



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A multivariate analysis of diachronic variation in a bunch of NOUN

Citation for published version:

Shao, B, Cai, Y & Trousdale, G 2019, 'A multivariate analysis of diachronic variation in a bunch of NOUN: A construction grammar account' *Journal of English Linguistics*, vol. 47, no. 2, pp. 150-174. DOI: 10.1177/0075424219838611

Digital Object Identifier (DOI):

[10.1177/0075424219838611](https://doi.org/10.1177/0075424219838611)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of English Linguistics

Publisher Rights Statement:

The final version of this paper has been published in *Journal of English Linguistics*, Vol/Issue, Month/Year by SAGE Publications Ltd, All rights reserved. © <Author's name>, year of publication. It is available at: <http://<Acronym>sagepub.com/>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



**A multivariate analysis of diachronic variation in
a bunch of NOUN: a construction grammar account¹**

Authors:

Bin Shao¹, Yingying Cai¹ and Graeme Trousdale^{2*}

Affiliations

¹Zhejiang University of Finance and Economics, ²University of Edinburgh

* = Corresponding author

Introduction

English has a set of constructions, such as binominal phrases like *a jug of wine* or *a loaf of bread*, whose function seems to align with sequences of measure words and nouns in classifier languages like Chinese (Allan 1977; Xiao & McEnery 2010:41-42). Phrases like these are usually labeled as instances of the English partitive construction, which is defined by Smitterberg (2006) as a noun phrase consisting of a partitive noun (i.e., the ‘part’), followed by the preposition *of* and a nominal complement, denoting the ‘whole.’ This paper is concerned with recent developments in a particular instance of this kind of pattern, namely *a bunch of NOUN*, whose traditional use – where *bunch* is the head noun – can be seen in (1); emphasis is added in all examples unless otherwise stated:

(1) He has, therefore, made the curtain in the background of the same crimson color, and the white is diffused by a letter which lies on the table; and ***a bunch of flowers*** is likewise introduced for the same purpose. (1851, Non-fiction [COHA])

More recent uses of *a bunch of NOUN* have provoked some public debate. For instance, during one of the 2016 Prime Minister’s Questions sessions in the United Kingdom House of Commons, former Prime Minister David Cameron used the phrase *a bunch of migrants* to refer to refugees at a camp in Calais. He was subsequently criticized by Labour MPs and members of the general public on Twitter, who considered his expression as ‘dehumanising’, ‘callous’ and ‘inflammatory’ (Love 2016). Such criticisms may be associated with speakers’ awareness that *bunch* typically co-occurs with nouns denoting inanimate referents as in *a bunch of flowers* in (1) above. In this study, we explore,

among other things, how such negative prosody may have arisen.

Previous studies have already shed some light primarily on the synchronic uses of *a bunch of NOUN*, on the basis of corpus data (Brems 2003, 2010, 2011; Francis & Yuasa 2008); for instance, Brems (2011:176-191), using data extracted from the Collins Wordbank corpus, considers three functions of *a bunch of NOUN* in contemporary English, i.e., lexical head use, quantifier use and valuing quantifying use (see further section 2 below). The present paper complements such research by conducting a multivariate analysis of recent change in *a bunch of NOUN* sequences, based on data from the Corpus of Historical American English (COHA; Davies 2010). In particular, we explore relationships between semantic, pragmatic and discourse properties of both partitive and quantifier uses and their diachronic development, explaining such changes in the framework of construction grammar.

The approach to diachronic construction grammar we adopt in this article is essentially that of Traugott & Trousdale (2013), albeit with the addition of a quantitative component. We take the construction (i.e., a conventional pairing of form and meaning) to be the basic unit of linguistic analysis; constructions vary in complexity and schematicity, and are arranged in a hierarchical inheritance network, known as the constructicon; a language user's constructicon is the product of lifelong generalization across usage events (see further Langacker 1987; Goldberg 1995, 2006, 2013; Croft 2001; Hoffmann & Trousdale 2013). From the perspective of language change, we focus here on what Traugott & Trousdale (2013) refer to as 'post-constructiconalization changes,' i.e., the changes in morphosyntax and in various domains of meaning (semantics, pragmatics and discourse context) which may take place following the creation of a new

construction). In particular, we consider how a detailed quantitative analysis might refine some of the ideas that Traugott & Trousdale (2013) propose regarding post-constructionalization changes (given that the account in Traugott & Trousdale (2013) is qualitative). We therefore see the present article as having an equal focus on two central issues: (i) a quantitative, corpus-driven analysis of the recent history of *a bunch of NOUN* in American English and (ii) the consequences of the results of such a study for our understanding of the nature of aspects of constructional change.

The remainder of the paper is organized as follows. In section two we sketch some influential previous studies concerned with partitive and quantifier constructions from various theoretical perspectives, with particular focus on historical developments. Section three describes the method and the data of our study. Section four is the results section, focusing on changes in meaning. In section five we provide an explanation of the results using principles of diachronic construction grammar, along with some further data regarding morphosyntactic changes. Finally, section six contains a summary and some concluding thoughts.

Previous studies on *a bunch of NOUN* and related patterns

When analyzing *a bunch of NOUN*, descriptive grammar books tend to focus on the word class of *bunch*. Quirk *et al.* (1985:250), for example, regard *bunch* as a partitive noun and *a bunch of* as a partitive construction which is often used with plural count nouns, as in *a bunch of flowers/keys*. It is stated that *bunch* can also modify nouns referring to people, as in *a bunch of teenagers*, but this usage

is said to be a feature of informal language. On the basis of the data from the Longman Spoken and Written English Corpus, Biber *et al.* (1999:248) contend that *bunch* is a collective noun with a fairly general meaning and is normally combined with a plural count noun, so it can be termed as a quantifying collective. In addition, such corpus data indicate that *a bunch of*, functionally similar to *a group of* and *a set of*, is highly productive in terms of its collocations, occurring together with more than 100 different nouns.

From the viewpoint of some typological linguists, *bunch* can be considered as having features somewhat like those of classifiers. It is used to identify objects in close distribution, as in *a bunch of roses*. For example, Lehrer (1986) claims that the semantics of *bunch* is related not just to quantity but also to arrangement.² In terms of quantity, *bunch* refers to a large number of items; with respect to the arrangement dimension, *bunch* deals with either objects arrayed within a close distance to one another (*grapes* and *bananas*, etc.) or those bound up together (*carrots*, *parsley* or *flowers*, etc.). Taking a functional-typological perspective, Zhang (2017:60-61) finds that many English classifier-like words, as in *a head of cattle* and *a sheet of paper*, “encode similar physical aspects and share most semantic parameters with Chinese numeral classifiers proper.” Consequently, English classifier-like words, including *bunch*, are called ‘quasi-numeral classifiers’. Zhang (2017:40) also points out that collective-arrangement classifiers have both quantitative and qualitative semantic features. The Chinese classifier *chuan*, as in *yī chuàn pútáo* ‘a bunch of grapes’ is a good example of this, and it can be inferred that *bunch*, the English equivalent of *chuan*, might also be located on a quantity-quality semantic continuum.

Cognitive linguists have also explored the semantics of the English expression

a bunch of from a more theoretical perspective. Langacker (1991:88), for instance, points out that initially *a bunch of carrots* had the literal sense of a bundle of vegetables tied together in some fashion. However, more recently *bunch* has taken on a different sense which foregrounds the ‘quantifier’ meaning. As such, *a bunch of carrots* may refer to a certain number of carrots irrespective of whether they are bundled together. Langacker (2008:343-344) provided a different analysis of partitive constructions. He points out that *flock of geese* can be profiled as either a bounded unit (flock) or the mass of geese, indicating that such a pattern may have two possible construals (one partitive, the other quantificational); this may be a factor in the regularly attested pathway of grammaticalization of quantifiers from partitives. He also suggests that the semantic changes affecting *lot* and *bunch* are so great that they have become quantifiers, to the extent that *a {lot/bunch} of Noun* can only “profile the quantified mass (NOUN), not the quantifying unit (*lot* or *bunch*)”.

As noted above, the properties of *a bunch of*, and of related English binominal strings which provide the source of new quantifiers in English, have typically been explored using corpora of present-day English, with claims regarding diachronic development extrapolated from the synchronic variation. Brems (2003), based on corpus data from The Bank of English, suggests that the development of *a bunch of* can be divided into two principal phases. Like Langacker, Brems suggests a literal meaning in the pre-grammaticalization stage, in which collocates of *a bunch of* are restricted to a rather limited set of nouns, e.g., *flowers, grapes, bananas, carrots*, etc. The second stage is one of systematic grammaticalization, involving bleaching of *bunch* into a more or less purely grammatical quantity meaning. This process of bleaching opens up possibilities

for collocational expansion, and thus more types of nouns, e.g., *studies*, *practicing* and *guys* etc., can occur in the NOUN slot. But since a synchronic corpus is used as the data source, it is not clear when the collocational expansion in *a bunch of NOUN* began in the history of English. In her more recent work, Brems (2010) proposes that size noun constructions (in her terms) like *a bunch of* have three major uses: a head use, as in *a bunch of flowers*, a quantifier use, as in *a bunch of people*, and a valuing use, as in *a bunch of gobbledygook*, in which the referent is primarily evaluated rather than quantified. She also argues that the three uses have to be studied as “collocationally constrained constructions in that the semantico-syntactic parsing of each use corresponds with specific collocational patterns” (Brems 2010:83). While the first of the three uses (i.e., head use) involves collocations with concrete nouns, quantifier uses co-occur with concrete as well as abstract nouns (and thus with count and mass nouns), while valuing uses correspond to concrete animate and abstract nouns, which are typically negatively evaluated. Brems (2012) uses *heap(s)* and *lot(s)* as case studies to explore the establishment of quantifier constructions for size nouns. It is found that *heap* and *lot(s)* appear in an early partitive construction, in which they serve as head and have a collective meaning. *Heap(s)* and *lot(s)* appear as partitives in the Middle English period, with quantifier meanings having emerged by the Late Modern English period (see also Traugott 2008 for further treatment of the development of binominal quantifiers in English, from the perspective of diachronic construction grammar).

Francis & Yuasa (2008) takes *a bunch of* as an example to illustrate the diachronic development of English quantificational nouns. From the Middle English period to the Early Modern English period, the word *bunch* refers to the

‘bundle’ sense only. From the early seventeenth century onwards, generalization occurs, after which *bunch* may be used to refer to a group or collection of people or things; this gives way in the recent history of English to the purely quantificational meaning. Consequently, Francis & Yuasa hypothesize that *bunch* undergoes the following semantic extensions: bundle > collection > large quantity, based on a qualitative analysis of examples from the history of the language (Francis & Yuasa 2008:50). Crucially, Francis & Yuasa argue that there has been a semantic reanalysis but no syntactic reanalysis in the string *a bunch of NOUN* and similar expressions: “Q[uantifier] N[oun]s have retained the syntactic category and (internal) phrase structure properties of the source construction” (Francis & Yuasa 2008:55).

In summary, the above studies are illuminating to the extent that they reveal the semantic extension of *a bunch of* from different perspectives, most of which rely on quantitative synchronic corpus data, or qualitative diachronic analysis. We extend the discussion by providing a quantitative analysis of recent change in collocations. In particular, our study is bottom-up and data-driven. A data-driven approach allows the data to reveal points at which there are important changes in frequency and collocational patterns, thus allowing the data to determine the divide into historical periods (Fitzmaurice 2016). Our method allowed us to retrieve all the instances of the string *a bunch of NOUN* from COHA, and then identify the stages of the diachronic development based on a particular clustering method. We annotated sampled instances in terms of their semantic, pragmatic and discourse features. Based on the annotated data, we provide in sections 4 and 5 an analysis of the diachronic changes involved and calculate the relevant correlations, including some further discussion of the nature of related

morphosyntactic changes.

Data and methodology

In this section, we discuss how we extracted instances of *a bunch of NOUN* from our corpus and how we identified the periods of the data on the basis of their token frequency.

Data Collection

The present paper uses COHA as the source of data. COHA contains more than 400 million words in more than 100,000 texts which represent material from 20 sequential decades from the 1810s to the 2000s. It is balanced by genre and sub-genre in each decade, and it has been lemmatized and tagged for part-of-speech using the CLAWS tagger system. However, considering the corpus composition presented in Davies (2012), texts for the 1810s and the 1820s are scant, meaning that data extracted in these decades could be biased. From the 1830s onwards, each decade has more than thirteen million words and roughly the same number of words per decade. Consequently, this case study uses the data ranging from the 1830s to the 2000s. In all, 3883 instances of *a bunch of* were retrieved from COHA³, and Table 1 shows the raw token frequencies and normalized token frequency (per million words) of the construction.

[TABLE 1 HERE]

It can be observed from Table 1 that overall the token frequency of *a bunch of*

has been increasing over the past 180 years. For example, the frequency in the 2000s is four times higher than that in the 1830s. However, whether we are able to distinguish particular stages in the change by frequency is still a question. It should be noted that much previous work in diachronic corpus linguistics has been carried out on the basis of even-sized sequential periods, such as centuries or half-centuries. Considering the data of *a bunch of* from COHA, it is not difficult to find that even-sized periodization has certain deficiencies. On the one hand, describing the changes observed in the data decade by decade may fail to make broad generalizations about the overall trend of the change. Focusing on each individual decade, we take the risk of ‘not seeing the wood for the trees’. On the other hand, arbitrary periodization invites the problem of subjectivity. For instance, the whole process could be randomly separated into 2, 3, 6 or even 9 periods, which, nevertheless, will lead to different conclusions. If the data are divided into 6 even time-sized periods, the two decades with similar frequency, like the 1940s and the 1950s, will be separated and fall into different periods. In a nutshell, we need to have a more objective method to identify relevant periods for analysis on the basis of the data themselves, i.e., a data-driven period identification. It should be pointed out that here we take the overall token frequency of *a bunch of* as the starting point for analysis, without considering the possible different interpretations it may have. The main motivation for this is the capacity for language users to register the string frequency of a sequence such as *a bunch of NOUN* regardless of the varied interpretations it may have. We started from the position that, before we are able to determine the nature of particular changes, it is reasonable first to look at the change in token frequency: as mentioned by Hilpert (2013b:462), “a look at raw frequencies is a common starting point for analyses of constructional change”. In

the following part, we use this token frequency as the basis for data periodization.

Data Periodization

The data periodization method used in here is Variability-based Neighbor Clustering (VNC). VNC is a “data-driven bottom-up clustering method for the identification of stages in diachronic corpus data” (Gries & Hilpert 2008:59). This method establishes time periods where the frequency of a particular sequence in a corpus is similar, merging together adjacent periods only, i.e., it groups together time periods on the basis of similar frequency distributions, but only merges adjacent time periods (see further Gries & Hilpert 2008; Hilpert & Gries 2009; Hilpert 2013a).

The sequence of normalized frequencies, as shown in Table 1, is used as the input for VNC; Hilpert (2013a:35) explains that a standard deviation is calculated to measure the variability between neighbors. For every adjacent period, a standard deviation is calculated. The periods with the smallest standard deviations consist of a group, within which the normalized frequencies are closest. In the present study, the 1930s and the 1940s are the two neighboring periods with minimum standard deviation (0.04), so they are grouped in the first iteration, merging into a single data point. The result of subsequent iterations of the VNC algorithm for the data as a whole can be seen in Figure 1.

[FIGURE 1 HERE]

Three pieces of information are represented in Figure 1: a dendrogram which represents the various clusters generated by the algorithm, the general change of token frequency of *a bunch of*, and the mean frequency of each cluster marked by

gray horizontal lines. The dendrogram, however, does not determine the periodization of the development of *a bunch of NOUN*; while this approach to periodization may appear objective, some subjectivity remains, as the analyst ultimately decides how many periods are optimal as a basis for analysis. Hilpert (2013a:37) observes, “the analyst could, in principle, draw a horizontal lines across [the figure] at any height and take the crossing vertical lines as representatives of successive historical stages.” In the present study, we identified five clusters: period 1 from the 1830s to the 1900s, period 2 from the 1910s to the 1960s, period 3 the 1970s and 1980s, period 4 the 1990s and period 5 the 2000s. In the following paragraph, we clarify the grounds for our decision to work with a five period clustering.

The five-cluster solution we adopt can be quantitatively justified using a scree plot (see Hilpert 2013a:37 for further discussion of this method).

[FIGURE 2 HERE]

In order to create a workable periodization, it is helpful to try to establish a balance between explicitness (which typically involves a larger number of periods) and economy (which will typically reduce the total number of periods). The most explicit number of periods is 18: explicitness correlates with proximity to the x-axis, as noted by Hilpert (2013a:37), who writes “the closer the numbers get to the x-axis, the more information is accounted for.” By inspecting the scree plot depicted in Figure 2, we see there are significant differences between the mergers 1 to 5: this is where the curve is steepest. Accordingly, Figure 2 motivates the choice of a five-cluster solution. Thus the diachronic change in string frequency of *a bunch of* can be partitioned

into 5 periods, which is illustrated in Table 2.

[TABLE 2 HERE]

The string frequency in Table 2 displays a clear trend. In the first period from the 1830s to the 1900s, the normalized frequency is 4.61 per million words. The number doubles in the next period from the 1910s to the 1960s. The frequencies of the three following periods increase by 50 percent, 30 percent and 16 percent respectively. Consequently, via the data periodization, we get a clearer picture of the frequency change of *a bunch of NOUN*.

The next step is to explore whether there are semantic changes affecting the schematic part of the string, namely the NOUN slot. We retrieved 200 random samples from each period for analysis⁴. In all, 1000 nouns in the string *a bunch of NOUN* were semantically tagged. We annotated the data based on the three types of properties, or dimensions, associated with meaning: semantic, pragmatic and discourse-functional. Firstly, we attempted to characterize semantic groups for the elements in the NOUN slot, resulting in five categories: plants; (inanimate) objects; animals; people; and abstractions. ‘Plants’ includes nouns such as *violets*, *flowers* and *grapes*, ‘objects’ like *keys* and *candles*, ‘animals’ like *cattle* and *pigs*, ‘people’ like *kids* and *guys*. ‘Abstract concepts’ include nouns such like *lies* and *ideas*. For the semantic tagging, we initially used Wmatrix, a web-based corpus tool developed at Lancaster University for the automatic semantic analysis of texts. Wmatrix provides a web interface to the English USAS (UCREL Semantic Analysis System) corpus annotation tools. After the automatic tagging, we checked the results manually and merged minor subdivisions into the five major

categories listed above. Pragmatic properties refer to the evaluative connotation of the *NOUN*, or the semantic prosody of the construction as a whole⁵. The evaluative meaning is tagged as POS (positive), NEG (negative), and NEU (neutral). Discourse-functional properties here mainly refer to the genres of texts in which the construction appears. COHA has four genres, that is, fiction (FIC), popular magazines (MAG), newspaper texts (NEWS), and non-fiction books (NF). The genre information had already been tagged in the corpus. The tagging of the ‘meaning’ of the elements that filled the NOUN schema is illustrated in Table 3.

[TABLE 3 HERE]

Results

In this section, we first describe changes in each property of the meaning of the NOUNs in the string one by one, and then measure the associations between them by means of correspondence analyses. The change in distribution of semantic properties of the NOUNs in the construction is presented in Figure 3.

[FIGURE 3 HERE]

Figure 3 illustrates the diachronic change in the distribution of semantic properties of the nouns in the string. In the first period, 88 percent (i.e., an overwhelming majority) of collocates with *a bunch of* consists of nouns referring to ‘plants’ and ‘objects’. In the second period, *a bunch of* shifts to collocate more with nouns denoting

‘people’, which accounts for 46 percent of all the tokens, more than fifteen times that of the first period. It can be inferred that in the second period *a bunch of* intensifies its meaning shift from literally ‘bundle’ to ‘group’. A plant or an object can either grow in clusters or can be tied together as a ‘bundle’, as in *a bunch of grapes/flowers/keys*, but this is not true of *a bunch of people/kids/Indians*, who can only be ‘grouped’ together. What makes the third period distinct from the earlier two is that since the 1970s more nouns expressing abstract concepts have entered the slot of the string, as in *a bunch of baloney/noise/times*. *A bunch of* used in this context cannot be interpreted as ‘bundle’ or ‘group’ any more. Rather, similar to *many* and *much*, *a bunch of* here refers to a large quantity. That is to say, in the third period, the quantifying function of *a bunch of* comes to be further entrenched, having a more bleached meaning than in the earlier periods. This meaning is strengthened in the 1990s and the 2000s. It can also be observed that the distributions of the five semantic categories are quite similar between the later three periods and the meanings of ‘a group of’ and ‘large quantity’ account for more than 75 percent, which means that the bleached meaning begins to be predominant in the 1970s-1980s and it has remained stable since then. The only difference between the final three periods lies in their token frequencies, as shown in Table 2. One interesting phenomenon that should be pointed out here is that in the last two periods, some nouns literally referring to animals actually denote people, representing the conceptual metaphor MEN ARE ANIMALS, as shown in (2) and (3):

(2) The undergraduate males are *a bunch of donkeys*, mostly, and you're going to look sleek and worldly to the coeds, and they're going to look pretty good to you. Some of them will sit in the front row in short skirts with paradise twinkling at you. (1993, Fiction [COHA])

(3) Of course I'm okay, you cartoon cretin. (Taking in the gawking crowd.)

Jesus, such *a bunch of mortal maggots*. (To Kate Spoon.) Get out of my way. Move, you sow! Move! (1994, Fiction [COHA]).⁶

The two curves in Figure 4 present the diachronic variation in pragmatic properties, which concerns the evaluative meanings as stated above. In the first period, all 200 nouns have neutral attitudinal meaning only. In the second period, the negative nouns account for nearly 40 percent, and during the last three periods the negative nouns have remained stable, accounting for 45 to 55 percent of the total number. It should also be noted that in the final period only five instances of positive evaluation were detected. However, that proportion is quite low compared with that of neutral and negative ones. So overall, the second period also demonstrates a shift regarding the pragmatic properties of the nouns in the string, with a change from exclusively neutral nouns in the slot in the first period to approximately half neutral and half negative in the second.

[FIGURE 4 HERE]

For the discourse properties, the column chart in Figure 4 shows that more than 80 percent of the instances of the string are located in the ‘fiction’ genre in the first period. In the following periods, ‘fiction’ accounts for 76 percent of the total in the second period, 71 percent in the third, 68 percent in the fourth and 62 percent in the final. It is evident that the proportion of ‘fiction’ declines and is complemented by genres of ‘magazine’ and ‘news’. The high proportion of ‘fiction’, however, cannot be taken as a definitive indication of the predominant occurrence of *a bunch of* in ‘fiction’; rather, it may largely be due to the corpus composition itself: about 50 percent of the texts in COHA are

fictional texts. So the change in relative proportions is what concerns us here. On the whole, it can be concluded that *a bunch of NOUN* increasingly widens in terms of its occurrence in different text types, appearing in diversified genres.

As indicated above, such diachronic data can be used to trace semantic change; however, the complexity of the relationship of the three properties has been insufficiently unraveled. For example, could a change in semantic properties be accompanied by a change in pragmatic properties? In order to determine possible associations between the different kinds of ‘meaning’, we measured the association between the semantic, pragmatic and discourse properties as well as between the variables in each property. Using SPSS 25.0, we carried out three pair-wise analyses of the association between meaning categories and applied multiple correspondence analysis (MCA) to visualize the relationship between variables in these different categories. MCA aims at reflecting aspects of co-occurrence of different linguistic elements. The merits of MCA lie in its capacity to deal with multiple variables, which is conducive to the multidimensional nature of constructional change (Hilpert 2013a, b; Traugott & Trousdale 2013). Table 4 presents the correlations between semantic and pragmatic properties, which is represented graphically in Figure 5. Since the frequency of each cell in the ‘POS’ column is below 5, ‘POS’ is combined with ‘NEU’ and labeled as ‘NEU & POS’ in order to ensure the validity of chi-square test.⁷

[TABLE 4 HERE]

The data in Table 4 shows that there are certain statistically significant associations between semantic properties and pragmatic properties ($\chi^2 = 300.490$,

$p < 0.0005$). Further information regarding these associations can be gleaned from a multiple correspondence analysis, as in Figure 5.

[FIGURE 5 HERE]

In a correspondence analysis, the output is a ‘map’ in which differing degrees of relatedness of linguistic elements is shown by variable distance. The closer the variables are distributed in the map, the more correlated they are. In Figure 4, dimension 1 and dimension 2 respectively capture 77.4 percent and 50 percent of the variance. It is demonstrated that ‘animals’, ‘objects’ and ‘plants’ in the construction are closely associated with ‘neutral’ evaluation, which indicates that language users seldom attach any evaluative meaning to *a bunch of NOUN* constructs where the nouns of those categories are used. Additionally, the figure shows that ‘people’ is closely related with ‘negative’ attitude, which means nouns denoting ‘people’ usually have a certain derogatory connotation when they appear in the string *a bunch of NOUN*. As for ‘abstractions’, corpus data⁸ shows that ‘abstractions’ is relatively more associated with nouns with negative meaning, though in Figure 5, it is isolated from all variables in the pragmatic dimension.

Similar associations may be observed in the relationship between semantic property and discourse property ($\chi^2 = 69.486$, $p < 0.001$). A multiple correspondence analysis is displayed in Figure 6.

[FIGURE 6 HERE]

In Figure 6, dimension 1 and dimension 2, respectively, capture 61 percent and

25 percent of the variance. As Figure 6 indicates, the fiction genre is located at the center of the graph, which is closely related to all five semantic variables, particularly with ‘objects’, ‘animals’ and ‘abstractions’. The other three discourse variables are located in the corners of the map, which indicates they may be positively related to one or two semantic variables only. For example, the magazine genre is positively related to ‘abstractions’, just as non-fiction is with ‘plants’, and the newspaper genre is with ‘people’ (albeit more loosely). Note that the centeredness of the fiction genre may be partially due to the uneven distribution of genres in the corpus itself, as mentioned earlier.

Finally, there is also a marginally significant correlation between the pragmatic property and discourse property ($\chi^2 = 7.71$, $p = 0.05$). The multiple correspondence analysis in Figure 7 illustrates particular associations.

[FIGURE 7 HERE]

In Figure 7, dimension 1 and dimension 2 respectively account for 54.4 percent and 50 percent of the variance. Fiction and non-fiction genres are closely associated with neutral or positive evaluation. Magazine and newspapers are relatively more associated with negative evaluation.

From these data, it can be observed that there are interesting relationships between variables in the semantic, pragmatic and discourse properties of the *NOUNs* in a bunch of *NOUN* during its development in the Late Modern English period. In order to observe further dimensions of change, temporal variables, namely, the five time periods, are introduced in the following analysis. The MCA result is displayed in Figure 8 (see further Glynn 2014 for a discussion of factor

maps):

[FIGURE 8 HERE]

Here, dimension 1 and dimension 2 respectively account for 53.1 percent and 29.5 percent of the variance. If we look at each period in detail, a clearer picture emerges. In the first period (1830s-1900s), the NOUN of the string is largely one which refers to ‘plants’ with neutral evaluation, and is more inclined to appear in such genres as non-fiction and fiction. It can be inferred that in this period *a bunch of* is typically used with its literal meaning ‘bundle’. During the second period (1910s-1960s), *a bunch of* tends to collocate with nouns denoting ‘animals’ and ‘objects’ with neutral attitudinal meaning, as well as ‘people’ and ‘abstraction’ with negative evaluation, more often appearing in fiction and magazines. It can be inferred that *a bunch of* in this period has multiple meanings while ‘group’ meaning is more typical. Compared with the second period, the third period (1970s-1980s) stands out, as it is more positively correlated with nouns denoting the meaning of ‘people’ and ‘abstractions’ with negative evaluation, appearing more in newspapers. Along with this change comes an inference that ‘group’ meaning in this period has more or less extended to quantifying meaning. In the 1990s and the 2000s *a bunch of* collocates with abstract nouns besides ‘people’, which means that during the last two stages, the ‘group’ meaning and ‘large quantity’ meaning of *a bunch of* have remained stable. At the same time, *a bunch of* tends to be associated with negative evaluation with a more frequent occurrence in fiction and magazines.

In sum, *a bunch of NOUN* changes its reference mainly from plants to people and to abstract things during this diachronic development, causing *a bunch of* to

extend its meaning from ‘bundle’ to ‘group’ and finally to a more bleached one of ‘large quantity’. Its attitudinal meaning changes from neutral to negative, particularly when referring to people and abstract things. There are marginal examples of positive nouns in the string, but their frequency is low. The genre association has extended from fiction to magazine and newspapers.

Implications for diachronic construction grammar

Thus far we have been concerned only with the nature of changes in the meaning correlates of string frequency. But as our discussion of the previous literature in section 2 above has shown, it is possible that we are dealing with two different constructions here: one is an instance of a partitive schema (with *bunch* as the semantic and grammatical head, and *of NOUN* a prepositional postmodifier), and the other is an instance of the quantifier schema (with *a bunch of* as a complex quantifier, and *NOUN* as its complement). How are we to relate these quantitative findings to meaning changes to the question of syntactic change affecting the noun *bunch*, and what does this tell us about the nature of constructional change? In the field of diachronic construction grammar, we observe two related but distinct points of view regarding the nature of constructional change. One is put forward by Hilpert (2013a:16):

“Constructional change selectively seizes a conventionalized form-meaning pair of a language, altering it in terms of its form, its function, any aspect of its frequency, its distribution in the linguistic community, or any combination of these.”

Another is proposed by Traugott & Trousdale (2013:1) who distinguish two types of change affecting constructions:

“(a) Changes that affect features of an existing construction, e.g., semantics, morphophonology, collocational constraints, etc. These changes do not necessarily lead to a new construction. We call them ‘constructional changes’.

(b) The creation of a form_{new}-meaning_{new} pairing. We call this type of change ‘constructionalization’.”

It is apparent that Hilpert’s definition of ‘constructional change’ as an umbrella term is broader in scope and does not distinguish between the coming into being of a new construction, and changes affecting an existing construction. While we hold that it is helpful to make the distinction between constructional change (as a change along one dimension) and constructionalization (as the creation of a new construction), our focus here is on the nature of form-meaning changes in established constructions, so we do not focus on constructionalization *per se*. Since a construction is defined as a ‘pairing of form and meaning’, it is natural for a construction to change in one dimension or in more dimensions.

Since we have provided a detailed analysis of the meaning change of *a bunch of NOUN*, let us turn now to matters of form, which includes the syntactic, morphological and phonological properties. In order to establish changes in form – and therefore to make connections between form and meaning changes, and the potential differentiation between the partitive construction and the quantifier construction – we must be selective about the examples we choose. Take for instance the clause *I saw a bunch of balloons in the sky*. There is no independent way here of assigning the string *a bunch of balloons* to either the partitive or to the quantifier construction: the balloons may be tied together, or simply be large in number. While this may be disambiguated in context, there is nothing in the form of the clause that leads us to favor an analysis of the construct as belonging to one

constructional type over another. However, there are certain contexts which may provide some clues as to whether the formal distinction between partitive and quantifier is becoming more clearly established.

To this end, we looked at two morphosyntactic properties: subject-verb concord and agreement between the noun and a coreferential pronoun, as illustrated in (4) and (5):

(4) Nonetheless, they did not propose to leave merely because ***a bunch of excitable foreigners were*** taking on. (1967, Magazine [COHA])

(5) Keyes has ***a bunch of questions*** Nixon wants to answer. He's written ***them*** in advance to make sure they're properly worded. (1969, Magazine[COHA])

While we recognize the issues regarding mismatch vs. alignment of semantic and syntactic heads in constructions of this kind, as raised by Francis & Yuasa (2008) and Börjars *et al* (2015), we nevertheless believe that the quantitative findings regarding concord are relevant to our argument regarding post-constructionalization changes. We focus on verbal concord with *a bunch of NOUN* where *bunch* and the complement of the preposition *of* do not agree in number, so the complement has to be a plural noun. If the verb agrees with *bunch*, we suggest that this is an instance of the partitive construction, where the head is followed by a postmodifying prepositional phrase; in contrast, if the verb agrees with the second noun, *bunch* can be characterized as part of a complex quantifier (see also Smitterberg 2006:260). In example (4) above, the plural verb *were* agrees with the noun *foreigners*, which serves as the control of concord, with *a bunch of* as the quantifier. Similarly, in example (5), *them* is used to refer to *questions*, rather than *bunch*, so *questions* serves as the control of number agreement.

Since just a small subset of instances of the string *a bunch of NOUN* provide the

right context for the analysis described above, we checked all 3883 tokens and found 514 instances that were amenable to analysis. Variation over time in the patterns of the two types of agreement in the five periods is listed in Table 5.

[TABLE 5 HERE]

As can be seen from Table 5, singular verb concord declines from 60 percent in the first period to 0 percent in the last period. That means an increasing number of verbs appear to be agreeing with the second noun in the string, rather than with *bunch*. To put it in another way, more nouns serve as the control of concord. It should be noted that our finding here is in line with the general tendency for most partitive constructions in the nineteenth century, as observed by Smitterberg (2006). Where the partitive noun was singular and the prepositional complement plural, the results indicate a tendency towards more verbal concord with the prepositional complement in the later period of the nineteenth century than the early period. In other words, in the nineteenth century there is a general tendency for some partitive nouns to go from being the head of the partitive construction to being part of a complex determiner of the new head, i.e., the prepositional complement (Smitterberg 2006), suggesting a realignment to the quantifier construction. The change in noun-pronoun agreement is even more radical. From the 1910s onwards, there are no instances of *bunch* controlling agreement on the pronoun. This quantitative analysis of the formal properties of the string lends weight to the claim that the semantic changes discussed in section 4 go alongside the greater entrenchment of *a bunch of* as a complex quantifier. What we see in the Late Modern English period, then, is the working out of an earlier

constructionalization (the creation of the complex quantifier *a bunch of*), in which the post-constructionalization constructional changes embed the new pattern more firmly in the system of English quantifiers. We illustrate the relationship between constructionalization and constructional changes now.

The phenomena we have observed demonstrate the gradual nature of post-constructionalization constructional changes: the 18 decades we are focusing on here is just a short episode of the whole diachronic change affecting the construction. The *Oxford English Dictionary* (OED) dates the sense of *bunch* as ‘a collection; a company or group of persons’ as far back as the 1620s. Consequently, it is reasonable to assume that the constructionalization took place prior to the beginning of the nineteenth century (i.e., before the period covered by COHA). Examples such as (6) from the Proceedings of the Old Bailey (POB) suggest that *a bunch of NOUN* was neoanalysed as [[a bunch of] NOUN] with its meaning of ‘collection’ or ‘large quantity’ (note the plural pronoun *them*):

(6) I heard a dispute between my maid-servant and the prisoners; the maid said Dowle had stole *a bunch of garnets*, and put them in her bosom; I found them on the counter. (POB; Trial of Mary Dowle and Anne Hinckley, 22 October 1766)

In terms of meaning alone, the ‘large quantity’ sense of *a bunch of* may have been available in the seventeenth century, as in (7) and (8):

(7) As for dropsies, he that said desires were dropsies, bid the mind take *a bunch of reason*, that grows in a well-temper'd brain. (1671 Margaret Cavendish. *Natures picture drawn by fancies pencil to the life* [EEBO])

(8) here *a bunch of pride* is growing, straighten and check there saies god. (1674 Thomas Hardcastle. *Christian geography and arithmetic* [EEBO])

A bunch of in (7) and (8) might be interpreted as ‘large quantity’, and as mentioned above, this sense came to be particularly associated with the *a bunch of NOUN* string in the 1970s, while its original ‘bundle’ sense persists to the present-day. However, it should be observed that both of these examples contain a form of the verb *grow*, which suggests that Cavendish and Hardcastle may also have been drawing on the earlier sense of ‘bundle’ associated with vegetation.⁹

Our last brief observation regarding post-constructionalization changes concerns the orthographic variant *buncha*, which represents a fusion of *bunch* and *of*. We found 46 instances of *buncha* in COHA, with its first occurrence in 1940s, as in (9):

(9) “Are you starting? Are you gonna start that, too?” She took her apron off. “I’m goin’ home, if you start that stuff. ***Buncha*** comedians.” (1949, Fiction [COHA])

All 46 instances of *buncha* have the meaning either ‘a group of’ (as in *a buncha kids/people*) or ‘large quantity’ (as in *a buncha games/crazy stories*), and it is more frequently used in a derogatory sense, as in *a buncha half-breed idiots/bastards/damn thieves*.

To summarize our account, we recognize that in the Early Modern English period, *bunch* in *a bunch of grapes* or *a bunch of flowers* has the lexically specific and collocationally restricted meaning of a constellation of entities growing or fastened together at one end. At that time, *bunch* is the head of a partitive construction, i.e., one which individuates a part as distinct from a whole. Similar constructions are *a slice of bread*, *a pinch of salt* and *a clod of mud*. However, in the passage of time, the string *a bunch of NOUN* begins to have the meaning of ‘group of X’ and ‘large quantity of X’ where X is the meaning conventionally associated

with the second noun in the string. Instances of use with one of these two meanings become increasingly frequent as our corpus data show.

Our study has articulated the nature of post-constructionalization changes in COHA by considering the string frequency of *a bunch of NOUN*, and the changes in concord marking with verbs and pronouns. We have shown that correlations between the semantic, pragmatic and discourse properties of the noun reveal how the quantifier construction becomes more strongly entrenched; this is also reflected in the morphosyntactic changes across the five periods we identified through a VNC analysis of the (normalized) string frequency. Yet we have also illustrated how the original partitive sense persists, particularly in certain collocations. The 18 decades we have focused on here is a period of co-existence where *bunch* forms part of two distinct constructions, one an instance of the more general partitive construction, and the other an instance of the more general quantifier construction, with the former becoming more marginal and reduced in its collocational range and the latter increasingly frequent and prominent, having undergone a series of expansions. We propose that the constructionalization from a partitive construction to a quantifier construction already took place before the 1830s, thus our focus here has been on the quantification of post-constructionalization constructional changes.

This case study has attempted to explore constructional change through frequency measurements of variant forms, which are provided by corpora. As stated by Wolk *et al.* (2013:383), “we premise that grammatical knowledge must have a probabilistic component, as the likelihood of finding a particular linguistic variant in a particular context in a corpus can be shown to correspond to the intuitions that speakers have about the acceptability of that particular variant, given the same context.” In other words, constructs found in corpus data reflect the

probabilistic linguistic knowledge that speakers have.

Conclusion

In this article, we have attempted to give a clearer account of the nature of post-constructionalization constructional changes, which typically involve expansion of collocations, and may also involve morphological and phonological reduction (Traugott & Trousdale 2013:27), in particular by providing a quantitative study of these phenomena, in contrast to the qualitative work that Traugott & Trousdale (2013) undertook.

Using the construction *a bunch of NOUN* as a case study, this paper presented a multivariate analysis of diachronic variation in English partitive and quantifier constructions on the basis of corpus data, and explicated the principles of constructional change. Specifically, our results were:

(1) the semantic expansion of *a bunch of* was particularly marked in the period from the 1910s to the 1960s when the original dominant meaning ‘bundle’ was overtaken by ‘group’, with the meaning ‘large quantity’ becoming increasingly frequent from the 1970s onwards. The three meanings co-exist even in the 2000s;

(2) there are statistically significant correlations between some of the meaning dimensions, i.e., the semantic, pragmatic and discourse properties of the noun in the string *a bunch of NOUN*. For example, when the noun refers to people, the string usually has a negative semantic prosody;

(3) changes affecting the string *a bunch of NOUN* in the Late Modern English period, i.e., from 1830 onwards, are best seen as part of a process of post-constructionalization constructional changes. Such changes are relevant not only to

the ‘new’ construction, but also to the ‘source’ construction. These changes are most clearly revealed through a quantitative analysis of both morphosyntactic and meaning (semantic, pragmatic and discourse) properties.

In sum, this article has applied a quantitative data-driven method to analyze the post-constructionalization changes to both the source construction and the new construction that becomes entrenched. This suggests that a quantitative approach provides insight into the nature of semantic-pragmatic expansion (Himmelmann 2004), entrenchment of schemas, and prototype formation, which Traugott & Trousdale (2013:238) noted as a potential contribution to future work in diachronic construction grammar.

Notes

¹ We are grateful to Martin Hilpert, Elizabeth Closs Traugott and audiences at the Universities of Edinburgh and Zürich for their constructive feedback on earlier versions of the research presented in this article. We would also like to thank the journal's editors, and reviewers of the first version of the article, for their helpful comments and suggestions for improvements. This work is supported by the National Social Science Foundation of China (Grant No. 14BYY001). All shortcomings are our own.

² Lehrer (1986) is prudent considering the word class of *bunch*. She admits that “English does not have a closed set of classifiers” (1986: 110) but states “consider the collective classifiers *bunch* and *cluster*” (1986: 118). So for her, ‘classifier’ may be just a convenient generalizing term to indicate the nouns functionally similar to classifiers. Allan (1977), despite admitting that English is less properly called a ‘classifier language’, claims that “perhaps all languages have classifiers” (1977: 286) and *bunch* is put in the category of classifier.

³ ‘A bunch of’ is used as the search phrase in corpus to make sure that expressions like ‘the/this/one bunch of’ will be excluded. Consequently, the results are ‘a bunch of (modifier) NOUN’. It should be noted that a modifier may be inserted between *of* and *NOUN* though the expression *a bunch of NOUN* is used for the sake of convenience.

⁴ The software ConcSampler (2010) developed by Beijing Foreign Studies University is used here for random sampling of concordances.

⁵ Semantic prosody is defined as “consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993:157). It reveals an evaluative potential of the

extended unit of meaning, chiefly in terms of a positive or negative evaluation. As is pointed out by Sinclair (1996), “a semantic prosody is attitudinal, and on the pragmatic side of the semantics/pragmatics continuum”. In our study, a pragmatic property includes either the connotative meaning implied in the NOUN itself as or the attitudinal meaning suggested by the speaker in context. For instance, (1) *The Moscow radio lambasted the new management as a bunch of rascals*. In sentence (1), *rascals* is tagged as NEGATIVE due to the negative connotation of *rascal* itself, along with the semantic associations of the verb *lambast*. (2) *How can you respect the world when you see it’s being run by a bunch of kids turned old?* In sentence (2), *kids* is annotated as NEGATIVE owing to the negative attitudes implied by the context in the sentence (the implication of absence of respect, along with the implication that adult leaders have the mindset of a child), though the word *kids* per se does not denote negative meaning.

⁶ For those cases where an animal noun is used negatively to refer to people, we annotate the noun as ‘animals’ in the semantic dimension, and ‘NEGATIVE’ in the pragmatic dimension.

⁷ The frequencies of ‘POS’ tagging for the plants, objects, animals, people, and abstractions are 0, 0, 0, 4, and 1 respectively. We are grateful to an anonymous reviewer for the suggestion of collapsing the two categories for purposes of statistical analysis.

⁸ For all the 75 instances of nouns denoting the meaning of ‘abstractions’ in the string of *a bunch of NOUN*, 44 instances are colored with negative evaluation while 30 with neutral meaning and only one instance of positive meaning.

⁹ We are grateful to an anonymous reviewer for drawing our attention to the use

of *grow* in these examples.

References

- Allan, Keith. 1977. Classifiers. *Language* 53(2). 285-311.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Brems, Lieselotte. 2003. Measure noun constructions: an instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics* 8(2). 283-312.
- Brems, Lieselotte. 2010. Size noun constructions as collocationally constrained constructions: lexical and grammaticalized uses. *English Language and Linguistics* 14(1). 83–109.
- Brems, Lieselotte. 2011. *Layering of size and type noun constructions in English*. Berlin: De Gruyter Mouton.
- Brems, Lieselotte. 2012. The establishment of quantifier constructions for size nouns A diachronic case study of heap(s) and lot(s). *Journal of Historical Pragmatics* 13(2). 202– 231.
- Börjars, Kersti, Nigel Vincent & George Walkden. 2015. On constructing a theory of grammatical change. *Transactions of the Philological Society* 113(3). 363–382.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Davies, Mark. 2010. The Corpus of Historical American English (COHA): 400 million words, 1810-2009. Available online at <http://corpus.byu.edu/coha/>.
- Davies, Mark. 2012. Expanding horizons in historical linguistics with the 400-

- million word Corpus of Historical American English. *Corpora* 7(2). 121–157.
- Fitzmaurice, Susan M. 2016. Semantic and pragmatic change. In Merja Kytö and Päivi Pahta (eds.). *The Cambridge handbook of English historical linguistics*, 256-70. Cambridge: Cambridge University Press.
- Francis, Elaine J. & Etsuyo Yuasa. 2008. A multi-modular approach to gradual change in grammaticalization. *Journal of Linguistics* 44(1). 45-86.
- Glynn, Dylan & Justyna A. Robinson. 2014. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.). *The Oxford handbook of Construction Grammar*, 15-34. New York: Oxford University Press.
- Gries, Stefan Th. & Martin Hilpert. 2008. The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora* 3(1). 59-81.
- Hilpert, Martin. 2013a. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2013b. Corpus-based Approaches to Constructional Change. In Thomas Hoffmann & Graeme Trousdale (eds.). *The Oxford handbook of Construction Grammar*, 458-475. New York: Oxford University Press.

- Hilpert, Martin & Stefan Th. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4). 385-401.
- Himmelman, Nikolaus P. 2004 Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelman, & Björn Wiemer (eds.). *What makes grammaticalization? A look from its fringes and its components*, 21-42. Berlin: Mouton de Gruyter.
- Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford handbook of Construction Grammar*. New York: Oxford University Press.
- Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, vol. 2: Descriptive application*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2008. *Cognitive Grammar a basic introduction*. Oxford: Oxford University Press.
- Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.). *Text and technology: In honour of John Sinclair*, 157-75. Amsterdam: John Benjamins.
- Love, Robbie. 2016. What's wrong with "a bunch of migrants"? Looking at the linguistic evidence. <http://cass.lancs.ac.uk/?p=1880>. [Accessed on January 28, 2016].

- Lehrer, Adrienne. 1986. English classifier constructions. *Lingua* 68. 109-148.
- Mvududu, Nyaradzo H. & Christopher A. Sink. 2013. Factor Analysis in Counseling Research and Practice. *Counseling Outcome Research and Evaluation*. 4(2). 75-98.
- Quirk, Randolph, Sidney Greenbaun, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.
- Sinclair, John. 1996. The search for units of meaning. *Textus* 9. 75–106.
- Smitterberg, Erik. 2006. Partitive constructions in nineteenth-century English. In Merja Kytö, Mats Rydén & Erik Smitterberg (eds.). *Nineteenth-century English: Stability and change*, 242-71. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs. 2008. The grammaticalization of NP *of* NP patterns. In Alexander Bergs and Gabriele Diewald (eds.). *Constructions and language change*, 23-45. Berlin: Mouton de Gruyter, 23-45.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in late modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419.
- Xiao, Richard & Tont McEnery. 2010. *Corpus-based contrastive studies of English and Chinese*. New York: Routledge.
- Zhang, Xu. 2017. *English quasi-numeral classifiers: A corpus-based cognitive-typological study*. Bern: Peter Lang.

TABLE 1Token Frequency of *a bunch of* in COHA

Year	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s
Frequencies	63	56	61	69	81	96	128	128	191
Frequency (n/MW)	4.57	3.49	3.70	4.05	4.36	4.73	6.21	5.79	8.41
Year	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
Frequencies	238	234	230	239	238	345	345	512	629
Frequency (n/MW)	9.28	9.51	9.45	9.74	9.93	14.49	13.63	18.32	21.27

TABLE 2

The Frequency of *a bunch of* in Each Period

Period	1830s-1900s	1910s-1960s	1970s-1980s	1990s	2000s
Frequencies	682	1370	690	512	629
Frequency (n/MW)	4.61	9.39	14.06	18.32	21.27

TABLE 3

Meaning Tagging of Nouns in *a bunch of NOUN*

Properties	Variables
Semantic Properties (SemP)	plants/objects/animals/people/abstractions
Pragmatic Properties (PragP)	POS/NEU/NEG
Discourse Properties (DisP)	MAG/FIC/NEWS/NF

TABLE 4

Association between Semantic and Pragmatic Properties

SemP	PragP		Chi-Square	P value
	NEU & POS	NEG		
plants	194	2	300.490	P=3.99E-10
objects	185	23		
animals	51	25		
people	171	274		
abstractions	31	44		

TABLE 5

The Change of Two Types of Agreement

Periods	Singular V(% <i>, n/N</i>)	Singular P(% <i>, n/N</i>)
1830s-1900s	60%(9/15)	14%(7/50)
1910s-1960s	23%(18/87)	0%(0/125)
1970s-1980s	11%(5/45)	0%(0/41)
1990s	5%(2/41)	0%(0/44)
2000s	0%(0/37)	0%(0/29)

Figure 1: Periods Identified through VNC

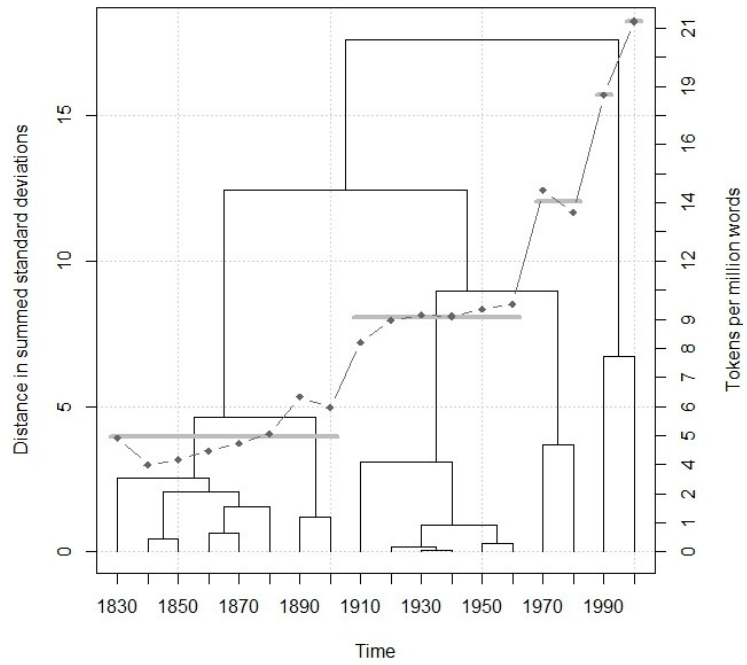


Figure 2: Scree Plot of VNC Results for *a bunch of NOUN*

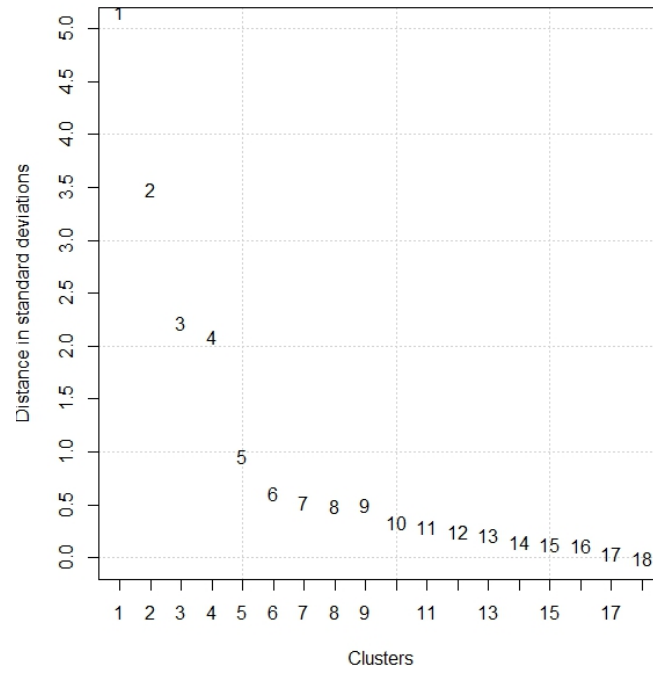


Figure 3: The Semantic Properties of the Noun in *a bunch of NOUN*

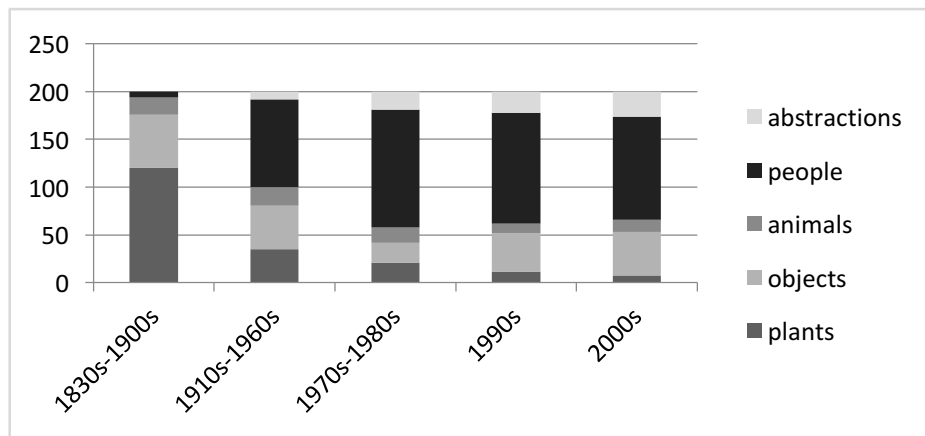


Figure 4: The Pragmatic and Discourse Properties of the Noun in *a bunch of NOUN*

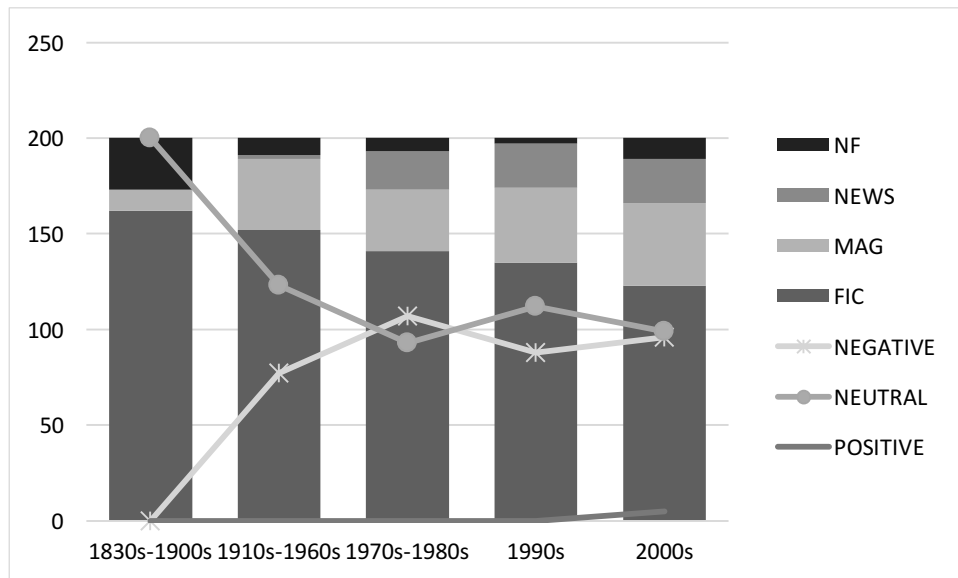


Figure 5: Multiple Correspondence Analysis (SemP and PragP)

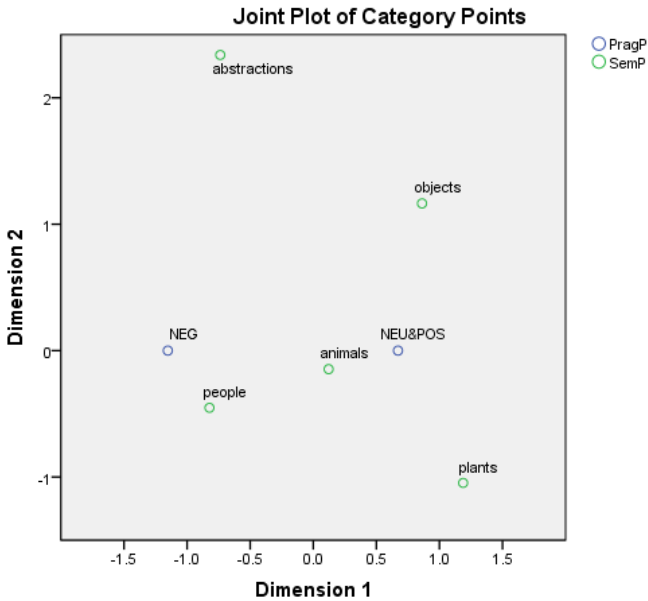


Figure 6: Multiple Correspondence Analysis (SemP and DisP)

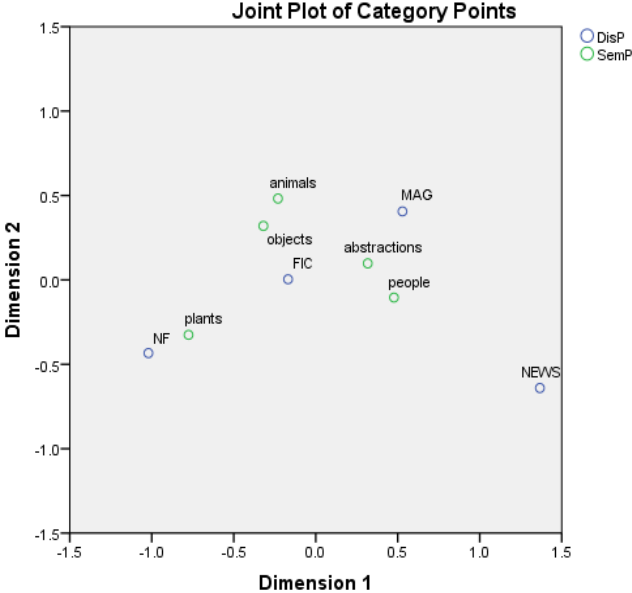


Figure 7: Multiple Correspondence Analysis (DisP and PragP)

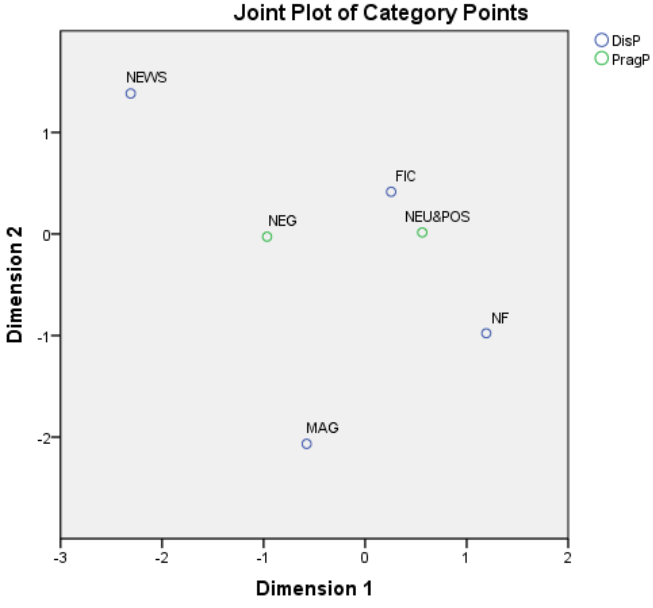
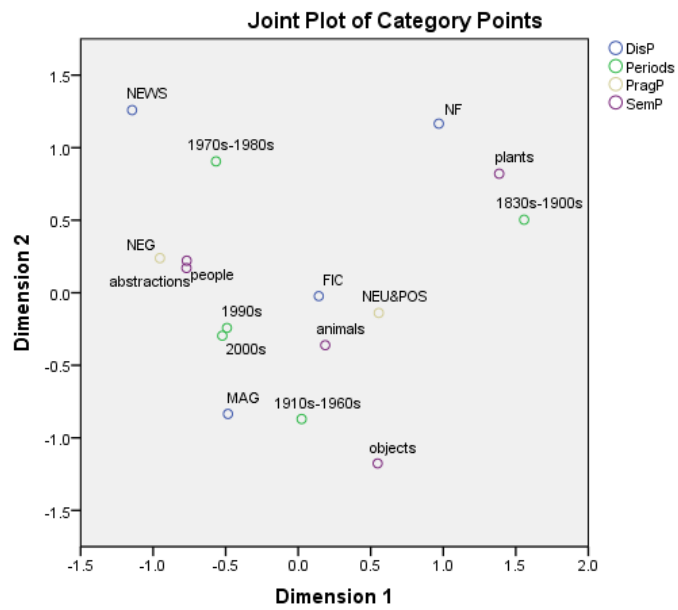


Figure 8: Multivariate Analysis of Variables of Nouns in *a bunch of NOUN*



Biographical statements

Bin Shao is a professor of linguistics in the School of Foreign Languages at Zhejiang University of Finance and Economics. His research interests involve the fields of lexical semantics and corpus linguistics.

Yingying Cai is a postgraduate student in the school of Foreign languages at Zhejiang University of Finance and Economics. Her research interests are corpus linguistics and historical linguistics.

Graeme Trousdale is a senior lecturer in the department of Linguistics and English Language at the University of Edinburgh. His research interests lie at the intersection of cognitive linguistics and historical linguistics.