Traces of Lexical Priming in English and German Uses of BE and HAVE Forms in Biographies

Michael Pace-Sigge[1]

Abstract

According to Sinclair, features where two varieties of English may systematically differ include: **collocation**, **semantic preference, colligation** and **lexically-driven grammatical patterns**. O'Donnell (2006), however, has highlighted recently the fact that corpus linguistic research is very Anglo-centric, and it has yet to be shown that these categories apply with equal effectiveness in German. This work would seek to supply this lack. More precisely, my objective is to see how far a related, yet structurally different language like German **can also provide existence of lexical priming**, a theory proposed by Hoey (2005) that makes use of the above categories and further categories. Using two small test corpora I investigate the lexical use *BE* and *HAVE (SEIN and HABEN)* forms, drawing on the features employed in lexical priming. One objective is to demonstrate that priming is determined not just by the specific language but also by genre– in this case, *biography*. The claim made is that speakers in different language communities use their lexicon in a way that makes the way they have been primed visible. In this paper I will focus on high frequency verb forms and compare words in their present tense and simple past uses and will report on the ways in which these items can be similar or different in these two Indo-Germanic languages. The focus shall be on how far the genre *biography* is primed to employ *be* and *have* and whether this genre-priming runs in parallel with language priming.

1. Introduction

John Sinclair pointed out that lexical choices are not free, are not fillers. Words are not just the bricks set into the steel structure of a modern building. His starting point was the English language. This paper tries to find application for some of his insights in English as well as in German.

To quickly re-iterate some points raised by Sinclair that will form the theoretical background of this research:

⇒ Many uses of words and phrases attract other words in strong collocation.
⇒ Many uses of words and phrases show a tendency to co-occur with certain grammatical choices.
⇒ Many uses of words and phrases show a tendency to occur in a certain semantic environment.

(Sinclair, 1991, p.112)

[1] University of Liverpool, UK
 e-mail: scouse@liv.ac.uk

These concepts have been developed since and Michael Hoey bases his theory of lexical priming (LP) on these properties of words. It goes beyond the mere preference or non-preference of words to each other:

> Every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word. If one of the effects of the initial priming is that regular word sequences are constructed, these are in turn primed.
>
> (Hoey, 2005, p.13)

This individual could be any language user, regardless of the language employed. It can cover the written text (as in Hoey's book, where he looked at *Guardian* newspaper texts) or spoken utterances (Pace-Sigge, forthcoming).

In 2006 Matthew O'Donnell queried in how far corpus linguistics can be seen as an Anglo-centric project. He highlighted how many of the corpus linguists are either from Britain or based in the UK. More importantly, much of the research is centred around the use of the English Language. Therefore, would categories that work so well in English work also in German? While O'Donnell proved that concepts and techniques can be applied even when researching very different languages, this paper attempts to support the view that the *Anglo* might be the initiator and inspirer of this research, yet is not the *centric* point.

This paper has in its title "traces" as it looks at first, rudimentary evidence of comparable priming in two languages.

As a caveat it needs to be said, however, that English and German are closely related languages, being used by people of closely related cultures.

For this paper I have chosen to look at a genre that is reasonably tightly defined. This genre is *biographies*, sets of texts that also have the virtue to be common both in German and English and therefore suitable for comparison. Amazingly, I could not find any research on language in biographies. *Wikipedia* in its style manual says:

> Biographies of living persons should begin in the present tense; biographies of deceased persons should begin in the past tense.

Similary, a textbook to teach German to non-native speakers:

> Das Präteritum wird vor allem im schriftlichen Deutsch verwendet, wenn man über etwas berichtet oder erzählt (Biographie, Märchen, Erzählung usw.)*

For this paper, I created two sub-corpora from larger, well-known corpora. Because of fairly restrictive copyright rules for German-published texts, I have to rely on older writings, which can be downloaded for free in an appropriate format from the Gutenberg Project.

The corpus of comparison is the BNC. As these two are very different corpora, I created two sub-corpora that are closer in their linguistic scope. One subsection of the BNC Written is made up out of biographies. Consequently, the sub-corpus from Gutenberg German texts consists of biographical writings only as well – albeit smaller in size – see Table 1.

	Ger-Biographie (Gbio)	Eng-Biography (Bbio)
Tokens used for word list	**648.817**	**3.489.804**
Ratio:	1:	5.4
autobiographies	**7**	**29**
biographies	**3**	**69**
Ratio:	7:3	3 : 7

Table 1.

Biography, in language terms, is a highly restricted form. Equally, this can be expected to be a formal constriction within biography-writing throughout the Western world. It presents itself therefore as suitable for making some comparisons that trace the priming in this text form.

2 Lexical Priming

2.1 Lexical Priming Hypotheses

Following John Sinclair's work, Michael Hoey gives the following relevant criteria for what makes Lexical Priming (LP):

- Every word is primed to occur with other words; these are collocates
- Every word is primed to occur with particular sets; these are its semantic associations
- Every word is primed to occur in association with particular pragmatic function; these are the pragmatic associations
- Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations
- Every word is primed for use in one or more grammatical roles; these are its grammatical categories

(Hoey, 2005, p.13)

2.2 Lexical Priming Genre-constrained hypotheses

Looking at *Biography*, a defined genre, it makes therefore sense to see what Hoey states with regard genre and LP:

- All these claims are in the first place *constrained* by genre.
- Words are primed in a range of social contexts of which account is taken.
- Certain kinds of features will only become apparent when one looks at more specialised data.

(Hoey, 2005: p.13)

This *specialised data* are my two *biography* corpora.

3. Findings for Primary Verbs

Going through the wordlists of both corpora, the forms of primary verbs BE (SEIN) and HAVE (HABEN) are amongst the most frequently occurring terms. Even more

importantly for this research, they appear as the highest-occurring verbs in either. They are also, as irregular verb-forms, easily identifiable as in their conjugational function.

Table 2 below gives the final forms discussed in the paper. However, as some findings are less conclusive, I will, for questions of space, concentrate on the most important findings in the framework of this paper only.

In the lists of most frequent items in both corpora, the primary verbs BE and HAVE stand out. This enables a straightforward comparison, based on the similar roots and functions of this term in English and German.

> As a main verb *be* – the **copula** – is the most important **copular verb** in English. It links the subject noun phrase with the subject predicative or an obligatory adverbial.
> (…)
> Surprisingly, the copula *be* differs from most lexical verbs because it is much more frequent in academic prose than in conversation, newspapers or fiction. (…) *be* occurs most commonly with a noun phrase as subject predicative. In these structures, *be* has two common functions: to characterise the subject noun phrase in some way, or to identify the subject noun phrase.
>
> (Biber *et al.* 2002: pp. 104 and 141)

The item BE appears in its copula function in the following discussion. What Biber *et al.* state for BE is also true for German *SEIN* (though I have no confirmation regarding frequency patterns for other genres).

Biber *et al.* make also clear that HAVE has functions beyond the possessive –

> (…) As a transitive main verb, *have* is as common as the most common lexical verbs in English. *Have* is most common in conversation and least common in academic prose. Within academic prose though, *have* is more common than any lexical verb.
> (Biber *et al.* 2002: pp. 136)

Given that biography has elements both of academic prose and of (recorded) conversation, its frequent use in both corpora under discussion does not surprise here.

3.1 The final forms

Table 2 deals with the highest occurring verb forms in both biography corpora. One difference between the languages becomes immediately obvious. Apart from the use of present tense forms *have – hatte(n)* and *are - sind*, the proportional occurrence frequency of each feature in both languages in the two corpora is more divergent. The proportional occurrence of all other forms are two – to three times more frequent in the English Language. This obvious friction will be focussed upon in this paper.

Eng	%	Deu	%	Tense form
WAS 65,476 occ.	1.84	WAR 4,763 occ.	0.73	1st/3rd pers. sg. simple past *be*
WERE 17,023 occ.	0.48	WARST *// WAREN 1,042 occ.	// 0.15	simple past 2nd pers. sing // pl. *be*
AM 1,127 occ.	0.03	BIN 428 occ.	0.07	present tense. 1st pers. *be*
IS 14,502 occ.	0.41	IST 3,533 occ.	0.54	Present tense 3rd pers. sing. *be*
ARE 5,134 occ.	0.14	BIST// SIND 122 // 1,016	0.02 // 0.16	Pres. Tense: 2nd. pers. sing// pl. *be*
BEEN	0.26	GEWESEN	0.06	past participle *be*
HAS 3,360 occ.	0.09	HAT 1,090 occ.	0.17	3rd pers. sing. prs. ts. *have*
HAVE 11,059 occ.	0.31	Habe (1090)/hast* //haben (1010)	0.17/ 0.02 //0.15	present tense *have* other
HAD 29,375 occ.	0.83	HATTE / / HATTEN 3112//554	0.39// 0.09	simple past *have*

Table 2.   *too few uses

3.2 To BE or not to BE?

Section 3.2 will look at the ways forms of BE co-occur in their clusters and colligations in both the BNC Biography (Bbio) and the German Biography (Gbio) corpora.

A conscious decision has been made to exclude the use of the participle in its entity. In both corpora, its use is marginal. Though this renders the PP still capable of collocational relevance, it is outside the limited scope of this initial study.

3.2.1 WAS / WAR

WAS and WAR are very specific verbs. These two forms of *be* indicate a clearly specified conjugation form: it is used, in both cases, for the first and also the third persons singular in the simple past only. As mentioned above, such congruence is the exception, rather then the norm when these two languages are compared.

There are 64,698 occurrences of WAS and 4173 of WAR in the respective corpora. Looking at the respective proportional percentages of the total words used, WAS occurs more or less exactly 2.5 times as often. (1.84% to 0.73% of the total word tokens).

It is this word association, the way these past tense markers string into clusters that really reveal how clearly this term follows the same pattern in its use. When put to direct comparison, (auto-)biographies appear to use the same 2-word clusters in a

regular sequence – and the proportional percentages of use overlap to a strong degree. This is shown by Table 3 below. That the same clusters appear so frequent in the genre biography indicates priming at work here.

CLUSTER Gbib	FREQ.	%	CLUSTER Bbib	FREQ.	%
ES WAR	453	10.9	IT WAS	9106	14
ER WAR	235	5.6	HE WAS	10171	15.7
WAR ES	215	5.2	WAS IT	174	.2
WAR EIN / WAR EINE	200 /121 (321)	7.7	WAS A	5991	9.2
ICH WAR	212	5.1	I WAS	4952	7.6
DAS WAR	184	4.4	THAT WAS	1174	1.8
WAR DIE / WAR DER / WAR DAS	177 / 175 /126 (478)	11.5	WAS THE	3426	5.2
WAR NICHT	98	2.3	WAS NOT	2130	3.3
SIE WAR	89	2.1	SHE WAS	2473	3.8
WAR AUCH	83	2.0	WAS ALSO	787	1.2
UND WAR	72	1.8	AND WAS	1665	2.4

Table 3.

First of all, it needs to be pointed out that the bulk of the 2-word WAS/WAR clusters in Table 3 above are taken from the 12 most-frequent clusters in both corpora. Exceptions are those clusters where a direct comparison seemed appropriate (WAR ES – WAS IT).

One obvious difference in use is the use of WAR ES / WAS IT, where the latter is extremely rare. It does not form part of longer clusters in English, while in German there are the phrases SO WAR ES and DAS WAR ES.

The use of HE WAS is more (proportionally) frequent than ER WAR. This reflects the greater proportion of biographies in Bbio (or the greater proportion of autobiographies in Gbio respectively).

Another – proportional – difference can be found in the use of indefinite and definite articles. While in English the indefinite article is slightly more prominent, the definite article is extremely frequent in German. This language characteristic has been observed previously.

Still: A clear structure emerges.

Taken together, all determiners (a / an / the / it – ein/e / der-die-das / es) cluster together with WAS and WAR at about the same rate.

Most biographies in either corpus deal with biographies of a male subject – hence ER WAR / HE WAS. This is followed by autobiographies or quotes – ICH WAR / I WAS. Female subjects are clearly prominent – but not as frequent – SIE WAR / SHE WAS.

Even more similar in proportional use are the overall second-most frequent 2-word clusters: ES WAR / IT WAS. A brief look at the concordance suffices to ascertain that this usual starts a sentence or clause – in both corpora. The same is true of the connector phrases WAR AUCH / WAS ALSO and UND WAR / AND WAS. These occur both at proportionally very similar frequencies and are still quite prominent uses of the terms WAR / WAS.

Intermediate conclusion: there are strong reasons to conclude that a certain genre as such determines specific use of key terms. This is even true just on the basis of 2-word clusters. This paper demonstrates that this is regardless of language.

Longer clusters do not appear with WAS / WAR – it can combine with such an array of options that even the most frequent 4-word clusters only amount to 0.1 - 0.2 % of its total usage. However, looking at the extensions of the above 2-word clusters and the related 3-word clusters, the similarity in use is as striking:

It becomes clear how similar the distribution is in both languages. Three high-occurring clusters at the top, then a flattening of the curve as the clusters further down the list appear with similar occurrence-rates.

Even more striking, however, is the overlap in the phrases occurring in proportional frequency in both languages. A direct comparison can be found below:

German Biographie 3-word clusters	English BNC Biography 3-word clusters
ES WAR EIN/E 102 - 2.4%	IT WAS A -1101- 1.7%
ES WAR DER – 16 - 0.38%	IT WAS THE – 583 - 0.9%
UND ES WAR – 24 – 0.6%	AND IT WAS – 450 – 0.7%
UND DAS WAR – 0.34%	AND THAT WAS – 115 – 0.2%
UND ICH WAR – 14 - 0.34%	AND I WAS – 332 – 0.5%
ABER ES WAR – 18 - 0.43%	BUT IT WAS – 425 - 0.6%
ES WAR NICHT – 11 – 0.24%	IT WAS NOT – 432 – 0.6%
ABER ES WAR – 18 – 0.43%	BUT IT WAS – 425 – 0.6%
ER WAR EIN – 0.8%	HE WAS A – 876 – 1.4%
ER WAR AUCH – 8 - 0.18%	HE WAS ALSO - 230 - 0.31%
ER WAR NICHT – 8 – 0.18%	HE WAS NOT – 223 – 0.31%
DAS WAR DIE / DER – 18 – 0.44%	THAT WAS THE – 136 – 0.22%
DA WAR DER – 5 – 0.11%	THERE WAS THE – 108 – 0.15%
SIE WAR EIN/E – 14 – 0.34%	SHE WAS A – 181 – 0.25%
WAR AUCH EIN – 11 – 0.24%	WAS ALSO A – 145 – 0.3%
DA WAR AUCH – 6 – 1.6%	THERE WAS ALSO – 88 – 0.12%

Table 4.

When the all the possible 3-word clusters that are formed from the above discussed 2-word clusters in the German corpus are compared with co-occurrences in the BNC corpus, the extent of congruence is telling.

Table 4 shows that there is one clear difference between 2-word and 3-word cluster structural use. While Table 3 describes the use of WAS THE as less than half as often occurring as its German equivalents, this is promptly reversed in the most occurring 3-word cluster *IT* WAS THE (*ES* WAR DER).

Yet the finding that is by far the most important one is this. Despite being a different language, despite being far more inflected, despite perceived as using different word order: the three word-clusters co-occur in content and proportional use as equivalents.

All of this just confirms the intermediate conclusion. As far as the simple past form of *be* – WAS / WAR is concerned, there are strong reasons to conclude that genre as such determines specific use of key terms. At least in the specific text form of *biography*, this paper demonstrates that this is regardless of language.

3.2.2 WERE / WAREN

The plural simple past forms of BE reflect the findings of 3.2.1 – yet, as Table 5 indicates, the proportional use of this term is more widely divergent.

However, WERE can also be used as the 2nd person singular – WARST. In both corpora the plural form (THEY WERE; WE WERE etc) is by far the most used. Accordingly, there is a relatively high occurrence of WAREN in German but only very few DU WARST (YOU WERE) – the singular form.

Rank	2w cluster	Occ.	%	Rank	2w cluster	Occ	%
1	They Were	2257	13.3	3	SIE WAREN	52	5.0
2	We were	1571	9.2	4	WIR WAREN	36	3.5
3	There were	1466	8.6	15	DA WAREN	18	1.8
6	Were not	568	3.4	11	WAREN NICHT	21	2.1
7	Were the	527	3.1	1	WAREN DIE	97	9.8
10	Were in	386	2.3	13	WAREN IN	19	1.8
11	Which were	382	2.3	22	DAS WAREN	22	2.1
	WERE Bbio 2w cluster				WAREN Gbio 2w cluster		

Table 5.

Table 5 shows firstly that, in biographies in both German and English the simple past plural BE is by far the preferred choice of writers compared to the simple past singular BE forms.

Still, WERE / WAREN might not be seen as a perfect examples for congruence of the two corpora. For a start, there is only an extremely limited amount of 3-word clusters in evidence. The closest cluster that direct translates directly either way is AND THERE WERE / UND DAS WAREN which appears in 0.59% and 0.29% of all uses of WERE / WAREN respectively.

Secondly, there is the dominant influence of German language priming (or, conversely, English language priming) that dictates possible and frequent word-order uses that can only occur in one language, but not the other. In the case of WERE / WAREN they frequently trump genre-specific priming.

For example, in German there are -

WAREN ES   34 occ.   3.3%   were it

WAREN UND   32   3.0%   were and

WAREN WIR   28   2.6%   were we

The occurrence figures show that these forms make perfect sense in German. The direct translation into English, however, indicates to an English speaker that such clusters are in rare use.

Conversely, the Bbio corpus has clusters like -

WHO WERE   524 occ.   3.1%

WERE TO   415   2.6%

These are, again, more common in English usage yet rare in German.

3.2.3 IS / IST – third person singular (present) occurrences

Moving on from the use of BE in its past tense forms, IS / IST is amongst the highest occurring items in most corpora – including Bbio / Gbio.

For this paper, 2-word clusters prove of little evidence, yet the 3-word clusters show some genre-specific usage evidence. Suffice to say that he highest occurring 2w clusters are IT IS (20.0%) and ES IST (11.1%) respectively and that the descriptive clusters THIS IS / THAT IS (4.7% / 3.8%) in English Bbio are mirrored by DAS IST (4.8%) in the German Gbio.

Bbio	Cluster 3N	Occ	%	Gbio	Cluster 3N	Occ.	%
	IT IS A/AN	336	2.3	27	ES IST EIN	41	1.2
26	IT IS NOT	192	1.3	74	IST ES NICHT	18	0.5
33	BUT IT IS	162	1.1		ABER ES IST	11	0.3
36	AND IT IS	150	1.0	41	UND ES IST	28	0.79
	IT IS THE	124	0.9		ES IST DER	12	0.3
	THIS IS THE	110	0.8		DAS IST DAS	13	0.4

Table 6.

Table 6 above demonstrates that there is a certain degree of congruence between the two corpora, yet there is little conclusive evidence that this is genre-specific in its priming.

Cluster Gbio	Occ	%	Cluster Bbio	Occ	%
SO IST ES AUCH	5	0.14	IT IS CLEAR THAT	52	0.36
UND DOCH IST ES	4	0.1			
SO VIEL IST GEWIß	4	0.1	THERE IS NO DOUBT	45	0.31
ABER ES IST NICHT	3	0.095	BUT THERE IS NO	32	0.2

Table 7.

Likewise, the 4-word clusters shown in Table 7 show clusters that are not literal translations but similar in meaning. All of these are marginal in their use yet reflect both language-specific priming (in the way words collocate and colligate together to form such clusters) and reflect in a wider sense genre-specific use.

The items IS / IST appear to be used with phrases of confirmation and arguing a case for rather than a case against.

3.2.4 AM / BIN – The power of genre priming

When AM / BIN is checked for proportional occurrence in Table 2, it does not stand out as particularly frequent in its use. Readers may assume that the first person singular (present tense) is fairly dominant in the Gbio corpus, given the higher amount of autobiographical material. Table 2 confirms that BIN appears about twice as often in Gbio than AM in Bbio.

Cluster Bbib	Occ.	%	Cluster Gbib	Occ	%
I AM	1012	90.0	ICH BIN	171	40.0
AM NOT	110	9.8	BIN NICHT	16	3.7
AM SURE	82	7.3	BIN ÜBERZEUGT	13	3.0
AM I	37	3.3	BIN ICH	133	31.1
AM IN	21	2.0	BIN IN	11	2.6
AM AND	17	1.8	BIN UND	17	3.7

Table 8.

Table 8 highlights that in English (at least within the strictures of the given corpus) I AM is the two-word cluster that is primed for near-total use when ever the item AM is used. In fact, it appears 27.2727 times more often than the reverse 2-word cluster AM I. This stands in strong contrast to the usage in Gbio. Gbio has also both clusters. Yet there ICH BIN is definitely less than 50% in use, and both ICH BIN and BIN ICH appear close in their proportional use.

Looking at genre-priming, however, the appearance of a key item is of far greater importance. The term AM / BIN is linked to the item SURE / ÜBERZEUGT in a way that highlights its strict preference. Evidence for this hardens up when the longer clusters are compared:

Cluster Bbib	Occ.	%	Cluster Gbib	Occ	%
I AM NOT	106	9.7	ICH BIN NICHT	15	0.36
I AM SURE	81	7.3	ICH BIN ÜBERZEUGT	11	2.6
AND I AM	60	5.7	UND ICH BIN	14	3.1
I AM A	41	3.6	ICH BIN EIN	5	1.3
BUT I AM	32	3.1	ABER ICH BIN	8	1.6
Bbio 3w AM			Gbio 3w BIN		

Table 9.

I AM SURE / ICH BIN ÜBERZEUGT remains a dominant cluster. This is all the more important when the items SURE *with* AM are looked at in a different context. In the nearly 5 times larger *BNC fiction* corpus (*I*) *AM SURE* appears proportionally less than half as often then in Bbio. This leads me to conclude that this term is primed for use in *biographies*.

In Bbio there appears also a high degree of defending the "I" speaker - I AM NOT / BUT I AM. This marks a contrast to the use of the terms IS / IST which is used more to argue a case in favour of than clusters with AM / BIN.

That this is proportionally far more widely used in Bbio than Gbib highlights cultural differences again, while its strong co-occurrence still makes a good case for genre-specific priming across language barriers.

3.3 Primary Possessive Verb: HAVE / HABEN

Section 3.3 will look how forms of HAVE co-occur in their clusters and colligations in both the BNC Biography (Bbio) and the German Biography (Gbio) corpora. Though HAVE / HABEN is another high-frequency primary verb, it remains to be

seen whether its occurrence patterns follow a similar pattern to BE. Furthermore, it is of interest to see where and when HAVE forms are functioning in parallel to HABEN.

3.3.1 HAD / HATTE(N) discrepancies

Unlike WAS and WAR, HAD / HATTE(N) does not show a clear pattern of concurrence in its use. For a start, most valid clusters are only 2-words short. While Table 10 below highlights that there are co-occurrences exist, these are mostly not matched by their proportional frequencies.

Cluster Bbio	Occ.	%	Cluster Gbio	Occ.	%
HE HAD	5501	18.7	ER HATTE	119	3.8
I HAD	4306	14.7	ICH HATTE	212	6.8
THEY HAD	1409	4.8	SIE HATTEN (pl)	38	6.9
SHE HAD	1216	4.1	SIE HATTE	52	1.8
WE HAD	1133	3.9	WIR HATTEN (pl)	42	7.6
HE HAD A	360	1.2	ER HATTE EINE	9	0.3
AND I HAD	167	0.6	UND ICH HATTE	14	0.45

Table 10.

Table 10 has the percentages for HATTE and HATTEN based on their respective totals. This highlights that, while the use of the singular forms diverge massively in frequency between the Bbio and the Gbio, the plural forms (blue) are in line with other uses shown above.

Likewise, there are co-occuring 3-word clusters (yellow), yet the frequencies are divergent, too. HAVE and HATTE(N) appear to demonstrate language patterning in English or German respectively is the main drive of there priming, subsuming genre-driven priming.

3.3.2 HAVE / HABE(N) – the dominance of the *to* Infinitive.

While HAVE and HABE(N) provide a little more evidence of correlation than the usage in the simple past shown above, it does throw up a peculiar use of language. Neither the Collins Cobuild Grammar Patterns (cf. Hunston *et al.,* 1996), nor Helbig / Buscha nor a third corpus used for comparison highlights any strong use of the TO HAVE / ZU HABEN clusters. It appears to be unique in its use in biographies.

Cluster Bbio	Occ.	%	Cluster Gbio	Occ	%
TO HAVE	1848	16.7	ZU HABEN	180	17.8
I HAVE	1263	11.4	ICH HABE	336	30.9
WE HAVE	384	3.5	WIR HABEN	36	3.6
YOU HAVE	343	3.1	DU HAST	31	25.4
HAVE THE	319	2.9	HABE DIE	34	3.1
			HABE DAS	29	2.7
			HABEN DIE	26	2.4
THEY HAVE	265	2.4	SIE HABEN	44	4.4

Table 11.

To start with, Table 11 demonstrates the idiosyncrasies of the Gbio corpus well. The cluster ICH HABE, using the *narrative I,* pinpoints that the majority of texts are autobiographical. Though a frequent cluster in both corpora, ICH HABE is found nearly three times more often in Gbio. This is in contrast with ICH BIN which is used less than half as often in Gbio than I AM is used in Bbio.

Furthermore a widely acknowledged German-language colligation – the use of the definite article where the three listed occurrences are outnumbering the English form HAVE THE by the factor three.

The one 2-word cluster that stands out, however, is TO HAVE / ZU HABEN. This is most frequent in Bbio and, proportionally, of similar frequency in the Gbio. Remarkable in itself, this becomes a marked feature of the genre *biography* when measured against a comparable corpus. In *BNC fiction* TO HAVE appears only in 10.9% of all uses of HAVE.

While HAVE TO (Biber et al) and HABEN + Infinitiv mit ZU (Helbig / Buscha) are described in their use (with HAVE / HABEN as auxiliary verb / Hilfsverb), it is beyond the scope of this paper to identify all the possible uses occurring here.
Yet an indication to its usage is given by Helbig / Buscha:

> *Haben* + Infinitiv mit *zu* bedeutet oefter eine Notwendigkeit, manchmal aber auch eine Moeglichkeit.
>
> (Helbig/Buscha. 1993. p. 128)

It describes often something of necessity, sometimes something that is a possibility. Given the possible context the genre biography offers, this would be a likely use.

3.3.3 HAS / HAT – the third person singular in the present

These items are rather infrequent in their use in either corpus. As above, there is a clear indication of the biography / autobiography split.

Rank	Cluster	Freq.	%	Rank	Cluster	Freq.	%
2	HE HAS	412	12.3	1	ER HAT	61	5.6
3	IT HAS	279	8.3	14	ES HAT	25	2.3
4	HAS A	214	6.4	20	HAT EINE	14	1.3
				23	HAT EIN	13	1.3
5	SHE HAS	208	6.2	16	SIE HAT	22	2.1
12	HAS THE	95	2.8	5	HAT DIE	43	3.9
				9	HAT DAS	33	3.0
				18	HAT DER	17	1.4
Bbio 2w cluster HAS				**Gbio 2w cluster HAT**			

Table 12.

The 3rd person possessive obviously used more often in biography (Bbio) than autobiographies (Gbio)

Within top 15 2-word clusters, direct equivalents appear in both corpora, yet not one of the clusters appears in equivalent proportional use. Because of the limited usage of these terms, there is no sufficient use of comparable 3-word clusters.

4. Conclusions

I believe that the above discussion gives first conclusive insights that there are traces where language-use appears *primed by genre* as well as language. This confirms Michael Hoey's assertion regarding context / genre-specific priming.

At the same time, there is overwhelming evidence for language-specific priming. Interestingly, while many English learners of German as an L2 point out that they find the different word-order confusing, this paper shows that not all verbs colligate in a way that dictates a word-order in German that is different to the word-order in English. The picture, as the saying goes, is more multi-facetted. (This may be of little consolation to the even more confused learner, however).

The initial points of difference are the differences of proportional usage. Overall, the BE and HAVE forms are appearing less frequently in the Gbio than in the Bbio. This is likely to indicate a wider variety of verb use in the German language – as far as these *biographies* are concerned at least. It is sometimes also down to language use. German does, for example, make much more frequent use of reflexive forms. These appear in the wordlists and some collocations, yet have been let out in the discussion above as there is no likely equivalent in the English-language corpus.

WAS/WAR in both 2-word and 3-word clusters appears broadly similar in its uses in both corpora. Given that it is one of the most frequent terms, the evidence is pretty solid.

By contrast, WERE / WAREN is far more language-specific in its priming. While the pattern of WAR use is similar to the English WAS use (and consequently employs a similar word order) the plural WAREN does not provide this option.

AM/BIN with SURE/ ÜBERZEUGT provide excellent insights how a genre functions in priming writers to develop *parallel language use* and use the same (i.e. equivalent in translation) keywords in a given context – here in combination with the first person singular BE.

This is in a wider sense comparable to IS / IST 3-word clusters. Furthermore, amongst the respective most frequent 4-word clusters, there are no mutual direct translations. However, very similar meanings are conveyed in very similar ways. This, again, hints at *genre-specific use.*

The other primary verb, HAVE, is used differently yet again. The past tense form, HAD / HATTE(N), only gives some idea of co-occurrence in their plural uses. This is also underlined by another singular form, the third person singular present tense HAS / HAT. While there are equivalent short clusters, the proportional frequencies are too divergent to provide solid evidence.

There appears therefore only one form of HAVE that is clearly primed dominantly by its genre: TO HAVE / ZU HABEN. These clusters seem to appear uniquely in biographies in such a high frequency as shown here. Neither the literature, nor other comparable corpora seem to mirror this specific use.

This paper is justifiably called *Traces*. While a number of valuable lines of thought have been opened up, it will need more time and detailed research of language use to have a sound foundation to make further conclusions.

References

Biber, D; Conrad, S; Leech, G. 2002. Longman Student Grammar. Edinburgh: Pearsson Education Ltd.
Helbig / Buscha. 1993. Deutsche Grammatik. Leibzig; Langenscheidt.

Hoey, M. 2005. Lexical Priming; London: Routledge.

Hunston, S.; Francis, G.; Manning, E. 1996. Collins Cobuild Grammar Patterns 1: Verbs   London: Collins Cobuild.

Hunston, S. and Francis, G. 2000. Pattern grammar : a corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.

O'Donnell Matt. 2006. "*Is Corpus Linguistics too Anglo-Centric*". Presentation. PG Seminar, University of Liverpool, October 2006.

Rundell, M. (Editor-in-Chief); Hoey, M. (Chief Adviser). 2002. Macmillan English Dictionary for advanced Learners. Oxford: Macmillan.

Sinclair, J.M. 1991. Corpus, Concordance, Collocation. Oxford: OUP.