Hans-Jörg Schmid and Annette Mantlik

Entrenchment in Historical Corpora? Reconstructing Dead Authors' Minds from their Usage Profiles

Abstract: Data from eight historical corpora spanning the period between 1250 and 1871 are investigated with regard to occurrences of the 'N+BE+*that*-construction' (as in *my concern is that* [...], *the idea was that* [...]). The formal, semantic, and pragmatic changes of this construction are described on the basis of 1,588 attestations retrieved from the corpora. Following this, the usage profiles of individual authors are examined. It is shown that even authors who are comparable in terms of period and genre show significant differences with regard to the frequency of use of the construction, collocational ranges and preferences, the use of semi-fixed lexical expressions manifesting the construction, as well as their functional preferences. These differences are interpreted from the perspective of the so-called 'Entrenchment-and-Conventionalization Model' (Schmid 2014a and 2015). It is argued that the usage profiles of individual authors can provide insights into the ways in which the construction under investigation was represented in these authors' minds, and that the observable collective long-term changes arise from the interaction of the cognitive processes in individual minds and the social processes taking place in the speech community.

Hans-Jörg Schmid, University of Munich
E-Mail: hans-joerg.schmid@anglistik.uni-muenchen.de
Annette Mantlik, Independent Scholar
E-Mail: annette-mantlik@gmx.de

1 Introduction

Language change is commonly conceived of as variation of the conventions of a language in time. It is generally considered to be a collective process taking place in speech communities and, as many linguists would insist on, a process that affects the structures and systems that make up language. While there is also considerable agreement that collective and systematic change is ultimately triggered and brought about by the linguistic activities of individual speakers, especially their innovations, this activity has largely remained outside the focus of the study of language change proper (Janda and Joseph 2003: 19). It is the study

of generalizable structures, relations, and systems that is regarded as the main task of linguists rather than the description of the more or less idiosyncratic utterances or routines of individuals.

Of course certain social groups have been identified as being more or less likely to instigate language change. Likewise the individual styles of key literary figures have been analyzed and described (see, e.g., Craig 1992 and 2002; Taavitsainen 1995; Cannon 2004), and the impact of particularly prominent writers, such as Chaucer or Shakespeare, has been investigated with regard to the development of the lexicon (see, e.g., Horobin 2007; Craig 2011). But on the whole the work of individual authors has not been deemed worthy of serious, i.e. systematic, linguistic research. This is also confirmed in a study by Raumolin-Brunberg and Nurmi (2011), which is explicitly devoted to the role of the individual in language change:[1] "The behaviour of the individual language user has not been among the key issues in the rapidly growing literature on grammaticalization. Nor has it been a popular topic in the general studies of language change" (Raumolin-Brunberg and Nurmi 2011: 251). The authors themselves investigate the use of the auxiliaries *do, will*, and *would* in the works of Sir Walter Raleigh, Philip Gawdy, and John Chamberlain and find that there is "a great deal of variation between individuals concerning their participation in ongoing linguistic changes" (Raumolin-Brunberg and Nurmi 2011: 262).

Given the scarcity of precursors, the present study can be considered programmatic in nature. Following up on the lead by Raumolin-Brunberg and Nurmi (2011), we introduce a fresh framework for the interpretation of individual differences between the usage profiles of individual historical authors and for comparing them to each other and to the collective usage tendencies typical of their respective historical periods and genres. The findings will be interpreted in terms of types and degrees of entrenchment. The following argumentative steps motivate the rationale behind this study:

1. Collective processes on the macro-level of speech communities (conventionalization) have to be separated from cognitive processes on the micro-level of individual authors[2] (entrenchment).

1 Individual authors are of course mentioned and quoted in many publications on language change and grammaticalization. Krug (2000), for example, explicitly includes the works of Daniel Defoe, Jane Austen, Charles Dickens, and Oscar Wilde and draws attention to special characteristics of their use of modal verbs. Overall, however, these references still are of an exemplary nature rather than providing systematic descriptions.

2 As this paper deals with texts written by historical authors, we will mainly use the term 'author' to refer to language producers. It is understood throughout that the theoretical background applies to both written and spoken forms of language production.

2. The linguistic activities of individual authors contribute to shaping the collective conventions and the way they develop.
3. However, authors differ with regard to the extent to which and the ways in which they share and exploit the conventions and play a part in their dynamic development.
4. Historical corpus studies can help us understand not only the collective long-term trends but also the role played by individual authors.
5. The usage patterns and profiles of individual authors can provide an indication of how their uses of a given construction were 'represented' in their minds in terms of entrenched patterns of associations.

Essentially, then, this paper is an investigation into individual differences in usage patterns between authors. These are interpreted in terms of degrees and types of individual entrenchment, on the one hand, and in relation to collective conventions and their changes, on the other. We will investigate the use and development of only one construction consisting of an abstract noun followed by a form of the copula BE and a *that*-clause (see Section 2). Our study is meant to show that individual differences in entrenchment are a key to understanding constructional change specifically and language change in general.

The paper is structured as follows: in Section 2 we briefly introduce the construction under investigation. Section 3 is devoted to outlining the so-called 'Entrenchment-and-Conventionalization Model' (EC-Model; Schmid 2014a and 2015). This will serve as a framework which allows formulating specific predictions concerning degrees and types of entrenchment. In Section 4 we detail our corpus-based methods and material. Section 5 provides the results of the corpus study in terms of the general historical development of the target construction. This is followed (Section 6) by a quantitative and qualitative investigation of the usage profiles of individual authors and an interpretation in terms of the EC-Model. Section 7 shows how the perspectives of entrenchment and conventionalization are linked.

2 The Construction under Investigation

As mentioned in the Introduction, the construction in focus, serving as a case study, has the form of an abstract noun, followed by a form of the copula BE, and a *that*-clause. We will refer to this construction as the 'N+BE+*that*-construction'. The beginning of an entry in Samuel Pepys' famous diary can serve as an illustration:

(1) 29th. [October 1664] All **the talk is that** De Ruyter is come over-land home with six or eight of his captaines to command here at home, [...]. (*The Diary of Samuel Pepys*, kept from January 1660 to May 1669, first published 1825)

In this example, the three parts of the 'N+BE+*that*-construction' are represented by *the talk*, by *is*, and by *that De Ruyter is come over-land* respectively. Following common practice in Construction Grammar (cf., e.g., Traugott and Trousdale 2013: 16), we will consider Pepys' utterance as a so-called 'construct' and assume that the production of this construct is licenced by one or more 'constructions' which were represented in Pepys' mind and activated while adding the entry to his diary. Several constructions on different levels of specificity could have motivated the construct in example (1):

a. a very general, schematic construction of the type Det+N+copula+*that*-clause, whose meaning/function could be glossed as 'THING-concept (encoded by the noun) encapsulates proposition (encoded by the *that*-clause)';[3]
b. a more specific but still schematic version, a sub-schema, which would take into account that *talk* is a linguistic noun; so the sub-schema would be Det+$N_{linguistic}$+copula+*that*-clause, associated with the more specific meaning 'speech-reporting noun encapsulates message';
c. a fixed expression, *all the talk is that*, i.e. a substantive, lexically filled construction, which could roughly be glossed as 'here is what people talk about these days'.

Which of these constructions motivated the use of (1) in Pepys' mind will of course remain in the domain of wild speculations. We believe, however, and will try to show in this paper, that the close scrutiny of data from historical corpora can indeed provide information on potential representations of constructions in the minds of historical authors or, couched in terms of the framework favoured here, on the relative degrees of entrenchment of different types of associations. It is to this framework that we will turn next.

3 We follow Langacker's (1987) ideas on word-classes and Schmid's (2000) concept of shell nouns here.

3 A Rough Sketch of the 'Entrenchment-and-Conventionalization Model'

The theoretical framework informing this study is the so-called 'Entrenchment-and-Conventionalization Model', or EC-Model for short, developed by one of the authors over the past years (cf. Schmid 2013: 106–107, 2014a: 242–254, Schmid 2015). Essentially, the EC-Model tries to provide a coherent account of how grammar emerges from usage in social situations and keeps changing under the influence of usage. We will only give a very rough sketch of the model to leave sufficient space for the empirical concerns dominant here. A summary of the theoretical aspects can be found in Schmid (2015).

The major elements of the model are summarized in Figure 1 (taken from Schmid 2015). Usage and the four types of activities invariably involved in it – sensory, motor, cognitive, and social activities – constitute the core of the model, thus marking the framework as belonging to the group of usage-based models. Its key assumption is that what we generally assume to be 'language' or 'the linguistic system' comes about and is continuously updated by the interaction of two types of processes: a limited set of cognitive processes operating in the minds of speakers, subsumed under the label 'entrenchment', and a limited set of socio-pragmatic processes operating in communities, subsumed under the label 'conventionalization'. There are three entrenchment processes – association, routinization, and schematization – and four conventionalization processes – innovation, co-adaptation, diffusion, and normation. The interaction of entrenchment and conventionalization processes depends on usage and on the activities involved in usage, and is influenced by a (probably open-ended) set of cognitive, emotive-affective, pragmatic, and social forces.

Both entrenchment and conventionalization can only take place if similar usage activities of all four types are repeated in similar situations, serving similar functions. Neither routinization on the cognitive side, nor diffusion on the social one, is possible without repetition. As is well known (cf., e.g., Bybee 2003 and Diessel 2007 for a survey), the frequency of production of and exposure to repeated usage is among the key determinants of entrenchment, and the frequency of occurrence of utterance types among those of conventionalization.

Generally speaking, frequency affects entrenchment as follows: the productive or receptive processing of a given usage event will leave a memory trace of the neuronal and cognitive patterns of activation required for processing it. If the same or a similar usage event recurs several times, the pattern of activation will be strengthened and/or the commonalities of these usage events will be 'stored' in the form of a second-order 'representation', i.e. a schema. The first of these two

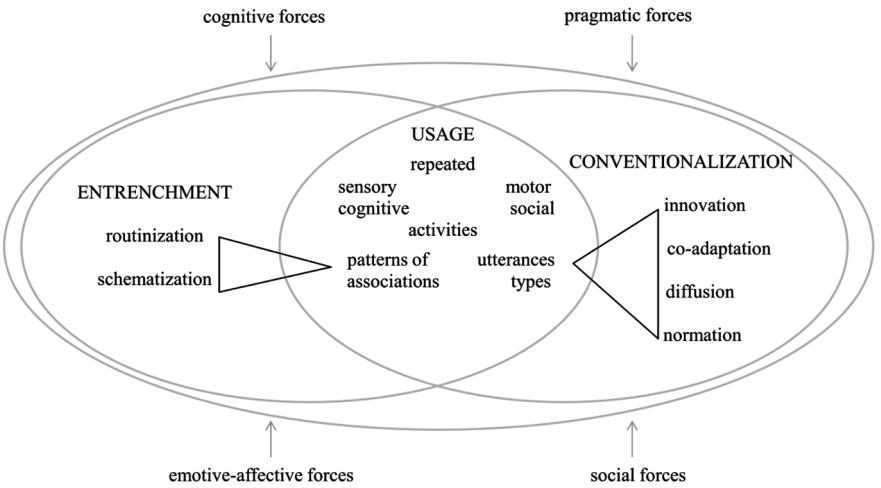

Figure 1: General outline of the EC-Model (taken from Schmid 2015)

processes is 'routinization', the second 'schematization'. In the EC-Model, both processes are considered to operate over patterns of different types of associations (see below). Note that the routinization part of entrenchment is essentially based on a form of associative or Hebbian learning, while schematization – explicitly treated as an entrenchment process, too (cf. Schmid 2014b) – builds on the cognitive abilities of categorization and analogy.

It is of crucial importance to understand that different kinds of repetitions will have different effects on these entrenchment processes. In the EC-Model, this differentiation is implemented by a distinction between four types of associations – symbolic, pragmatic, syntagmatic, and paradigmatic associations – and by formulating systematic predictions concerning the ways in which different types of repetition affect the way in which the processes of routinization and schematization unfold.

Symbolic associations link the forms and meanings of linguistic elements in the minds of language users and thus afford the symbolic power of language. Entrenched symbolic associations are the cognitive and neuronal substrate of what we traditionally call 'linguistic signs', i.e. morphemes, lexemes, and grammatical constructions. Symbolic associations are routinized by the repeated processing of **identical form-meaning pairings**, triggered, for example, by exact repetitions of the same word-forms or fixed expressions. This type of entrenchment, **cotext-free entrenchment**, will facilitate later activation of the same symbolic association and reduce the amount of time and effort required to retrieve the form and meaning of an element in production and comprehension.

Words and constructions that are ordered sequentially in a given utterance trigger **syntagmatic associations**. In language comprehension, syntagmatic associations are required for integrating meanings, while in production they are a major force in the sequential arrangements of the component parts of utterances. Syntagmatic associations link associations activated by the sequential processing of linguistic forms and meanings. They are strengthened by the repeated processing of **sequences of identical or similar linguistic elements**. This can result in a 'chunking' of the given sequence and the emergence of a symbolic association connecting the whole chunked form to one non-compositional meaning, accompanied by a reduction of the strength of the symbolic associations of its parts. Valency patterns, collostructions, collocations, idioms, and other types of co-occurrence patterns that allow language users to predict what will come next rely on more or less deeply routinized syntagmatic associations. The routinization of syntagmatic associations is referred to as **cotextual entrenchment**.

Paradigmatic associations link associations triggered by processing the forms or meanings of linguistic elements to alternative associations that are potentially co-activated. Paradigmatic associations thus activate what could have been said or meant instead of what was said or meant. Paradigmatic associations strongly interact with pragmatic and especially syntagmatic associations; in fact, they depend on them because paradigmatic alternatives only arise within a given linguistic cotext and situational context. Paradigmatic associations are thus essentially probabilistic expectations that depend on and are created by syntagmatic and/or pragmatic associations. Paradigmatic associations are routinized by the repetition of **different elements in an identical or similar cotextual or contextual environment**. Thus, paradigmatic associations connect the lexemes that can fill the variable slot in a schematic or partly schematic construction, for instance the nouns that can be inserted in the nominal slot in the 'N+BE+*that*-construction'. Paradigmatic associations also link the constructions that are potentially activated by pragmatic associations to recurrent types of communicative events, for instance *hello, hi, hi there, good afternoon* in a greeting situation. The cognitive process of analogy plays a key role in this routinization process, identifying the shared role elements play in a given cotext or context. Crucially, the paradigmatic associations between elements competing for a given slot in a grammatical environment and the syntagmatic or pragmatic associations creating this environment work together in the process of **schematization**. Syntagmatic associations are responsible for establishing the links between sequentially arranged elements or variable slots, e.g. the noun slot, the copula slot, and *that* in the 'N+BE+*that*-construction', paradigmatic ones for the links between the elements competing for occurrence in variable slots, i.e. the nouns and the different

forms of the copula. The schemas resulting from this process are strengthened by the **repeated processing of different elements instantiating the same schemas**.

Finally, **pragmatic associations** connect symbolic, syntagmatic, and paradigmatic associations and their component parts to mental states activated by perceptual input from the usage event and by subsequent spreading activation and inferential mechanisms. They are regarded as encompassing information about the physical (time, place, props, etc.) and social situation (participants and their social roles), about the larger preceding linguistic cotext (what was said or written before the current utterance), and about the pragmatic acts, moves, and intentions of discourse participants, including inferential mechanisms like implicatures or irony. Pragmatic associations become routinized by the repeated processing of **identical or similar linguistic elements under similar contextual circumstances**, resulting in **contextual entrenchment**. Faced with a specific situation, contextually more entrenched elements are more likely to be activated than elements that are not pragmatically associated, even if the latter show a higher degree of cotext-free entrenchment.

As pointed out before, a more detailed description of the EC-Model can be found in the publications mentioned above, especially Schmid (2015). The ways in which the model can be operationalized for the investigation of usage patterns in corpora will be formulated in Section 5. Before, however, we will introduce the material used in this study and explain how it was collected and analyzed.

4 Material and Methods

Several historical corpora were used to retrieve authentic attestations of the 'N+BE+*that*-construction'. The material ranges from 1250 up to the arbitrarily chosen point of 1871 (the publication of George Eliot's novel *Middlemarch*). The resulting body of data was used firstly to understand how the construction evolved and developed during this period, i.e. for the study of its conventionalization, and secondly, this macro-development and its stages serve as a benchmark against which data from individual authors can be assessed in terms of entrenchment. The eight corpora listed in Table 1 were accessed online or retrieved from the Internet. In addition to listing the sources, Table 1 renders the numbers of words, the text types contained in the corpora, the time span covered by the sources, as well as the numbers of tokens of the 'N+BE+*that*-construction' they yielded.

Table 1: Data acquisition: eight corpora accessed online or retrieved from the Internet[4]

Corpus	Approx. Number of Words	Text Types	Time	Number of Tokens
Gutenberg	18,500,000	Fiction, Essays, Chronicles	1250–1871	1,039
OED3	20,000,000 (est.)	Mixed	1250–1871	194
Helsinki	1,500,000	Mixed	730–1710	13
Old Bailey	17,000,000	Court Proceedings	1720–1871	153
Paston	200,000	Letters	1422–1509	9
PCEEC	2,200,000	Letters	c. 1410–1695	140
Shakespeare	900,000	Drama, Poetry	1590–1612	17
Jane Austen's Letters	140,000	Letters	1796–1817	23
				Total 1,588

The bulk of the raw text material for this study was downloaded from the Project Gutenberg website offering novels, essays, treatises, chronicles, and other texts. The work of 83 authors spanning the period from 1250 to 1871 is represented in the version of the Gutenberg sub-corpus used for this study. Another important contribution to our material is made by the quotation database of the *OED3*, which contains an estimated 20 million words in the quotations from that period (Rohdenburg 2013: 145). The Old Bailey corpus (see Huber 2007) collects the court proceedings of the Old Bailey court in London from 1720 to 1913. In addition, the classic historical corpora of the English language – the Helsinki Corpus, the Paston letter corpus, the Parsed Corpus of Early English Correspondence (PCEEC), the Online Shakespeare concordance, and the collection of Jane Austen's letters available at the University of Oxford Text Archive, were all used for data retrieval. As Table 1 shows, roughly two thirds of the material comes from the Gutenberg collection; the rest is scattered across the other corpora, with *OED3*, the Old Bailey corpus and PCEEC contributing quite substantial numbers of observations (194, 153, and 140 attestations respectively). The data was retrieved from these sources by means of different search queries adapted

4 Detailed references to the corpora can be found at the end of the paper.

to the corpus in question and its query language. The results of all queries were subjected to manual post-processing, leaving us with the 1,588 attestations of the 'N+BE+*that*-construction' from 139 known authors plus numerous more from the Old Bailey corpus and other sources. The calculation of relative, or normalized, frequencies per author is only possible for the data from Project Gutenberg, the complete works of Shakespeare and Jane Austen's letters because total numbers of words are known for these sources. The data from the Helsinki Corpus and the Old Bailey corpus do not allow for the calculation of author-related relative frequencies since data on overall amounts of words by authors are not available.

All 1,588 examples were coded in terms of the lexical, semantic, textual, pragmatic, and user-related variables listed in Table 2: the variables SOURCE, AUTHOR, and DATE are self-explanatory. All noun tokens were LEMMATIZED and classified on two levels of granularity according to their SEMANTIC TYPE, following the categories introduced by Schmid (2000). TEXT TYPES were also classified in broad terms and so were the PRAGMATIC FUNCTIONS of the utterance as a whole. In addition, it was recorded whether the particular token was a hapax (±HAPAX) within the scope of our dataset and whether the noun type is still used in the 'N+BE+*that*-construction' in Present-Day English (±OBSOLETE). Obsolescence was operationalized by checking the nouns against Schmid's (2000) inventory of nouns occurring in the 'N+BE+*that*-construction' in Present-Day English.[5]

Table 2: Data coding: variables and values

Variable	Values
SOURCE	name of corpus (Gutenberg, PCEEC, etc.)
AUTHOR	author name
DATE	year of publication of original source
NOUN LEMMATIZED (normalization of spelling variants)	noun lemmas, e.g. ISSUE for *issue* and *yssue* or INTENT for *intent* and *entent*

5 In addition, all 1,588 tokens were analyzed and coded with regard to the type of determiner preceding the noun (e.g. article, possessive pronoun, demonstrative pronoun), the tense of the copula, and whether or not there were words intervening between the noun and the copula (e.g the insertion *of my ring* in the example *and the virtue of my ring is that* [...] found in Thomas Malory's *Le Morte d'Arthur*, 1485). Since these variables are not discussed in this paper, however, we will not pursue them any further.

Variable	Values	
SEMANTIC TYPE (semantic classification)	circumstantial	e.g. *time, way, tradition*
	eventive	e.g. *event, step, action*
	factual	
	attitudinal	e.g. *difficulty, advantage*
	causal	e.g. *cause, effect*
	comparative	e.g. *difference*
	evidential	e.g. *sign, proof*
	neutral	e.g. *fact, thing, point*
	partitive	e.g. *feature, characteristic*
	linguistic	
	assertive	e.g. *claim, answer, report*
	directive	e.g. *order, commandment*
	propositional	e.g. *news, doctrine, rumour*
	mental	
	conceptual	e.g. *idea, thought, sense*
	creditive	e.g. *belief, assumption*
	emotive	e.g. *consolation, concern, hope*
	volitive	e.g. *desire, pleasure*
	modal	
	deontic	e.g. *duty, law*
	dynamic	e.g. *utility, privilege*
	epistemic	e.g. *truth, probability*
TEXT TYPE	chronicle	
	essay	
	fiction	
	letter	
	other (incl. diary)	
FUNCTION	argumentative	
	descriptive	
	directive	
	explanatory	
	expressive	
	reporting	
HAPAX	yes/no	
OBSOLETE	yes/no	

5 Findings I: The Historical Development of the 'N+BE+*that*-Construction'

We will provide the following pieces of information in order to describe the general historical development of the 'N+BE+*that*-construction': the earliest known attestation (Section 5.1); formal (5.2), semantic (5.3), and pragmatic changes (5.4); changes concerning usage frequencies (5.5); and sources of innovation (5.6). More details using a complementary method of analysis can be found in Mantlik (2011).

5.1 First Known Attestation

The earliest attestation of the 'N+BE+*that*-construction' in our database was found in the Helsinki Corpus and comes originally from *The Appeal of Thomas Usk against John Northampton* written in 1384:

(2) And, truly, **the ful entent** was that al the ordinances that wer ordeyned in hys tym [...]. (1384, Helsinki Corpus, *The Appeal of Thomas Usk against John Northampton*)

The year 1384 is thus the *terminus ad quem* – the time from when on we can be certain that the construction was actually used (cf. Mantlik 2011: 11).

As far as the origin of the construction in English is concerned, we assume that we are dealing with a case of structural borrowing from Romance (Mantlik 2011: 174, 195). Cognate constructions are attested in both Latin and Old French (cf. Mantlik 2011: 173–195). The heavy preponderance of nouns of Romance origin in the construction also points in this direction (cf. the examples listed in Table 2 above; see Mantlik 2013 on the etymology of shell nouns, and Mantlik 2011 for a detailed discussion of the predominant origins of the nouns occurring in shell-noun-constructions). Very early nouns of this type are *garnison* (1386, Chaucer), *proclamation* (1387, unknown author quoted in *OED3*), *counsel* (1393, Gower), *medicine* (1397, Trevisa), and *statute* (1425, Mandeville). Nevertheless, some of the nouns observed to occur early in the construction are of Germanic origin, among them *mete* (1430, Chaucer, see example 5 below), *will* (1430, Chaucer, example 4), *thing* (1430, Chaucer), *kind* (1425, Mandeville, see example 8), and *sooth* (1397, Trevisa, example 10; also used as many as five times by Chaucer, 1430).

5.2 Formal Changes

Overall, the formal properties of the inflexible parts of the 'N+BE+*that*-construction' have remained astonishingly stable. Two aspects seem worth reporting. Firstly, the insertion of a comma between the copula and *that* remained common, though unsystematic, until the eighteenth century.[6] Sporadic commas are found as late as 1871, in George Eliot's *Middlemarch*, at least in the edition downloaded from Project Gutenberg. Secondly, the omission of the complementizer *that* starts in the early seventeenth century with the noun *truth*.[7] The first attestation of the zero variant was found in *The Proficiency and Advancement of Learning* by Francis Bacon published in 1605.

(3) [...] and so goeth on in an irony. But **the truth is**, they be not the highest instances that give the [...]. (1605, Francis Bacon, *The Proficiency and Advancement of Learning*)

It took considerable time until other nouns, the next one being *fact*, were also used with complementizer omission and until the zero-complementizer variant began to occur with noteworthy frequencies (cf. Mantlik and Schmid forthcoming for more details).

A key indicator of the development of the construction is the question which nouns were used in the nominal slot. While details about the types of nouns will be discussed in the next section devoted to semantic changes (5.3), information on the quantitative aspects is better placed here. Overall, the 1,588 tokens in our database represent 293 types. As many as 133 of these are hapax legomena in our data. The introduction of new nouns in the variable slot – which are hapaxes by definition – reflects the vitality of the schematic construction and can be interpreted as a sign of its continuing productivity (cf. Hilpert 2013: 127–133). Remarkably, this productivity remains high throughout the whole period of our investigation. This is demonstrated by Figure 2, which shows the number of hapaxes per 50-year period.[8] The figure indicates that new collocates which were not repeated (at least until 1871) continue to be introduced.

6 To what extent the use of commas was influenced by editorial practices and decisions was not investigated.

7 As regards complementizer omission with verbs, Rissanen (1991) found a steady relative increase of zero complementizers from 14 per cent between 1350 and 1420 to almost 70 per cent in the period between 1640 and 1720. Finegan and Biber (1995) showed that this development is also subject to register variation.

8 The division of the data into 50-year periods is potentially problematic because it may be too coarse and runs the risk of introducing boundaries that are not motivated by the structure of the

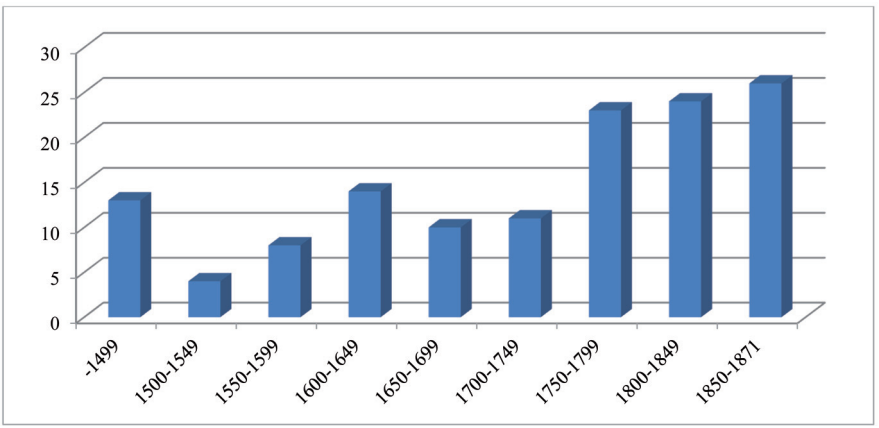

Figure 2: Absolute number of hapaxes in the dataset per 50-year period

The impression that the nominal slot of the construction is subject to considerable change is supported additionally by the large number of nouns in the dataset which are no longer used in the construction in Present-Day English. As many as 160 obsolete types are attested,[9] which corresponds to a proportion of 55 per cent of the 293 types found altogether. As can be expected, the relative number of tokens of obsolete types generally decreases over time. However, though quite close to Present-Day English, the final period from 1850 to 1871 still musters as many as 45 tokens of nouns that no longer occur in the construction today, among them *meaning*, *purpose*, *sense*, *substance*, and *wonder*.

Overall the findings on hapaxes and obsolete types clearly show that the changes in the collocational range of the nouns occurring in the construction are considerable, reflecting a dynamic development which seems to be fairly typical of language change in general.

---

data. In order to avoid these problems, we applied Gries and Hilpert's (2012) method of variability-based neighbour clustering, which allows for a bottom-up periodization of the data. However, for the present data the method turned out to be too strongly influenced by the large number of outliers (see Section 6.2 below) and suggested a division into no more than three periods spanning the years 1377–1566, 1567–1791, and 1792–1871 respectively. The method was then improved by including the nominal variables NOUN TYPE and FUNCTION (see Hilpert 2013: 38–42). This reduced the effect of outliers in terms of sheer frequency, but did not yield a more fine-grained division. We therefore decided that for the purposes of the present paper, 50-year periods would be a reasonable choice after all.

9 Obsolete types, i.e. nouns that are no longer used in the 'N+BE+*that*-construction', which occur more than five times in our data are *account*, *case*, *counsel*, *end*, *event*, *intent*, *law*, *matter*, *meaning*, *mind*, *object*, *pleasure*, *presumption*, *report*, *sooth*, *substance*, *talk*, *trust*, *will*, and *wonder*.

5.3 Semantic Changes

The changes of the collocational range also affect the semantic development of the construction, of course. While we cannot go into detail here, the most remarkable semantic change is demonstrated by the line charts provided in Figures 3 and 4. Both figures render the relative frequencies of tokens of different classes of nouns per 50-year period.[10] Figure 3 collects five classes whose relative frequencies decrease, Figure 4 four classes whose share of the total data per period increases.

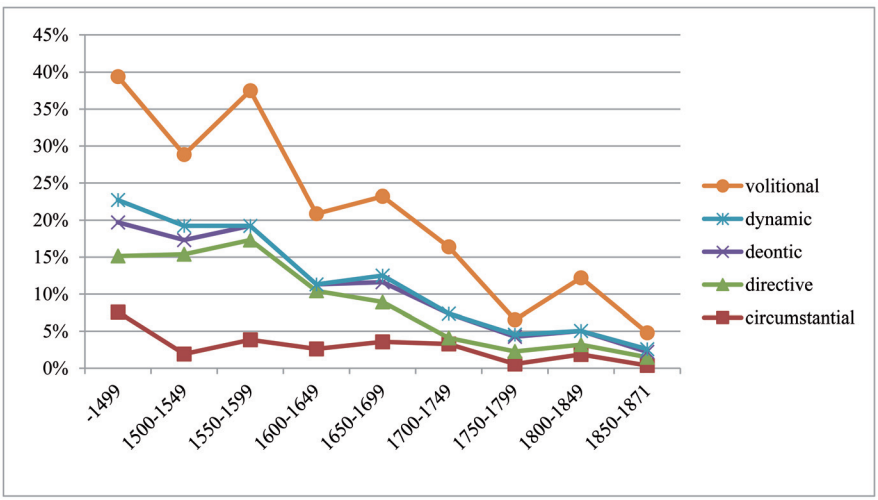

Figure 3: Decrease in relative frequency of nouns expressing volition, dynamic and deontic modality, directive illocutions, and circumstance

10 We refrain from discussing the statistical significance of the trends that can be gleaned from the two figures because they are co-determined by other variables, most notably by the highly variable availability of sources in different periods and their genres and text types. The overall trend seems strong enough. Given the comparatively low prominence of this aspect in the present paper, we do not think that it is necessary to subject the data to multivariate statistics such as regression models.

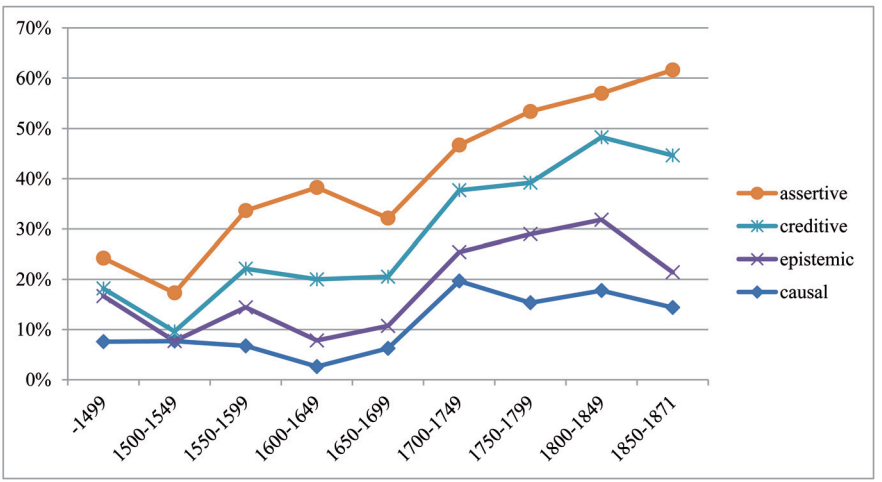

Figure 4: Increase in relative frequency of nouns expressing assertion, belief, fact, and cause

The types of nouns collected in Figure 4 share semantic components related to actions and events: volition (see example 4), dynamic modality (5), obligation (6), giving orders (7), and circumstance (8):

(4) volitional: **The wyl** of crist **was that** she sholde abyde. (Chaucer, *Canterbury Tales*, 1430)
(5) dynamic: **My mete** ['means of support of strength', *OED3*, s.v. *mete*] **is that** I do the will of him that sente me. (*Wycliffe Bible*, John 4:34, 1384, *OED3*)
(6) deontic: **The lawe** of Medis and Persis **is that** eche decree whiche the kyng ordeyneth be not leeful for to be chaungid. (*Wycliffe Bible*, Daniel 6:15, 1384, *OED3*)
(7) directive: **Our request is, that** you would by the bearer of these presents, [...]. (Hakluyt, *The Principal Navigations, Voyages, Traffiques and Discoveries of the English Nation*, Vol. 1, 1598)
(8) circumstantial: **The kind** of the mors **is that** they wole leefe noothing empty besides hem. (Mandeville, *Mandeville's Travels*, around 1425, *OED3*)

As shown in Schmid (2000), in Present-Day English all these nouns are much more likely to be complemented by *to*-infinitives than by *that*-clauses. In contrast, the classes of nouns brought together in Figure 4 are related to abstract states and relations, the epistemic domain, and linguistic statements rather than actions: states of believing (see example 9), epistemic modality (10), causal relations (11), and assertive illocutions (12).

(9) creditive: **My judgment is, that** they ought all to be despised. (Bacon, *Essays of Francis Bacon*, 1627)

(10) epistemic: [...] but **the sooth** ['truth'] **is, that** they were strong hores [...]. (Trevisa, *De Proprietatibus Rerum*, 1397)
(11) causal: **The cause is that** they labour do despyse. (Brand, *The Ship of Fools*, 1494)
(12) assertive: And **your reporte is that** you haue done it with my consent. (PCEEC, 1573)

Figures 3 and 4 reflect a trend in the history of English that resulted in a semantic distinction between verbs and nouns preferring infinitival complements and those usually taking *that*-clauses (cf., e.g., Givón 1990: 517–561). The former cluster, corresponding to Givón's "manipulative verbs", is associated with actions and events taking place in the real world, as well as obligations, wishes, aims, etc., the latter, corresponding to Givón's "cognition-utterance verbs", with ideas, thoughts, beliefs, and abstract relations such as causation. The two figures show that nouns belonging to the event-related group used to be complemented by *that*-clauses in the early period, but – as a comparison with their use in Present-Day English indicates – were later 'absorbed' by the infinitive. The increase in the proportion of cognition-utterance nouns is caused by larger numbers of both types and tokens manifesting this semantic cluster.

5.4 Pragmatic Changes

The main parameter of pragmatic change is the function of the utterance containing a given N+BE+*that*-construct. The data represented in Figure 5 show trends which partly correlate with the semantic development: first, the explanatory function is strong in the very first stage following the introduction of the construction, but begins to lose ground soon. This function is prominent in examples 5, 6, 8, 11 and 12. Second, being represented quite strongly in the data from the sixteenth century, the directive function, illustrated in example 7, virtually becomes extinct at the beginning of the eighteenth century, coinciding with the decline of uses of the 'N+BE+*that*-construction' for the cluster of nouns collected in Figure 3. Third, the opinative (cf. example 9) and especially the reporting functions (cf. example 12) gain in relative importance, which is matched by the increase in the use of creditive and assertive nouns shown in Figure 4.

As far as text types are concerned, we have only investigated the data from those sources whose extent – in numbers of words – we could identify with sufficient reliability for calculating normalized frequencies of occurrence. This part of the data yielded the relative frequencies rendered in Figure 6. The 'personal' genres biography, letter, and diary reach higher relative scores than essays and chronicles, followed by prose fiction (novel, romance, satire) and more formally restricted forms of literary writing (drama, poem). Overall, less formal

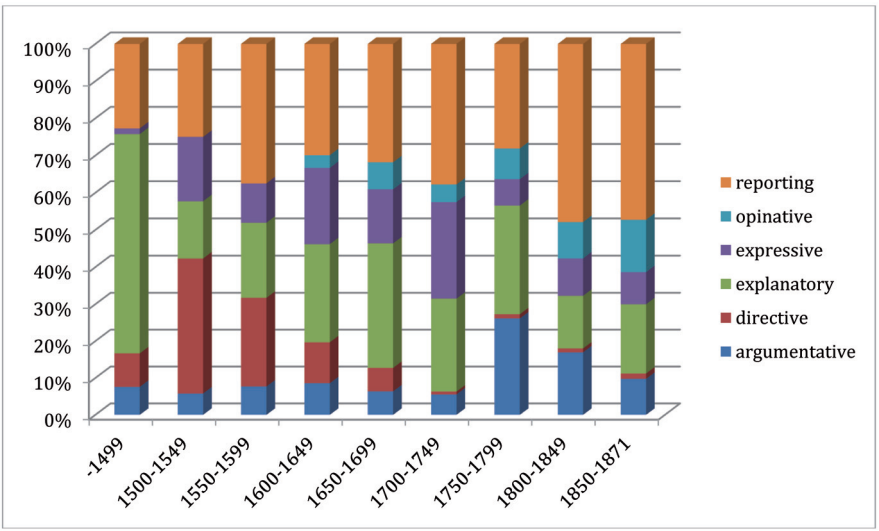

Figure 5: Changes in distribution across functions (four additional tokens which serve the descriptive function are omitted in the figure for the sake of clarity)

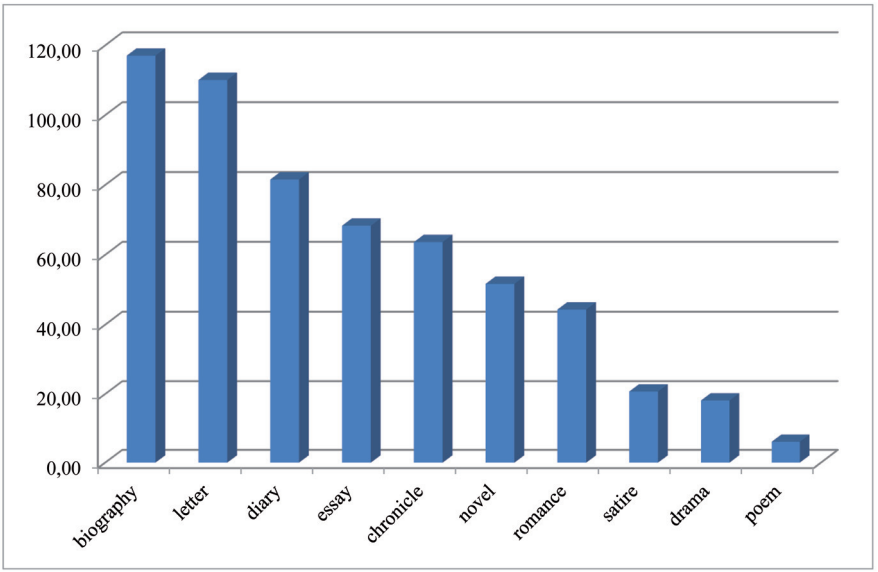

Figure 6: Frequency per million words in different genres (n = 1069)

genres seem to provide a better habitat for the construction to thrive in than more formal ones. Given notorious problems with genre classifications across centuries, and given the influence of other variables such as topic, style, and of course time, the reliability of these findings should not be overestimated, however.

5.5 Changes in Frequency

Figure 7 provides the relative frequencies per million words per 50-year period. Not surprisingly, the construction is used with increasing frequency, with a noticeable upward spike found towards the end of the eighteenth century. Since the availability of sources that could be added to our convenience sample and the genres of these sources of course influence the overall development in terms of frequency of use of the 'N+BE+*that*-construction', we refrain from venturing an interpretation of this finding. The main purpose of Figure 7 is to serve as a backdrop for the more fine-grained analysis reported in Section 6.2 below.

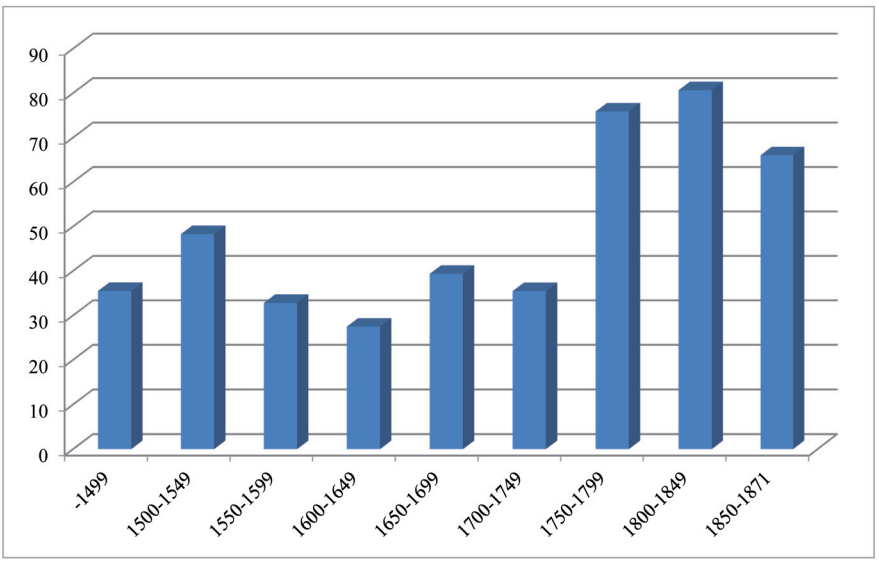

Figure 7: Relative frequency per million words per 50-year period

5.6 Locus of Innovation

Innovation regarding the use of the target construction will be described here in terms of extensions of the collocational range by introducing nouns not attested before. Table 3 provides the dates and sources of the earliest attestations of selected nouns. Anticipating the author-related perspective that will dominate in Section 6, the names of innovators – always relative to our database – are listed as well, if known. To give a more systematic presentation of the data is neither required nor reasonable, as it would again mainly be a reflection of the data situation. The main point to be gleaned from Table 3 is that it does not reveal a consistent trend concerning the question whether the semantic and pragmatic changes described in the previous sections are changes 'from above', as would be indicated by a preponderance of innovations in 'learned' writing, or 'from below', as suggested by innovations in personal letters or direct speech in fiction, especially when attributed to figures of low social standing. The preponderance of nouns of Romance origin already mentioned above speaks in favour of the idea that innovations come 'from above' rather than 'below'.

Table 3: Dates and sources of first attestations of selected nouns in the 'N+BE+*that*-construction'

Noun	Date	Type of Source	Author
condition	1430	fiction	Chaucer
meaning	1430	fiction	Chaucer
thing	1430	fiction	Chaucer
answer	1469	letter	
opinion	1470	letter	
cause	1474	essay	Caxton
advice	1517	letter	
truth	1519	letter	
desire	1548	fiction	Hall
conclusion	1567	fiction	Painter
charge	1574	letter	
wish	1585	letter	
reason	1589	letter	
difference	1591	letter	

Noun	Date	Type of Source	Author
news	1592	fiction	Shakespeare
comfort	1595	fiction	Shakespeare
hope	1604	letter	
belief	1651	essay	Hobbes
fact	1669	diary	Pepys
misfortune	1677	letter	
consequence	1709	essay	Berkeley
idea	1739	essay	Hume
result	1744	essay	Haywood
substance	1756	essay	Burke
probability	1762	essay	Burke
agreement	1790	essay	Burke
case	1791	essay	Paine
plan	1796	fiction	Austen
thought	1848	fiction	Thackeray

As the table indicates, a number of nouns are first attested in letters and then found to occur in published sources. Essays are a source of more formal types of nouns used mainly with argumentative and explanatory functions. Fiction contains a variety of innovations in the construction, ranging from linguistic, over factual, to modal nouns.

Summing up our bird's eye view of the diachronic development of the 'N+BE +*that*-construction', we can recapitulate that the construction has been subject to considerable change with regard to frequency, collocational range, meaning, and function.

6 Findings II: Usage Profiles of Individual Authors Interpreted in Terms of Entrenchment

Keeping the findings reported in the previous section in mind, we will now proceed to a comparison of the usage profiles of individual authors. Two goals motivate this endeavour: we want to show, firstly, that individual authors participate in and contribute to the collective development of the construction in very

different ways, and secondly, that the usage profiles of individual authors provide clues as to the way in which the construction may have been entrenched in their minds. We will begin by spelling out the rationale underlying our analysis and then report findings pertaining to the different types of entrenchment processes and associations discussed in Section 3.

6.1 Assumptions and Predictions

The basic assumption behind this investigation is that different degrees of entrenchment (in terms of routinization) and different types of entrenchment (in terms of types of associations and levels of schematization) correlate with and are manifested in the usage patterns of individual authors. This in turn rests on the conjecture that strongly entrenched patterns of associations are more likely to be activated than less strongly entrenched ones, and that therefore frequencies of patterns of utterances reflect patterns of entrenchment. For example, if, as will be shown, Samuel Pepys stands out by repeated uses of the expressions *all the talk is that* and *all the news is that*, then this can be interpreted as indicating that these expressions were licenced by strong syntagmatic associations resulting in a chunk-like representation of these sequences in his mind.

By reversing the perspective of the EC-framework outlined in Section 3, these assumptions can be translated into the following predictions:

1. General prediction: the usage patterns of comparable individual authors will differ. These differences can be interpreted in terms of different degrees and types of entrenchment.
2. Specific predictions:
   a. Authors will differ regarding the relative frequencies with which they use the 'N+BE+*that*-construction'. These differences reflect different strengths of symbolic associations linking the form(s) and meaning(s) of the construction.
   b. Authors will differ regarding the preferences and ranges of nouns used in the construction, and the paradigmatic relations between these nouns, and with regard to their inclination to use new nouns (innovations). These differences reflect different strengths of paradigmatic associations and different types of schemas.
   c. Authors will differ regarding their usage of semi-fixed strings. These differences reflect different strengths of syntagmatic associations between the component parts in their minds.

d. Authors will differ regarding the functions with which and the contexts and cotexts in which they use the construction. These differences reflect different strengths of pragmatic associations.

These predictions will be taken up in the following sections.

6.2 Frequency of Usage

To provide a first overview, Figure 8 plots the relative frequencies of the construction per work. Since the data represented in Figure 7 showed an average frequency of around 30 tokens of the 'N+BE+*that*-construction' per million words before 1750, which means that one can expect one occurrence in approximately 30,000 words, the plot includes only those works that contain more than 30,000 words. Zero attestations in texts that are shorter than 30,000 words are not informative.

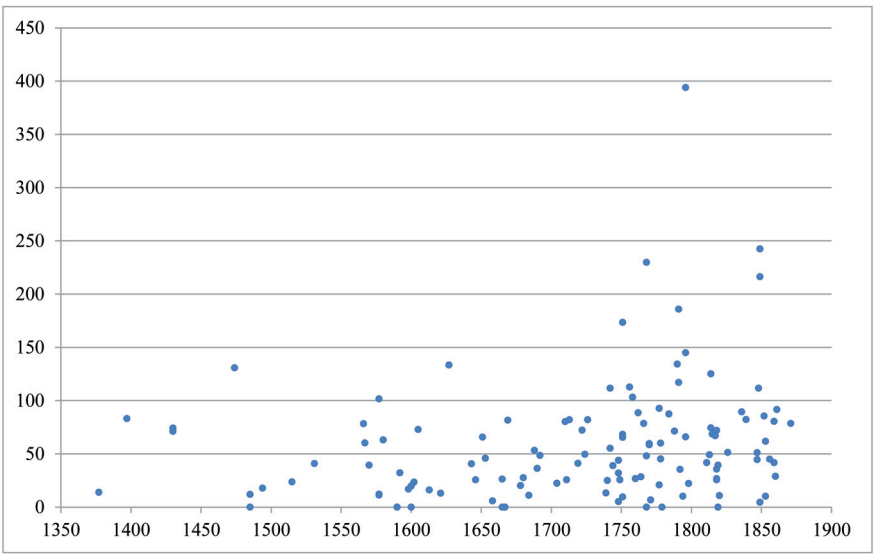

Figure 8: Frequencies of occurrence per million words per source (only works containing more than 30,000 words)

The plot confirms the overall trend to be observed in Figure 7. It also shows, however, that as expected (see note 8 above), reporting average frequencies per 50-year period runs the risk of glossing over quite substantial authorial differences. In fact, there is considerable variation between the sources, with several of

them not containing a single attestation of the construction on the one extreme, and an outlier of almost 400 occurrences per million words – Thomas Paine's *The Age of Reason* (1796), boasting as many as 26 occurrences in 65,986 words – on the other. Small or zero scores at the low end of the scale only seem remarkable if they diverge significantly from the general trend of the period, and if this happens in spite of a large output. Cases in point are Thomas Malory, who produces a mere two instances in the 352,791 words making up the two volumes of *Le Morte d'Arthur* (1485); John Milton with not a single one in almost 80,000 words in *Paradise Lost* (1667); and Tobias Smollett with only one occurrence each in his two hefty volumes *The Adventures of Peregrine Pickle* (1751; 317,864 words) and *The Expedition of Humphrey Clinker* (1771; 150,473 words). While nobody would conclude that Milton did not have command of the 'N+BE+*that*-construction', these very low frequencies of usage may indicate that the construction was not sufficiently deeply entrenched for it to be activated in language production more often, at least as far as the genre represented by *Paradise Lost* is concerned (see Section 6.5 for more details on the role of genres).

Of course, Figure 8 can give only a very crude picture, because it lumps together authors and works of very different types. In order to get a more detailed view of individual differences, sources which are comparable in terms of period and genre should be selected. Data for four groups that lend themselves to such comparisons, three of contemporary fiction writers and one of philosophers, are collected in Table 4:

Table 4: Comparison of authors from selected periods writing the same genre

Author	Dates of Publications	Number of Words in Source	Number of Occurrences of N+BE+*that*	Relative Frequency per Million Words	χ^2 Test Results
Mid-18th-Century Fiction					
Richardson	1740–1748	1,406,495	42	29.86	$\chi^2 = 19.1447$ df = 3 p = 0.0002552
Fielding	1742–1751	691,821	33	47.70	
Haywood	1744–1748	170,858	7	40.97	
Smollett	1748–1771	659,690	5	7.58	
Early 19th-Century Fiction					
Austen	1795–1818	887,680	71	79.98	$\chi^2 = 12.7824$ df = 2 p = 0.001676
Shelley	1818–1826	250,729	11	43.87	
Scott	1814–1820	983,630	41	41.68	

Author	Dates of Publications	Number of Words in Source	Number of Occurrences of N+BE+*that*	Relative Frequency per Million Words	χ^2 Test Results
Mid-19th-Century Fiction					
Brontë, E.	1847	117,276	6	51.16	$\chi^2 = 21.037$ df = 3 p = 0.0001034
Brontë, C.	1847–1856	395,351	12	27.82	
Dickens	1836–1861	1,409,404	144	102.17	
Thackeray	1848–1852	491,855	50	101.66	
Late 18th-Century Philosophers					
Hume	1739–1779	137,954	6	43.49	$\chi^2 = 58.3611$ df = 2 p = 2.124e-13
Paine	1776–1796	179,138	48	267.95	
Burke	1756–1796	1,515,549	129	85.12	

In addition to authors' names, Table 4 renders dates of publications, overall numbers of words, numbers of attestations of the target construction, and normalized frequencies. Considerable differences can be observed in all four groups. Particularly extreme cases are the contrasts between Smollett and Fielding (7.58 vs. 47.70 occurrences per million words), and Hume and Paine (43.49 vs. 267.95), in both cases in texts of comparable extents (137,954 and 179,138 words respectively). The right-most column reports the results of chi-square tests which indicate that the differences are significant in all four groups. This demonstrates that even authors writing comparable types of texts around the same time differ significantly in terms of the frequency with which they use the construction.

In the terms of the predictions formulated above, this could be interpreted as indicating that symbolic associations connecting potential ideas to be encoded and the form N+BE+*that* in production are not entrenched equally strongly in the minds of these authors. However, these differences in plain frequencies are of course a very crude measure of cotext-free entrenchment which would perhaps not even be worth reporting as such if they were not substantiated by the qualitative findings to be reported next.

6.3 Ranges of Nouns, Preferences for Nouns, Paradigmatic Relations, and Innovations

In this section we are concerned with ways of judging the nature of schemas representing a 'construction' in a given author's mind. According to the EC-Model, this can be done by looking at the following parameters:

- the range of nouns used in the construction, both in terms of different types of nouns as such and in terms of semantic classes;
- the distribution of tokens across types, indicating, among other things, preferences for certain nouns;
- paradigmatic relations between the nouns that are used and preferred by a given author;
- innovations, i.e. nouns that are not attested in the data before a given author's use, indicating the availability of a productive constructional schema that can be used to extend the potential of the construction.

The examination of our data along these parameters suggests a distinction into four basic types of author profiles. These are illustrated by typical examples in the following and summarized in Table 9.

The first type is exemplified by Edmund Burke (see Table 5), whose prolific writing contains 129 tokens of the construction spread across as many as 58 different nouns. This corresponds to a noun type-token ratio of 0.45 tokens per type, indicating considerable consolidation of tokens across type in spite of the large number of hapaxes. The rank list in terms of frequency rendered in Table 5 shows a strong preference for the noun *fact*. The epistemic modal noun *truth* and different types of factual nouns (e.g. causal: *consequence*, *reason*; neutral: *thing*; attitudinal: *fault*) further dominate the list, but mental (*opinion*, *consolation*) and linguistic nouns (*objection*, *answer*, *charge*) are found as well.

Table 5: Collocational range and usage preference for Edmund Burke

Tokens	Nouns
21	*fact*
11	*opinion*
8	*thing, consequence*
6	*objection*
4	*fault, reason, truth*

Tokens	Nouns
3	*answer, substance*
2	*charge, condition, consolation, difference, effect, expression, principle, probability, property*
1	*account, accusation, agreement, article, assertion, authority, circumstance, claim, complaint, concern, confirmation, consideration, defence, dishonour, event, evidence, fear, grievance, idea, issue, justification, limitation, maxim, mischief, notion, object, opprobrium, order, part, plea, pleasure, point, proposal, purpose, regulation, requisition, step, use, view*

As far as paradigmatic relations are concerned, two semantic clusters of closely related nouns can be observed, namely *fact* and *truth*, and *consequence, reason*, and *effect*. Burke introduces three nouns into the construction which do not occur in our dataset before him, namely *agreement, probability*, and *substance* (cf. Table 3). Overall, this leaves the impression that Burke had a strongly entrenched high-level schema linking the form N+BE+*that* to a range of meanings, particularly strongly to the expression of epistemic modality and factuality. The high frequency of the noun *fact* actually indicates that the epistemic meaning variant may have been available as a strong sub-schema.

The second type of profile is represented by Samuel Pepys (Table 6; see the examples in (13) below). His diary contains 30 tokens of the construction manifesting 13 types, which yields a noun type-token ratio very similar to that of Burke. Pepys' semantic range of nouns is much less wide than Burke's, however. He shows a very clear preference for linguistic nouns: as many as 17 of the 30 tokens instantiate the nouns *news, talk, discourse*, the three top-ranking nouns, as well as *answer* and *piece of news*. Despite this strong semantic focus on linguistic nouns, Pepys is on record in our database as the first user of the important noun *fact* (cf. Table 3) and the hapax *expedient*. We will come back to this finding in Section 7. Pepys' usage profile suggests that he does have a high-level schema licencing his two innovations, complemented by a strongly entrenched specific schema linking the form N+BE+*that* to the meaning of reporting linguistic utterance and illocutionary acts (again see the examples rendered in (13) and the discussion in Sections 6.4 and 6.5 below). As will be shown below, the latter is supported by strong syntagmatic relations between the construction and selected nouns.

Table 6: Collocational range and usage preference for Samuel Pepys

Tokens	Nouns
7	*news*
4	*talk*
3	*discourse, fear*
2	*answer, reason, rule, thing*
1	*business, expedient, fact, piece of news, truth*

The third type, represented by Samuel Richardson (Table 7; examples in (14)), is essentially a variant of the second one. There is also a strong focus on a specific meaning variant of the construction, namely the encoding of states of emotion (*fear, consolation, desire, hope, pleasure*, etc.) and attitudes (*misfortune, advantage, disgrace, excellence*, etc.), but this focus is spread across a much wider range of nouns than in the case of Pepys. Paradigmatic associations and analogy seem to play an even greater role here. In addition, the overall range of nouns used by Richardson is wider, which is also reflected in a higher noun type-token ratio of 0.67. This suggests that Richardson's high-level schema was possibly a little stronger than Pepys', but one should not forget that the amount of material contributed by Richardson to our database is much more extensive than that by Pepys.

Table 7: Collocational range and usage preference for Samuel Richardson

Tokens	Nouns
4	*fear*
3	*consolation, result*
2	*comfort, condition, desire, hope, misfortune, pleasure, thing*
1	*advantage, concern, disgrace, doubt, end, excellence, felicity, improbability, intention, mind, news, objection, occasion, opinion, point, trouble, truth, view*

The fourth and final type is represented by Charlotte Brontë (Table 8). Contributing almost 400,000 words to the raw material, she nevertheless produces only 12 instances of our target construction.

Table 8: Collocational range and usage preference for Charlotte Brontë

Tokens	Nouns
1	*consequence, difference, event, fact, news, persuasion, probability, proof, result, thing, thought, worst of the matter*

What is more, not a single one of the nouns she uses in the construction occurs twice, thus yielding a noun type-token ratio of 1, which indicates that she has no collocational preferences. From a semantic point of view, her output makes up a fairly mixed bag of various factual nouns, as well as epistemic, linguistic, and mental ones. We interpret this as evidence for a weakly entrenched and extremely unspecific schema, which hardly manages to reach the level of activation required for frequent production. It is not surprising that just like Samuel Richardson, Charlotte Brontë is not responsible for an innovation in the nominal slot of the construction.

The four types and the profiles are summarized in Table 9. Brief glosses of the interpretation of the profiles in terms of entrenchment are included. Further authors belonging to the four classes are listed as well.

Table 9: Summary of the four types of schematization profiles

	Type 1	Type 2	Type 3	Type 4
Author Words	Burke 1,515,549	Pepys 367,747	Richardson 1,406,495	C. Brontë 395,351
Range Number of Types and Tokens	very large 58 types 129 tokens	limited 13 types 30 tokens	large 28 types 42 tokens	limited 12 types 12 tokens
Noun Type-Token Ratio	0.45	0.43	0.67	1
Semantic Range	wide	limited, but one focus	wide, but one focus	limited, no focus
Paradigmatic Clusters	several, but not very prominent	one prominent cluster: 'linguistic'	one prominent cluster: 'emotive/ attitudinal'	no cluster
Innovations	*agreement, probability, substance*	*fact, expedient*	–	–

	Type 1	Type 2	Type 3	Type 4
Interpretation	strong high-level schema plus several more specific sub-schemas	possibly weak high-level schema, strong sub-schema ('linguistic')	weak high-level schema, strong sub-schema ('emotive/attitudinal')	weak high-level schema
Further Authors of this Type	Austen Chaucer Dickens Eliot Johnson Paine	Berkeley Boswell Cromwell (letters)	Painter Johnson Thackeray	E. Brontë Hobbes Locke Shakespeare Shelley Smollett Walpole

Although the four types are of course idealizations across massive variation, most authors find a natural place in one of the four categories. Perhaps not surprisingly, the general trend is that authors that are particularly strongly represented in our data belong to Type 1, but there are also numerous exceptions: prolific writers such as Samuel Richardson, Tobias Smollett, and Charlotte Brontë, who nevertheless represent the other types, indicate that the allocation of authors to the four types is not just an artefact of the convenience-driven data-structure.

6.4 Semi-Fixed Strings and Patterns

The findings and interpretations reported in the previous section receive additional support by the analysis of semi-fixed strings.

To begin with, the different preferences of individual authors for certain nouns are also a sign of strong syntagmatic associations between the constructional schema and the nouns filling the variable slot: Burke's preference for *fact* and Pepys' preference for *news* are cases in point. Austen favours *truth* and *consequence*, but underuses, among others, *fact*, while Thackeray fills as many as 22 out of his 50 tokens of the construction with the two nouns *truth* and *fact*. This may seem trivial, given the fact that *truth* and *fact* are the most frequent types across centuries, but it turns out to be significant once we realize that as many as 110 out of the 139 authors that are identified as contributors to the database do **not** use the noun *truth*. We will come back to this observation in Section 7.

Some authors, especially those representing the schematization profiles of types 2 and 3, show interesting semi-fixed patterns extending beyond the confines

of the construction as such. In order to show this, we quote selected material from three authors: Pepys, Richardson, and Thomas Cromwell. The selection is determined by the goal to demonstrate patterns. Attestations by these authors that do not illustrate the respective patterns are omitted.

(13) Selected concordance lines from Pepys

therein. Here **all the discourse** is, that now the King is of opinion to have the Parliament
depart the town. **All the discourse** now-a-day is, that the King will come again; and for all I see
the reason of all this. **The great discourse** now is, that the Parliament shall be dissolved
died the last week. **All the news** now is, that Sir Jeremy Smith is at Cales with
her part mighty well. **All the news** now is that Mr. Trevor is for certain to be Secretary
which was made him. **All the news** from London is that things go on further towards a King.
to the office, where **all the news** this morning is that the Dutch are come with a fleet
monk's soldiers. 4th. **All the news** to-day is, that the Parliament this morning voted the House
gentleman. 27th. **All the news** this day is, that the Dutch are, with twenty-two sail of ships
leave of him. But **the best piece of newes** is, that instead of a great many troublesome Lords
they must be purified. **The worst news** he tells me, is that Mr. Chetwind is dead
and hath good rest. **All the talk** is that my Lord Sandwich hath perfected the peace
which vexed me. **The great talk** is, that the Spaniards and the Hollanders do intend to set
good night. 16th. **The talk** upon the 'Change is, that De Ruyter is dead, with fifty men

(14) Selected concordance lines from Richardson

to your honest friends: and **all my pleasure** is, that I can and will make you amends
tribute to her memory. **All my pleasure** now is, that she knew not half my wicked pranks
protection as I could find. **All my comfort** is, that your advice repeatedly given me to
think so. **All the comfort** I know of in children, is, that when young they do with
as you think fit in it. **All my concern** is, that this daring and foolish project, if carried on
always been my principal care. **All my fear** is, that, when she comes to the point,
let him say of me what he will. **All my fear** is, that, as he knows I am in disgrace
of his lady.' And now, Belford, **all my hope** is, that this fellow (who attended us in
her refusal to receive her. **All her consolation** is, that her unhappy situation is not owing to
was no more? **Her principal consolation**, however, was, that she should not long
to my corpse; in this case **my desire is**, that it may be interred in the churchyard
think your going away a fault. **The hope** is, that things will still end happily, and

(15) Selected concordance lines from Cromwell

the Marches of Calays, **the kinges pleasure** is that ye and thother Commyssioners shall
the pleasure of his highnes is that the due dettes of the said howses well prouyde
be enhabited, **his graces pleasure** is that your lordship shall cause some oone or two
The kinges pleasure is that you shal sende a spedy aunswer herof,
remedy whereof **the kinges graciouse pleasur** is that ye shal cause the said persones
his gracious pleasure and commandment is that forasmoch as we hope
herafter **The kinges Maiestes pleasure** is that you shal vieu his graces howse

> Kinges Maieste whose graciouse **pleasure** is that with all celerite ther shalbe the worst may be provided for, **his graces pleasure** is that it shalbe diligently forseen vnto youe, whiche **myn advise** and counsail is that youe shall in any wise ensue, that my poure and frendelie **aduise** is that his grace shall liberally wryte to **the kinges maiestes will** is that ye shal make the shortiest abode there ye can his gracious pleasure and **commandment** is that forasmoch as we hope chiefly myn advise and **counsail** is that youe shall in any wise ensue,

As shown in (13), Pepys shows a strong tendency to use his favoured linguistic nouns *news*, *talk*, and *discourse* in the schema *all the* N ($Adv_{temporal}$) *is that* [...]. The meaning associated with this schema can roughly be glossed as 'rumors are that'. While the schema typical of Richardson (see 14) also includes the quantifier *all*, this is followed by the first-person possessive pronoun and an emotive noun. Richardson's use of this schema is associated with passages where his female protagonists talk or write about their emotions. Finally, Cromwell's use as shown in (15) is marked by the genitive *King's* preceding the directive illocutionary noun *commandment*, the volitive noun *will* or, most frequently, the emotive noun *pleasure*. Overall, the pronounced differences support the impression that the three authors rely on different variants of the constructional schema.

In terms of the framework described in Section 3, the repetition of such semi-fixed sequences of content and function words in and around the target construction can be interpreted as a result of the routinization of syntagmatic associations. Such usage patterns are particularly prominent in the writing of authors of types 2 and 3, who seem to have a strongly represented semantically specific sub-schema of the more general construction. These author-specific usage patterns provide an insight into the individual cognitive processes that contribute to the emergence of chunks, prefabs, and usage patterns on the macro-level of the speech community (cf. Bybee 2010: 33–56). In all three cases we can observe a combination of fixed chunks with some paradigmatically motivated variability. This would mean that alongside chunks the authors have access to a strongly entrenched and highly productive low-level schema, licencing variation in the nominal slot in spite of the strong syntagmatic association.

6.5 Functions, Cotexts, and Context: Pragmatic Associations

The concordance lines in the previous section provide excellent examples of the way in which the functions of uses of the 'N+BE+*that*-construction' can motivate the emergence of usage patterns and schemas. Pepys' pattern is supported by the reporting function ('here is what people say'), Richardson's by the expressive function ('here is how I feel'), and Cromwell's by realizing the directive function

by means of indirect illocutionary acts: *the kinges pleasure* or *his graces pleasure* can be glossed as 'this is what the King wants you to do'.

Differences with regard to preferences for certain functions can not only be observed in these usage patterns, however, but basically for all authors. These in turn seem to be influenced quite strongly by genres. Table 10 renders the distribution across functions for all 10 authors who produced 40 or more tokens of the construction.

Table 10: Author-specific distribution of functions (only authors with n > 40 are included)

	Reporting	Argumentative	Explanatory	Expressive	Directive	Opinative	Total
Fiction							
Richardson	26%	2%	12%	53%		7%	
Austen	44%	21%	17%	15%	1%	1%	100%
Scott	59%	20%	10%	12%			100%
Dickens	44%	14%	17%	13%	2%	9%	100%
Thackeray	42%	40%	12%	2%	2%	2%	100%
Eliot	47%	13%	20%	13%	2%	5%	100%
Expository Writing							
Johnson	12%	40%	38%	8%		2%	100%
Burke	26%	29%	28%	6%	2%	9%	100%
Paine	23%	27%	35%			15%	100%
History							
Macaulay	66%		32%			2%	100%

As the numbers for some intersections of AUTHOR x FUNCTION are very low, we do not see much sense in subjecting these data to inferential statistical procedures. A number of trends and outliers can be observed, however: not only the three authors of expository texts, Johnson, Paine, and Burke, but also Thackeray, who wrote fiction, share a relatively large proportion of uses serving the argumentative function. Not surprisingly, Macaulay, as the author of a history of England, stands out with the largest proportion of reporting uses, but Scott comes quite close to his score. Johnson, Burke, Paine, and Macaulay differ from the fiction authors with regard to the comparatively high proportion of explanatory uses. So preferences for certain functions can extend across genre boundaries. The reverse is also true: authors producing texts of the same genre do not necessarily share the

same functional preferences. For example, a large proportion of Richardson's uses, as we have seen, is motivated by the expressive function, which in turn is, however, almost irrelevant for Thackeray.

We interpret all these differences and tendencies in terms of pragmatic associations which link genres, reasons for the production of a text, and functions of utterances to the activation of the construction. Topics are likely to play a role as well. Very simplistically speaking, when Pepys sits down to make an entry in his diary, and when he is about to record the rumours of the day, then the linguistic sub-schema of the 'N+BE+*that*-construction' shows a particularly high activation potential and comes to his mind (see the concordance lines collected in (13) above). It is certainly remarkable that in as many as four of the concordance lines quoted in (13), his favourite expression stands at the very beginning of the entry for a fresh day. Likewise, when Richardson puts himself in the mind of his protagonists and plans to portray their emotions, utterances of the type *my fear is* or *my concern is* (see (14)) are readily activated.

That pragmatic associations influence the activation potential of sub-schemas of entrenched constructions can also be demonstrated by comparing the usage profiles of one author while writing different types of texts. Jane Austen can help to show this since our database includes both her novels and letters. Given that the corpus of Jane Austen's novels is of course much larger than that represented by her few remaining letters, it is not very surprising that 15 nouns that occur in her novels are not found in her letters. However, that the letters include 12 nouns used in the construction which are not found in the novels, namely *account, anxiety, history, intention, meaning, message, occasion, opinion, reason, report, substance*, and *wish*, is quite remarkable. It seems that the differences in terms of style, envisaged addressees, writing conditions, and topics activate nouns that do not make it to the surface in her fiction.

To wrap up this section, we think it is justified to conclude that individual differences between authors are much more pronounced than expected, not only between those writing different types of texts and/or writing in different periods, as could be expected, but also between authors that are very similar with regard to these parameters. The different parts of this section have demonstrated how these findings can be interpreted within the part of the EC-Model devoted to entrenchment processes. Implications of the findings for the link between entrenchment on the micro-level of individuals and conventionalization on the macro-level of the speech community will be discussed in the next section.

7 The Link between Entrenchment and Conventionalization

To begin with, we would like to come back to the observation that Pepys contributes the first attestation of the noun *fact*. What is remarkable from the point of conventionalization is that the next occurrence of *fact* in our database is found as many as 73 years later in the work of Fielding. From this point onwards we find a massive increase in the use of *fact*. For the 50 years following Fielding's use in 1742 we have as many as 27 attestations of *fact* in our target construction. Of course it is likely that the noun was in use in the intermitting period, too; but if it happened to be true that it was not, then the reason why Pepys' innovation did not catch on at first could be that his diary was of course secret. Following the noun's first use in fiction by Fielding, its frequency of occurrence begins to rise. The fate of Pepys' favourite nouns *news, talk* and *discourse* support this idea. *Talk* and *discourse* are only used in the construction by Pepys, at least in our database. They remain hapaxes. As regards *news*, the first attestation of this noun in our data comes from Shakespeare's *Richard III* (1592). Shakespeare himself uses it a second time in *The Tempest* (1611). Overall, the noun is used 20 times between 1592 and 1866, with Pepys accounting for as many as 7 tokens. Neither *news*, nor for that matter *talk* and *discourse*, are exactly success stories. While these nouns were closely associated with the construction in Pepys' mind, there was no effect on the speech community, presumably because his secret diary could not contribute to the diffusion of these usage patterns in the speech community.

This observation on Pepys' favourite nouns belongs to the realm of speculative anecdotal evidence, of course. Unfortunately, even the 1,588 data points that we have collected are not sufficient for sound quantitative investigations. Only if the major writers of nineteenth-century fiction are lumped together, an opportunity, though a poor one, arises. Even if we focus on the most productive users of the construction, we are left with nouns which are used highly frequently by some authors and not at all by others. Keeping these reservations in mind we can interpret the data presented in Tables 11 and 12. Table 11 ranks the nouns that are used most frequently by five nineteenth-century fiction writers.

Table 11: Comparison of frequencies of top-ranking collocates of 19th-century fiction authors (only authors with sufficiently large numbers of tokens)

	Austen	Dickens	Eliot	Scott	Thackeray	Total
truth	10	9	1	9	12	41
fact	3	13	3	4	10	33
consequence	5	8	3	1	2	19
result	3	8	2	2	0	15
thing	1	9	2	1	0	13
opinion	0	8	1	0	1	10

Table 12 gives the results of pairwise correlations (Kendall's tau). Regarding the comparison between the five authors, only the correlations between Austen and Scott (0.64), Austen and Thackeray (0.50), and Dickens and Scott (0.48) are quite high, while all the others are low. Yet three out of the five authors – Austen, Scott, and Thackeray – show a strong correlation with the collective (consisting of the data of the other four authors), namely 0.78, 0.83, and 0.69 respectively.

Table 12: Kendall's tau: rank correlations among authors and between authors and collective

	Austen (Fiction Only)	Dickens (Fiction Only)	Eliot	Scott	Thackeray	Sum Minus Target Author
Austen		0.16	0.23	0.64	0.50	0.78
Dickens	0.16		0.17	0.48	0.32	0.39
Eliot	0.23	0.17		0.08	0	0.30
Scott	0.64	0.48	0.08		0.50	0.83
Thackeray	0.50	0.32	0	0.50		0.69
Sum Minus Target Author	0.78	0.39	0.30	0.83	0.69	

This indicates that what Zimmerer, Cowell and Varley (2011) describe in the following quote could very well be quite common instead of being only a worst-case scenario: "in the worst case, group data describe a behavioral pattern that does not occur within a single individual" (Zimmerer, Cowell and Varley 2011: 492; cf. also Barlow 2013: 444). While this insight is to some extent predicted by the EC-Model, it suggests a new research agenda which requires a greater focus on the usage preferences and habits of individual authors and speakers and, more

importantly, on the precise ways in which these habits conspire to structure and change grammar.

8 Conclusion

We have tried to show that historical corpus data can be used to investigate individual differences and that these differences can be interpreted in terms of different degrees and types of entrenchment. To come back to the question raised in Section 2, in light of his usage profile and in light of the differences to other authors, it seems likely that Pepys' utterance in (1) was licenced by the activation of a lexically-filled chunk resulting from repeated usage under similar pragmatic conditions. As this chunk still has internal variability – recall that analogous phrases such as *all the discourse is that* [...] are also attested – we can assume that the paradigmatic associations between the nouns in the nominal slot are not totally lost and that a sub-schema associating linguistic nouns to the constructional schema is also entrenched. While the usage profiles of other authors differ substantially, it was possible to classify them according to a number of parameters and allocate them to four different types. The comparison of the profiles of authors with similar characteristics with regard to period and genre suggested that there are strong inter-individual but less strong individual-to-collective differences. Although with 1,588 attestations, the database is quite substantial, it turned out that more data will be needed to investigate the nature of the links between individual authors and the collective and to show how the linguistic behaviour of authors affects change on the macro-level.

We hope that the separation between entrenchment and conventionalization processes has contributed to demonstrating the ultimate sources of many of the sub-processes traditionally regarded as being involved in language change. Analogy (cf., e.g., Hock 2003; Fischer 2011; Bybee 2015: ch. 5), as we have shown, begins to take shape in the minds of individual speakers on the basis of paradigmatic associations linking elements competing for occurrence in a syntagmatic sequence or a pragmatic frame. Analogy is thus first and foremost a cognitive process, and neither a linguistic nor a social one (Fischer 2011: 40–42). Likewise, chunking (Bybee 2010: 33–56, 2015: 238–239) starts out in individual minds rather than in speech communities and relies on the routinization of syntagmatic associations, often supported by pragmatic associations. The process of schematization, too, is genuinely cognitive in nature: since communication is the exchange of utterances and not the exchange of constructions, rules, or syntactic structures, the abstraction required for the formation of a schema must take place in individual minds (cf. Schmid 2015). Whether or not different speakers process the same

utterance by activating the 'same' schema, whatever this could mean, is a different story. The nucleus of change by context-induced interpretations (Heine, Claudi and Hünnemeyer 1991: ch. 3) and invited inference (Traugott and Dasher 2004: 34–41) resides in the pragmatic associations carried over from multiple communicative events involving indirect interpretations, resulting in a consolidation of pragmatic associations into symbolic associations. Constructional changes triggered by newly incoming collocates have their source in the output of individual speakers. Whether these innovations catch on and spread in the speech community is again an entirely different issue on the level of conventionalization, as the large number of hapaxes in our database has shown. We consider it unlikely that other classic types of language change, e.g. bleaching or pragmaticalization, take place within the lifetime and linguistic history of individual speakers. Our impression is that for those processes to happen, a more long-term interaction between entrenchment and conventionalization processes seems to be required. Further research is required to address this issue.

Finally, should these findings and their interpretation be of interest to anybody who studies linguistic structures and language change proper, i.e. on the level of the linguistic system? One possible answer would be a clear no: the task of linguists is to search for the principles and rules that underlie grammar and the way in which they change over time; the performance of individual speakers has no role to play in this quest. If one agrees with this view, then one could still take home the message that beneath the surface of idealized speaker-hearers, individual variation is rampant. If one subscribes to a usage-based approach which assumes that grammars emerge from usage-events, then one would be well advised to take these insights more seriously, since from this perspective the usage patterns of individual speakers under specific pragmatic circumstances are likely to have an effect on grammar and the way in which it changes. If one believes that usage is at least co-determined by entrenchment, and that usage can therefore be used as a diagnostic of degrees and types of entrenchment, then one can read our paper as a methodological model for more research into individual differences using data from historical corpora.

Works Cited

Corpora

Austen = Chapman, Robert William (ed.). 1881–1960. *Jane Austen's Letters to her Sister Cassandra and Others*. 2nd ed. 1952. London/New York: Oxford University Press. <http://ota.ox.ac.uk/desc/1520>.

Project Gutenberg = <https://www.gutenberg.org>.
Helsinki = *The Helsinki Corpus of English Texts*. 1991. Compiled by Matti Rissanen (Project leader); Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English). Department of Modern Languages, Helsinki: University of Helsinki. <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus>.
OED3 = *Oxford English Dictionary Online*. 2000–. 3rd ed. online. Oxford: Oxford University Press. <http://www.oed.com>.
Old Bailey = Huber, Magnus, Magnus Nissel, Patrick Maiwald and Bianca Widlitzki (eds.). 2012. *The Old Bailey Corpus. Spoken English in the 18th and 19th Centuries*. Giessen: Giessen University. <www.uni-giessen.de/oldbaileycorpus>.
Paston = Gordon, Eric Valentine (ed.). 1953. *Paston Letters and Papers of the Fifteenth Century*. Oxford: Clarendon. <http://ota.ox.ac.uk/desc/1685>.
PCEEC = *Parsed Corpus of Early English Correspondence*, parsed version. 2006. Annotated by Ann Taylor, Arja Nurmi, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. <http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/pceec.html>.
Shakespeare = <http://www.opensourceshakespeare.org>.

Secondary Sources

Barlow, Michael. 2013. "Individual Differences and Usage-Based Grammar". *International Journal of Corpus Linguistics* 18: 443–478.
Bybee, Joan. 2003. "Mechanisms of Change in Grammaticalization: The Role of Frequency". In: Brian D. Joseph and Richard D. Janda (eds.). *Handbook of Historical Linguistics*. Oxford: Blackwell. 602–623.
Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
Bybee, Joan. 2015. *Language Change*. Cambridge: Cambridge University Press.
Cannon, Christopher. 2004. "Chaucer's Style". In: Piero Boitani and Jill Mann (eds.). *The Cambridge Companion to Chaucer*. Cambridge: Cambridge University Press. 233–250.
Craig, Hugh. 1992. "Authorial Styles and the Frequencies of Very Common Words: Jonson, Shakespeare, and the Additions to *The Spanish Tragedy*". *Style* 26: 199–220.
Craig, Hugh. 2002. "Common-Words Frequencies, Shakespeare's Style, and the Elegy by W. S.". *Early Modern Literary Studies: A Journal of Sixteenth- and Seventeenth-Century English Literature* 8: 1–42.
Craig, Hugh. 2011. "Shakespeare's Vocabulary: Myth and Reality". *Shakespeare Quarterly* 62: 53–74.
Diessel, Holger. 2007. "Frequency Effects in Language Acquisition, Language Use, and Diachronic Change". *New Ideas in Psychology* 25: 108–127.
Finegan, Eduard and Douglas Biber. 1995. "*That* and Zero Complementizers in Late Modern English: Exploring ARCHER from 1650–1990". In: Bas Aarts and Charles F. Meyer (eds.). *The Verb in Contemporary English*. Cambridge: Cambridge University Press. 241–257.
Fischer, Olga. 2011. "Grammaticalization as Analogically Driven Change". In: Heiko Narrog and Bernd Heine (eds.). *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press. 31–42.

Gries, Stefan and Martin Hilpert. 2012. "Variability-Based Neighbor Clustering: A Bottom-Up Approach to Periodization in Historical Linguistics". In: Terttu Nevalainen and Elizabeth Closs Traugott (eds.). *The Oxford Handbook on the History of English*. Oxford: Oxford University Press. 134–144.

Givón, Talmy. 1990. *Syntax: A Functional-Typological Introduction*. Vol. 2. Amsterdam: Benjamins.

Heine, Bernd, Ulrike Claudi and Friederike Hünnemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago, IL: University of Chicago Press.

Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.

Hock, Hans Henrich. 2003. "Analogical Change". In: Brian D. Joseph and Richard D. Janda (eds.). *Handbook of Historical Linguistics*. Oxford: Blackwell. 441–460.

Horobin, Simon. 2007. *Chaucer's Language*. Basingstoke: Palgrave Macmillan.

Huber, Magnus. 2007. "The Old Bailey Proceedings, 1674–1834: Evaluating and Annotating a Corpus of 18th- and 19th-Century Spoken English". In: Anneli Meurman-Solin and Arja Nurmi (eds.). *Annotating Variation and Change*. Helsinki: VARIENG. <http://www.helsinki.fi/varieng/series/volumes/01/huber/>.

Krug, Manfred. 2000. *Emerging English Modals: A Corpus-Based Study of Grammaticalization*. Berlin: Mouton de Gruyter.

Janda, Richard D. and Brian D. Joseph. 2003. "On Language, Change, and Language Change – or, of History, Linguistics, and Historical Linguistics". In: Brian D. Joseph and Richard D. Janda (eds.). *Handbook of Historical Linguistics*. Oxford: Blackwell. 3–180.

Langacker, Ronald W. 1987. "Nouns and Verbs". *Language* 63: 53–94.

Mantlik, Annette. 2011. *The Historical Development of Shell Nouns: A Diachronic Study of Abstract Noun Constructions in English*. Unpubl. PhD dissertation, Munich University.

Mantlik, Annette. 2013. "An Etymological Analysis of Shell Nouns". In: Hans Sauer and Gaby Waxenberger (eds.). *Recording English, Researching English, Transforming English*. Frankfurt am Main: Lang. 133–161.

Mantlik, Annette and Hans-Jörg Schmid. forthcoming. "*That*-Complementizer Omission in N+BE+*that*-Clauses – Register Variation or Constructional Change?". In: Alex Ho-Cheong Leung and Wim van der Wurff (eds.). *The Noun Phrase in English: Past and Present*. Amsterdam: Benjamins.

Raumolin-Brunberg, Helena and Arja Nurmi. 2011. "Grammaticalization and Language Change in the Individual". In: Heiko Narrog and Bernd Heine (eds.). *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press. 251–262.

Rissanen, Matti. 1991. "On the History of *that*/Zero as Clause Object Links in English". In: Karin Aijmer and Bengt Altenberg (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman. 272–289.

Rohdenburg, Günter. 2013. "Using the *OED* Quotations Database as a Diachronic Corpus". In: Manfred Krug and Julia Schlüter (eds.). *Research Methods in Language Change and Variation*. Cambridge: Cambridge University Press. 136–157.

Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin/New York: Mouton de Gruyter.

Schmid, Hans-Jörg. 2013. "Is Usage more than Usage After All? The Case of English *not that*". *Linguistics* 51: 75–116.

Schmid, Hans-Jörg. 2014a. "Lexico-Grammatical Patterns, Pragmatic Associations and Discourse Frequency". In: Thomas Herbst, Hans-Jörg Schmid and Susen Faulhaber (eds.). *Constructions – Collocations – Patterns*. Berlin: Mouton de Gruyter. 239–293.

Schmid, Hans-Jörg. 2014b. "A Framework for Understanding Linguistic Entrenchment and its Psychological Foundations in Memory and Automatization". <http://www.anglistik.uni-muenchen.de/personen/professoren/schmid/schmid_publ/entrenchment-intro.pdf>; accessed 30 April 2015. [To appear in: Hans-Jörg Schmid (ed.). *Entrenchment, Memory and Automaticity: The Psychology of Linguistic Knowledge and Language Learning*. Boston, MA: APA and Walter de Gruyter.]

Schmid, Hans-Jörg. 2015. "A Blueprint of the Entrenchment-and-Conventionalization Model". *Yearbook of the German Cognitive Linguistic Association* 3: 5–27.

Taavitsainen, Irma. 1995. "Narrative Patterns of Affect in Four Genres of the Canterbury Tales". *The Chaucer Review: A Journal of Medieval Studies and Literary Criticism* 30: 191–210.

Traugott, Elizabeth Closs and Richard Dasher. 2004. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs and Graeme Trousdale. 2013. *Constructionalization and Constructional Change*. Oxford: Oxford University Press.

Zimmerer, Vitor, Patricia Cowell and Rosemary Varley. 2011. "Individual Behavior in Learning of an Artificial Grammar". *Memory and Cognition* 39: 491–501.