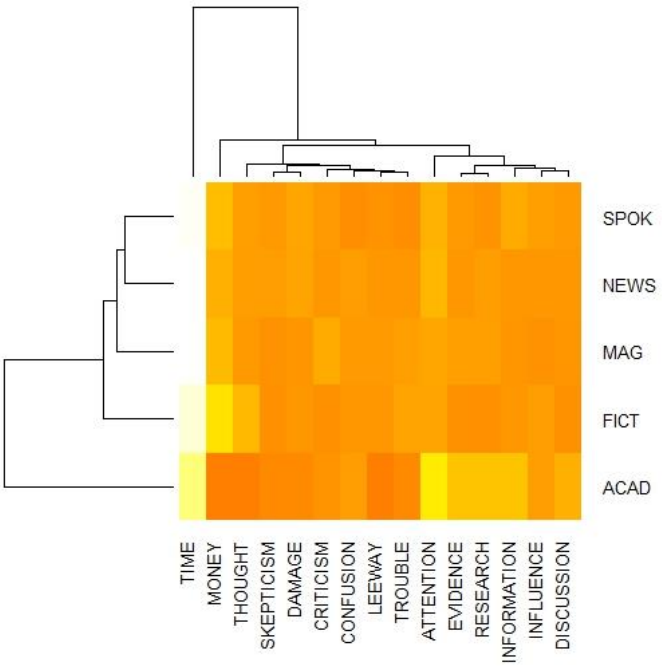

Figure 8: Heatmapping of lexeme-register interaction in *a good deal of N*

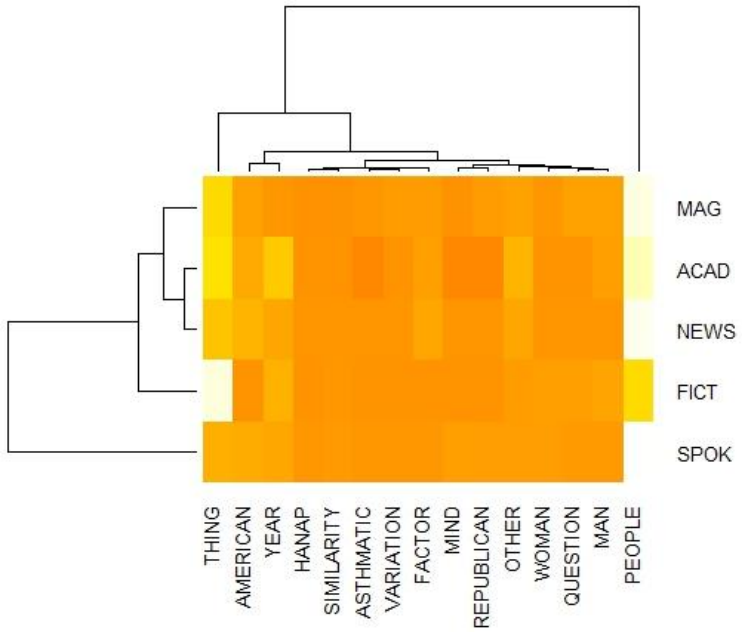

Figure 9: Heatmapping of lexeme-register interaction in *a great many N*

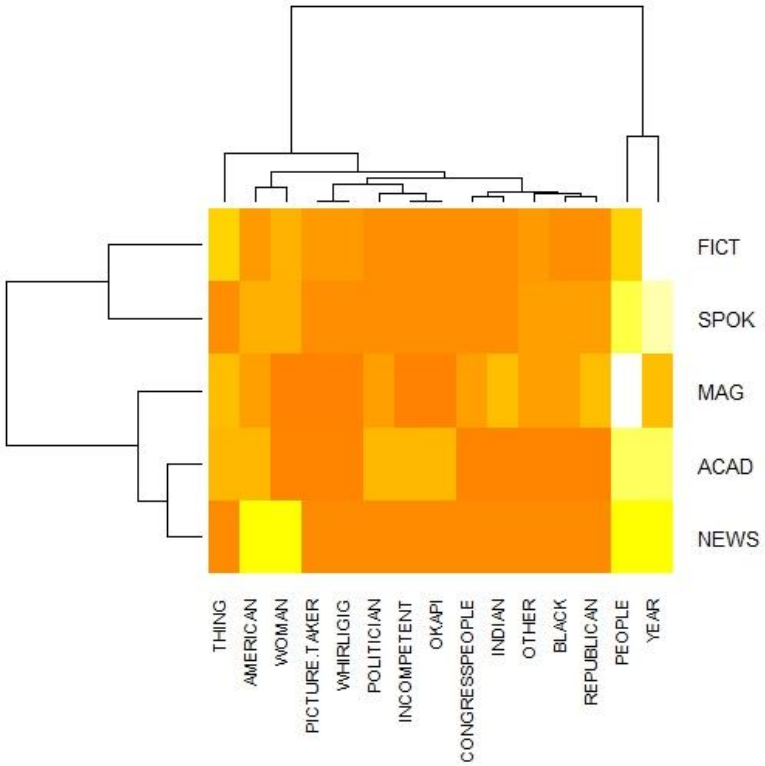

Figure 10: Heatmapping of lexeme-register associations in *a good many N*

In *a great many N* and *a good many N*, *thing* is more strongly associated with the fiction register than with the other registers. *American* is has a weaker association with fiction in *a great many N* than in the other registers and a slightly stronger association with the newspaper register than with the other registers. In contrast, the association between *American* and the newspaper register is much stronger in *a good many N*, but, in contrast, *American* seems

to have the weakest association with the fiction register. *Year* is generally strongly associated with all registers except the magazine register *in a good many N*, while, in *a great many N*, it is more strongly associated with the academic register than the others; in comparison with *a good many N*, *year* is generally more weakly associated with all registers in *a great many N*. This should not be a surprise seeing that the distinctive collexeme analysis in Table 5 indicates that time-period nouns seem to prefer *a good many N* over *a great many N*. Note also that *people* is strongly associated with all registers except fiction in Figure 9, and it appears in its own cluster; this may suggest that *a great many people* is a lexically filled subconstruction (probably subsumed under the slightly more schematic *a great many* N_{PEOPLE} subconstruction) which is not register-sensitive when it comes to the non-fiction registers. In contrast, *people* and *year* form a cluster together in Figure 10. This does not mean that the two form one subconstruction, but, as the color coding indicates, the two N-elements are more unevenly distributed over the COCA-registers in *a good many N* than in *a great many N*.

The multidimensional scaling analyses and the heatmap analyses suggest that neither *a great/good deal of N* nor *a great/good many N* is monolithic, as the constructions display cross-register variation in terms of the N-elements. These analyses also suggest that the *great*- and *good*-variants of either construction display differences in terms of construction-lexeme interaction across the five registers above and beyond semantic preferentiality.

5.4 Constructional productivity

Turning to constructional productivity, we are interested in comparing the N-position in either construction in terms of productivity as defined by Bybee (1995). In Figure 11, we see the lexical growth curves of *a great/good deal of N* vs. *a great/good many N*. As the figure shows, the former is slightly more lexically rich than the latter, which means that *a great/good deal of N* occurs across the corpus with a wider range of lexemes in the N-position. Consequently, the N-position in *a great/good deal of N* is more productive than the N-position in *a great/good many N*.

While this is interesting as it is, Figure 11 does not account for the *good*- and *great*-variants. Remember that, in Table 5, there were traces of differences at subconstructional level in that *a good many N* differs from *a great many N* in being preferred by time-unit nouns. Seeing that the productivity of a constructional position is ultimately also related to construction-lexeme interaction, it might be interesting to compare all four constructional variants in terms of the productivity of the N-position. As Figure 12 shows, while there is only a very small difference in productivity between *a great deal of N* and *a good deal of N*, the difference between *a great deal of N* and *a good many N* roughly corresponds to that between the two overarching constructions. More interestingly perhaps, the difference between the *great*- and *good*-variants of *a great/good many N* is larger than the variants of *a great/good deal of N*, with *a good many N* being more productive than *a great many N*.

While admittedly somewhat of a sketchy analysis, and further analysis based on more data is needed, Figure 12 does seem to suggest that the four

constructional variants are not equally productive in American English with *a great many N* being the least productive variant while *a good deal of N* is the most productive one.

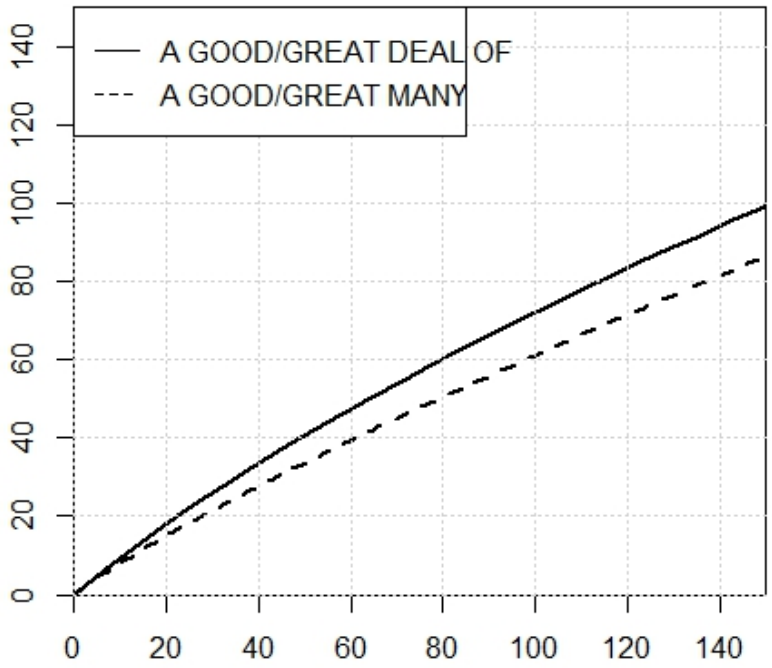

Figure 11: Lexical growth curves for *a great/good deal of N* vs. *a great/good many N*

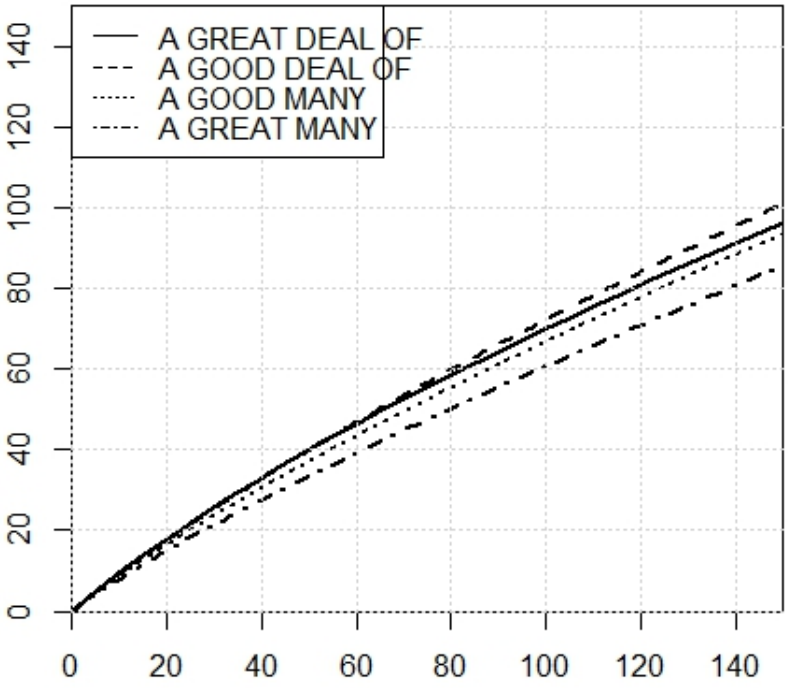

Figure 12: Lexical growth curves for each constructional variant

6. Concluding remarks

Hjulmand & Schwarz (2015: 137) advise Danish learners of English to translate *en hel del N* into *a great/good deal of N* with non-count nouns and *a great/good many N* with count nouns but do not offer other information on how the two English constructions are used. Thus, the two constructions are presented in a rather monolithic fashion. Prompted by this, this paper presents a corpus-based study, anchored in usage-based construction grammar, of the two constructions in COCA (Davies 2016) which assesses whether this statement holds up in terms of countability and studies further patterns of language use.

We found that, in terms of countability, the basic assumption – and probably received wisdom – that the statement in Hjulmand & Schwarz (2015: 137) is based on does not appear to be wrong. However, in applying colllostructional analysis (Gries & Stefanowitsch 2004; Stefanowitsch & Gries 2003), multidimensional scaling (Borgatti 1997), heatmapping (Perrot et al. 2015), and lexical growth curves (Zeldes 2013; Shibuya 2015), we found that the constructions are not monoliths, as they display cross-register variation in construction-noun interaction. Moreover, the interaction between the constructions and N-elements involve details beyond mere countability, as the constructions clearly prefer both individual nouns and specific semantic classes of nouns. This means that the constructional networks in question are probably rich in item-specific and item-class-specific lower-level subconstructional exemplars. Furthermore, the constructions are not used uniformly across the registers documented in COCA. We also learned that neither construction is particularly frequent in American English at least.

Ultimately, then, we can ask whether Danish learners of English are done a service or a disservice when advised to use *a great/good deal of N* and *a great/good many N*, seeing that neither construction is very frequent. Moreover, one could make the point that, if instructed to use those two constructions, learners would benefit greatly from also learning how they are actually used, what N-elements they tend to appear with, and what contexts they are used in (and how they are used in these contexts). Perhaps, it would be pedagogically wiser to introduce learners to a range of different English nominal constructions that Danish learners could use when they would use *en hel del N* in Danish rather than just *a great/good deal of N* and *a great/good many N*. Of course, this requires that pedagogical grammar books be based on actual empirical observations of language in use rather than received wisdoms about aspects of English grammar. Hopefully, the present paper has shown that a good theoretical-methodological framework for this is usage-based construction grammar and corpus linguistics.

References

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Borgatti, S. P. (1997). Multidimensional Scaling. Retrieved December 21, 2016, from http://www.analytictech.com/borgatti/mds.htm

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J. L. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49–69). Oxford: Oxford University Press.

Bybee, J. L., & Hopper, P. J. (2001). Introduction to frequency and the emergence of linguistic structure. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (Vol. 45, pp. 1–26). Amsterdam: John Benjamins.

Croft, W. A. (2000). Countability in English nouns denoting physical entities: A Radical Construction Grammar analysis. Retrieved from https://www.unm.edu/~wcroft/Papers/Count.pdf

Croft, W. A. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.

Croft, W. A. (2003). Lexical rules vs. constructions: A false dichotomy. In H. Cuyckens, T. Berg, R. Dirven, & K.-U. Panther (Eds.), *Motivation in Language: Studies in Honor of Günter Radden* (pp. 49–68). Amsterdam: John Benjamins.

Croft, W. A. (2005). Logical and typological arguments for Radical Construction Grammar. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions* (pp. 273–314). Amsterdam: John Benjamins.

Croft, W. A. (2009). Toward a social cognitive linguistics. In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 395–420). Amsterdam: John Benjamins.

Croft, W. A., & Wood, E. J. (2000). Construal operations in linguistics and artificial intelligence. In L. Albertazzi (Ed.), *Meaning and Cognition: A Multidisciplinary Approach* (pp. 51–78). Amsterdam: John Benjamins.

Davies, M. (2010). Corpus of Historical American English (COHA). Retrieved December 30, 2016, from http://corpus.byu.edu/coha/

Davies, M. (2016). Corpus of Contemporary American English (COCA). Retrieved December 22, 2016, from http://corpus.byu.edu/coca/

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, *64*(3), 501–538.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Gregory, M. (1967). Aspects of varieties differentiation. *Journal of Linguistics*, *3*(2), 177–198.

Gries, S. T. (2014). *Coll.analysis 3.5. A Script for R to Compute/Perform Collostructional Analyses*.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on "alternations." *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Halliday, M. A. K., McIntosh, A., & Strevens, P. (1964). *Linguistic Sciences and Language Teaching*. London: Longman.

Harder, P. (2010). *Meaning in Mind and Society: A Functional Contribution to the Social Turn in Cognitive Linguistics*. Berlin: Mouton de Gruyter.

Harder, P. (2015, February). *Substance(s) and the Rise and Imposition of Structure(s)*. Presented at the Substance and structure in linguistics, University of Copenhagen, Copenhagen, Denmark.

Hjulmand, L.-L., & Schwarz, H. (2015). *A Concise Contrastive Grammar of English for Danish Students* (4th ed.). Frederiksberg: Samfundslitteratur.

Jensen, K. E. (2015, September). *A Great Many Things and a Great Deal of Stuff: A Synchronic and Diachronic Corpus Study of Two Nominal Constructions in American English*. Presented at the Academic Day, Aalborg University.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. vii–xxviii). Stanford, CA: Stanford University Press.

Levshina, N. (2014). Geographic variation of *quite + ADJ* in twenty national varieties of English: A pilot study. *Yearbook of the German Cognitive Linguistics Association*, *2*(1), 109–126.

McArthur, T. (1992). Variety. In T. McArthur (Ed.), *The Oxford Companion to the English Language* (pp. 1081–1082). Oxford: Oxford University Press.

Patten, A. L. (2014). The historical development of the *it*-cleft: A comparison of two different approaches. In N. Gisborne & W. B. Hollmann (Eds.), *Teory and Data in Cognitive Linguistics* (pp. 87–114). Amsterdam: John Benjamins.

Perrot, A., Bourqui, R., Hanusse, N., Lalanne, F., & Auber, D. (2015). Large interactive visualization of density functions on big data infrastructure (pp. 99–106). *IEEE Symposium on Large Data Analysis and Visualization 2015 October 25–30, Chicago, Il, USA*.

Pütz, M., Robinson, J. A., & Reif, M. (2014). The emergence of Cognitive Sociolinguistics: An introduction. In M. Pütz, J. A. Robinson, & M. Reif (Eds.), *Cognitive Sociolinguistics: Social and Cultural Variation in Cognition and Language Use* (pp. 1–22). Amsterdam: John Benjamins.

Quirk, R. (1989). Language varieties and standard language. *JALT Journal*, *11*(1), 14–25.

Schönefeld, D. (2013). It is … quite common for theoretical predictions to go untested (BNC_CMH). A register-specific analysis of the English go un-V-en construction. *Journal of Pragmatics*, *52*, 17–33.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language* (30. print). Cambridge: Cambridge University Press.

Shibuya, Y. (2015, July). *Lexical and Constructional Richness of Adjectives: A Diachronic Study*. Presented at the 13th International Conference on Cognitive Linguistics, Northumbria University, Newcastle upon Tyne, United Kingdom.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243.

Stefanowitsch, A., & Gries, S. T. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, *1*(1), 1–43.

Talmy, L. (2000). *Toward a Cognitive Semantics – Vol. 1: Concept Structuring Systems*. Cambridge, MA: MIT Press.

Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Zeldes, A. (2013). *Productivity in Argument Selection: From Morphology to Syntax*. Berlin: Mouton de Gruyter.