BE and HAVE: Qualities and Shortcomings

Pierre Labrosse

E.A. Formes-Discours-Cognition and C.A.T.I., Université de Paris IV-Sorbonne, Paris, France

Abstract

This article examines the hypothesis that a speaker uses the $HAVE$ + noun construction when (s)he wants to express somebody's positive qualities. This hypothesis was tested with two different corpora using two different retrieval programs. The numerous steps that are necessary show that it can take a considerable amount of time to reach sound conclusions and suggests the need to improve the building of those corpora or their retrieval software programs. χ^2 and z-score tests were used to analyse the results. The above hypothesis proves to be too strong: although the verb $HAVE$ is never significantly associated more frequently with shortcomings, the verb BE is sometimes associated more frequently with qualities. The lists of qualities and shortcomings expressed with $HAVE$ or with BE show that BE collocates with a wider range of qualities than shortcomings. These lists also indicate that the qualities often quoted by the theoretical linguist are not those most frequently found in corpora. They also point to differences between the qualities that collocate with $HAVE$ and those that collocate with BE. The use of corpora enabled me to widen the range of qualities or shortcomings that are said to collocate with $HAVE$ or BE, and revealed that, contrary to the hypothesis, the $HAVE$ + noun construction can apply to inanimate subject referents.

Correspondence:
Pierre Labrosse,
UFR d'anglais,
Université de Paris IV –
Sorbonne, 1 rue Victor
Cousin, 75230 Paris
Cedex 05, France.
E-mail:
pierre.labrosse@club-internet.fr

1 Introduction

P. Cotte, a French linguist, remarks that the verb *be* collocates with qualifying adjectives having a positive or negative connotation (i.e. *BE courageous, BE cowardly*) whereas the verb *have* is followed by nouns expressing positive qualities[1] only (i.e. *HAVE courage* versus * *HAVE cowardice*).[2] This $HAVE$ + Noun construction conveys a characterization of the subject with nouns expressing modal qualities (e.g. *courage*) which are focalized. Contrary to the usual expression of possession in which the object refers to an entity different from the subject, these qualities are part and parcel of the subject referent. In this construction, the verb $HAVE$ seems to introduce

qualities that differentiate this referent of the subject from other possible referents, including the speaker. But it does not introduce his/her shortcomings (* *HAVE cowardice*), most probably because of an empathy akin to that delineated in Boisson (1987). The use of nouns preceded by a zero determiner to express the qualities of the subject referent implies that these characteristics can be shared by people other than that referent.[3]

On the other hand, the verb BE can somehow distance the speaker from the subject referent, although it is usually held to express an identity between the subject and its complement. This is probably due to the fact that its history is linked to deixis. The use of adjectives to express the

Table 1 Various forms of *HAVE* or *BE* and qualities or shortcomings investigated

HAVE forms	Noun	*BE* forms	Adjectives
have VHB	+ courage/cowardice	*am* VBB	+ courageous/cowardly
has VHZ	+ patience/impatience	*are* VBB	+ patient/impatient
	+ compassion/callousness	*is* VBZ	+ compassionate/callous
had VHD	+ beauty/ugliness	*was* VBD	+ beautiful/ugly
	+ intelligence/stupidity	*were* VBD	+ intelligent/stupid
	+ tact/tactlessness	*be* VBB	+ tactful/tactless
had VHN		*been* VBN	
have VHI		*be* VBI	
having VHG		*being* VBG	

characteristics of the subject implies that these characteristics apply to the subject referent only, not to the speaker.

2 Does *Have* Collocate More Frequently with Nouns Expressing Qualities?

2.1 Hypotheses

P. Cotte advanced two hypotheses. The first hypothesis adopted an onomasiologic (from mind to signs) point of view: when a speaker wants to express somebody's quality/qualities, (s)he will use the *HAVE* + noun construction. The second, or corrected hypothesis, took a semasiologic (from signs to mind) point of view: the verb *BE* can be used to express a person's qualities or shortcomings. The obvious consequence, from an onomasiologic point of view, is that when a speaker wishes to express somebody's shortcoming(s), (s)he uses the verb *BE*.

I tested these hypotheses with the British National Corpus World Edition December 2000 (hereafter BNC) for some of the qualities quoted by P. Cotte, namely: *beauty, courage, intelligence*, and *patience*.[4] I have added two qualities taken from a vocabulary list: *compassion* and *tact*.

2.2 Queries

Using SARA (the SGML Aware Retrieval Application developed by T. Dodd for the BNC project), I built queries combining various forms of *HAVE* + singular quality/shortcoming noun or *BE* + corresponding quality/shortcoming adjective as listed in Table 1 subsequently.[5]

I accepted 3,910 hits: 85% for *BE* and 15% for *HAVE*. These proportions are slightly different from the proportions of *BE* and *HAVE* in the BNC: out of a total of 4.8 million hits, *BE* represents 76% of hits and *HAVE* 24%. A certain number of hits were rejected, but I did not keep count or a record of those because I selected the acceptable hits with the 'thin' option of SARA, an option which erases the hits that do not meet the researcher's requirements.

2.3 Results: the χ^2-test

2.3.1 Individual comparison

I first used the χ^2 test to compare the number of hits of *HAVE* + noun with the number of hits of *BE* + adjective for a given characteristic. This resulted in a table with two columns (*HAVE* or *BE*) and two rows (quality or shortcoming) for each characteristic. Such a table can be easily drawn on an EXCEL spreadsheet, as exemplified in Table 2 subsequently.

If we accept a risk of error equal or inferior to 5% ($\chi^2 = 3.841$ in the above case), we find a significant difference (with various risks of error) for all the characteristics studied except *tact/tactful* versus *tactlessness/tactless*, to which the χ^2-test cannot be applied because the number of hits expected for both characteristics with *HAVE* is lower than five.

2.3.2 General comparison

I then tried to use the χ^2-test to compare the number of hits of *HAVE* + noun and the number of hits of *BE* + adjective for all the characteristics studied above. This yields a table with two columns (*HAVE* or *BE*) and twelve rows (qualities or shortcomings), but the test is not applicable

Table 2 χ^2-test for *be courageous/cowardly* versus *have courage/cowardice*

Hits counted				Hits expected			χ^2-calculation		
Characteristic	BE	HAVE	Total	Characteristic	BE	HAVE			
Courage	72	295	367	*Courage*	94	273	4.986	1.707	
Cowardice	29	0	29	*Cowardice*	7	22	63.099	21.604	χ^2:
Total	101	295	396	Total	101	295	68.085	23.311	91.396

Table 3 Paired qualities and shortcomings with *HAVE* and *BE*

Hits counted				Hits expected			χ^2-calculation		
Characteristic	BE	HAVE	Total	Characteristic	BE	HAVE			
Courage/Cowardice	101	295	396	*Courage/Cowardice*	335.5	60.5	163.939	909.763	
Patience/Impatience	455	140	595	*Patience/Impatience*	504.2	90.8	4.792	26.593	
Tact/Tactlessness	96	5	101	*Tact/Tactlessness*	85.6	15.4	1.269	7.042	
Compassion/Callousness	71	27	98	*Compassion/Callousness*	83.0	15.0	1.745	9.683	
Beauty/Ugliness	1406	66	1472	*Beauty/Ugliness*	1247.2	224.8	20.206	112.134	
Intelligence/Stupidity	1184	64	1248	*Intelligence/Stupidity*	1057.4	190.6	15.145	84.047	χ^2:
Total	3313	597	3910	Total	3313.0	597.0	207.096	1149.263	1356.359

Table 4 Reduced deviation test for *BE courageous* versus *HAVE courage*

Verb	Hits counted	Percentage of hits counted (p_o)	Percentage in the BNC (p)	$n * p$	Reduced deviation	Significant at
BE	72	20	76	278.04	25.09750	More than 1 per billion
HAVE	295	80	24	88.96	25.09750	More than 1 per billion
Total (n)	367	100	100	367.00		

because there are not enough hits expected for *cowardice* and *tactlessness* with the verb *HAVE*. In such a case, the usual (mathematical) approach is to group the two characteristics. From a linguistic point of view, however, it would be difficult to group those two shortcomings and find a semantic common ground between them.

Two other solutions are available. The first is to exclude those two characteristics, in which case the χ^2-test is significant at more than 0.001. But such a solution is not very satisfactory from a linguistic point of view. The second (Table 3) is to group opposites in a semantic field, in keeping with Freud's studies of the subconscious (Freud, 1988, especially pp. 95–96).

The χ^2-test is again significant at more than 0.001, but it does not prove that certain qualities or shortcomings co-occur more frequently with *HAVE* or *BE*. It merely proves that some semantic fields expressing a certain quality and its opposite shortcoming appear more frequently with *HAVE* or with *BE*, namely *HAVE courage/cowardice*, *BE beautiful/ugly* and *BE intelligent/stupid*. But we do not know, for example, whether it is the collocation of *courage* or of *cowardice* with *HAVE* that yields this trend. A more discriminatory test is needed to answer this question.

2.4. The *z*-score

In France this test is called the 'reduced' deviation test, which seems to be the equivalent of the *z*-score. The 'reduced' deviation test is the difference between the percentage observed (p_o) and the

percentage expected (p) over the standard deviation σ (Table 4).

This test highlights differences with various degrees of significance: *HAVE* is more frequently found with certain qualities (*compassion, courage, patience*), whereas *BE* is more frequently associated with all the shortcomings and also with some qualities (*tactful, beautiful, intelligent*).

The fact that the verb *HAVE* is never significantly associated more frequently with shortcomings is consistent with either version of the hypothesis. So is the fact that the verb *BE* is always associated more frequently with shortcomings. But the fact that *BE* is also sometimes associated more frequently with qualities invalidates the strong (or first) version of the hypothesis (i.e. when a speaker wants to express somebody's quality/qualities, (s)he uses the *HAVE* + noun construction). Therefore, to express qualities, the speaker has a choice between *HAVE* + noun or *BE* + adjective, whereas (s)he seems to have no such choice when expressing shortcomings.

These results prompt two questions: is there a link between the qualities more frequently associated with *HAVE*? And similarly, is there a link between the qualities more frequently associated with *BE*? It is quite obvious that six qualities is not enough to establish either link and that more qualities should be studied. Although I had not studied all the qualities quoted by P. Cotte, I widened the range of the eleven qualities he quotes. To do so, I needed to find out what qualities or shortcomings can collocate with *HAVE* or *BE*, and therefore, to list all the hits for *HAVE* + noun and *BE* + adjective. Such a query raises the question of how one defines a quality or a shortcoming.

3 What Adjectives or Nouns are to be found with *BE* or *HAVE*, Respectively?

The other reason that led me to study the qualities and shortcomings associated with *BE* and *HAVE* generally was that I wanted to check whether the qualities or shortcomings quoted by P. Cotte were the most frequent to appear in a corpus. This could

prove to the theoretical linguist that corpus linguistics can be of some help.

3.1 Limitations

SARA, however, imposes some limitations on such a study. First of all, it is not possible to list all the hits for the *BE* + adjective and *HAVE* + noun constructions because it does not allow queries on grammatical categories. Moreover, the sheer number of hits for the six qualities already studied (nearly 4,000) makes it quite difficult to check and examine numerous new qualities or shortcomings closely. More precisely, the number of hits for *BE* + adjective ranges from 30 (e.g. *be tactless/cowardly*), a number which is quite easy to handle, to 1,250 (e.g. *be beautiful*), an amount which is more difficult to deal with. If I selected 100 qualities or shortcomings for statistical purposes, examining each and every occurrence of *BE* + adjective could prove to be an unmanageable task.

3.2 Solution

The solution I adopted was to use the BNC Baby (version 1.0) as a corpus and XAIRA (XML Aware Indexing and Retrieval Architecture) as a retrieval software program. This enhanced version of the SARA program allows queries on grammatical categories. Another advantage is that the size of the corpus is much smaller (four million words instead of 100 million), which entails a smaller number of hits. But the correlated drawback is that the range of media is also smaller, being limited to spontaneous conversation, academic prose, written fiction, and newspaper texts.[6]

3.3 Queries

The queries with XAIRA are quite similar to those delineated in Section 2.2, except that the different adjectives or nouns given in Table 1 can be replaced with the grammatical categories 'adjective' or 'noun'. So far, my research has concentrated on six verb forms: *is* VBZ, *be* VBB, *be* VBI, *have* VHB, *has* VHZ, and *have* VHI. These results, therefore, are only provisional since the past forms, the past participles and the gerunds have not yet been considered. Most importantly, there also remain two present indicative

forms of *BE* (*am* VBB and *are* VBB). But these queries yielded lists of hits (in XML format, easily converted into rtf or txt format), out of which I kept 378 hits on *HAVE* and 2,031 hits on *BE*. Each and every hit had to be carefully checked, and sometimes rejected, a process that I would like to expand upon in order to rebut the idea that the use of a computer gives results at the click of a mouse. This account might also prove useful to the BNC Baby corpus builders and to the developers of XAIRA.

3.4 Steps to process the answers to queries

To find out what qualities and shortcomings collocate with *BE* and *HAVE* in the BNC Baby, one needs to go through three steps:

- Checking that the hits comply with the construction studied.
- Extracting all the adjectives or nouns.
- Extracting the adjectives or nouns expressing qualities from the adjectives or nouns extracted in step 2, which implies a definition of quality or shortcoming.

3.4.1 *Checking*

As with the hits listed with SARA and the BNC, some occurrences need to be weeded out. I had to discard 19% of the 466 hits on *HAVE* + noun, which seems a rather high proportion. This proportion reached more than a quarter of all hits (27.4%) in the case of *has* + noun. Most of the occurrences discarded complied with the query, so they were not 'errors' strictly speaking. The problem was that, as far as I know, the query could not be formulated in such a way as to eliminate this noise. Here are examples of these unwanted hits along with their references in the BNC Baby (bold type added):

- Plural compound nouns

He said the regional health authority has **air disaster contingency plans** at Stansted. (hit text='CFC' $n =$ '57')

- Causative constructions:

Sometimes she had to **have water fetched** from miles up in the mountain. (hit text = 'FAJ' $n =$ '3,185')

- Object complements:

You've got to **have** food **available**. (hit text = 'KBD' $n =$ '7,345')

- Interrogative constructions:

Why **has** History **gone** wrong? (hit text = 'CTY' $n =$ '111')

- Subject-auxiliary inversion due to adverb fronting:

'**Never has** homo faber better understood that he has made history and never has he felt so powerless before history'. (hit text = 'CTY' $n =$ '473')

- Verb + complement deletion and subject-auxiliary inversion:

Seagram, last year's winner, has been a long way behind that excellence this season, as **has Bonanza Boy**, the 1991 favourite . . . (hit text = 'AHC' $n =$ '120')

- Time adverbials:

This is the emotional strain ambulance men have **day in, day out**. (hit text = 'A7W' $n =$ '511')

- Quantifiers:

She and Matthew must have **plenty** to talk about. (hit text = 'J54' $n =$ '2,593')

- Qualitative modifiers:

Yeah I mean that's the same as I did I must admit I did have **sort of** an argument with him once over his music. (hit text = 'KB7' $n =$ '2,377')

- Genitives:

And they have **granddad's money**. (hit text = 'KD3' $n =$ '87')

- Titles:

For a start, you have **goalkeeper Simon Tracey**, signed from Wimbledon last year. (hit text = 'A1N' $n =$ '489')

- Vocatives:

What we gonna have **dad**? (hit text = 'KBW' $n =$ '3,344')

- Proper names:

But . . . here [Ms Tomei] is the fully-fledged leading lady of a romance which appears

to have **Beauty and the Beast** and Tarzan and Jane as dual inspirations. (hit text = 'K37' n = '184')

- Phrasal verbs:

Injury-hit Rangers could **have** debutant England right-back David Bardsley **back** after a three-match absence with an ankle injury. (hit text = 'CH3' n = '4,986')

- Idioms:

He would **have** Christmas Day **off** … (hit text = 'CB5' n = '197')

- Formulae:

C will have Cpq = 0 and may also have **cuv = 0** (if [formula]); but this situation will not persist. (hit text = 'EWW' n = '1,296')

- Grammatical mistakes/errors:

A one, ju you know, I mean you have **complex of apartments** … (hit text = 'KCV' n = '1,860') (one could expect: *you have a complex of apartments*)

- Tagging errors:

'… we have **gurus** a dime a dozen'. (hit text = 'AL0' n = '199')

As for the 2,156 hits on BE + adjective, I rejected a much smaller proportion of occurrences (5.8%), the higher proportion (6.2%) relating to the '*s* form of the verb again. As can be seen from the following examples, there seems to be more errors due to the building of the corpus or the retrieval software program:

- Adjectives + nouns:

… all that is left is **abstract philosophy**. (hit text = 'K5C' n = '375')

- Compound nouns:

So one of those is real cheese and one is **cream cheese**. (hit text = 'KBW' n = '18,564')

- Genitives modified by an adjective:

This is **English law's rather clumsy attempt** to reflect the element of partial justification in the doctrine of provocation. (hit text = 'ACJ' n = '208')

- Passive constructions:

… the situation **is complicated by** the fact that some large ammonites have inner whorls looking rather like Promicroceras. (hit text = 'AMM' n = '590')

- Interrogative constructions:

Should there be proof of actual danger to someone, or **is potential danger** enough? (hit text = 'ACJ' n = '636')

- Subject-auxiliary inversion due to adverb or adverbial fronting:

Also back on the scene **is veteran spinner** Keith Madeley … (hit text = 'K3A' n = '902')

- Boundary between two clauses (square brackets added):

It was in their interests … not to let it get about [that simply [by keeping their hands [where they ought to be]] **young people could have healthy teeth** and gums for life]. (hit text = 'CDB' n = '1,705')

- Grammatical mistakes or transcription errors:

No you **be good boy**. (hit text = 'KD3' n = '10') (instead of *Now you be a good boy*)

- Tagging errors:

(i) adverb + compound adjective: … I'll **be bloody pig sick**! (hit text = 'KCU' n = '4,774')
(ii) progressive form: Midfielder Donnelly **is interesting** Villa boss Ron Atkinson … (hit text = 'CH3' n = '6,662')
(iii) preposition: I actually come from the Dales from Middleham which **is near** Leyburn … (hit text = 'KD7' n = '64')

- Typing mistakes:

But … there is also a need for serious spending on infrastructure and services, **is well as** for start-up finance or subsidy. (hit text = 'AS6' n = '148') (instead of *as well as*)

- Unexplainable results[7]:

Alright what do we say about throwing toys eh? (hit text = 'KBW' n = '12,670') (there is no verb *is* or adjective)

original as far as I know. (hit text = 'KBD' *n* = '3,579') (there is no verb *is* and the sentence is incomplete)

I discarded 214 hits out of a total of 2,622 hits, which amounts to an overall percentage of 8.2% and proves that checking is still a useful, albeit time-consuming, step.

3.4.2 Extracting all the nouns or adjectives

This step is a preliminary to finding the adjectives or nouns expressing qualities or shortcomings with *BE* or *HAVE*. At this stage, one can notice that the ratio of different nouns or adjectives over the number of hits, which could be called 'variance', is around 42% for *BE* and 55% for *HAVE*. When one considers the three different verb forms studied for each verb, this variance is also more consistent in the case of *BE* (40–44%) than in the case of *HAVE* (49–70%). There seems to be no obvious reason why the range of adjectives with *BE* should be more restricted than the range of nouns with *HAVE*, apart from the frequent use of constructions expressing modality (chiefly *be able to/capable of*). Further research could try to categorize those nouns or adjectives into different semantic fields, an undertaking that would beg the question of semantic categories for nouns or adjectives.

3.4.3 Extracting the adjectives or nouns expressing qualities/shortcomings

3.4.3.1 A question of semantics

This question is also relevant to our study, as we have to sort out the nouns or adjectives expressing qualities or shortcomings.

The first source of information one can think of is the conceptual index of general grammars. In Huddleston and Pullum, 2002, no mention is made of 'quality' in general or in the semantic classes of adjectives, whereas a concept such as 'quantity' and the category of 'noun/NP denoting quantity' are mentioned. This is also true of the index of Biber *et al.* (1999), which includes a reference to 'quantifying noun'. In Quirk *et al.* (1985), 'quality' is mentioned as a semantic classification for partition (e.g. *a new kind of computer, delicious sorts of bread*: 249) and for verbs, whereby 'qualities are relatively permanent and inalienable properties of the subject referent' (e.g. *Mary is Canadian/has blue eyes*: 200). This type of quality, however, is not what P. Cotte seems to have in mind when comparing *have courage/* * cowardice* and *be courageous/cowardly*.

One can also look up the definition of quality or shortcoming in a dictionary. In the *Oxford Advanced Learner's Dictionary* (2000), the most appropriate definition of quality for our purposes is 'a thing that is part of a person's character, especially sth good: *honesty and generosity*' (p. 1035), and a shortcoming is 'a fault in sb's character, a plan, a system, etc.' (p. 1186). These definitions take into account moral and, probably, intellectual qualities, but they do not seem to include physical qualities like *beauty*.

A third source of information could be the eleven examples quoted by P. Cotte, which we mentioned at the end of section 2.1, but they do not seem sufficient to extrapolate a general definition. The last means to sort out qualities could be a test proposed by a lexicographer[8], namely that a quality noun should collocate with *(s)he is gifted with …* and *she lacks …*. I personally think that we should say '*(s)he is gifted with …* OR *she lacks …*,' because *(s)he is gifted with (sex) appeal* does not seem right, contrary to *(s)he lacks (sex) appeal*. This test, of course, would not apply to shortcomings.

It is quite obvious that a more precise, if not convenient, definition of those two concepts is required. It is clear, however, that these qualities or shortcomings should apply to human beings. This is inherent in the explanation given by P. Cotte for the more frequent use of the *HAVE* + noun construction to express qualities. Yet another problem occurs: although one can check the definitions of each and every 'quality/shortcoming adjective,' how should adjectives with multiple meanings like *common* be dealt with? Do we check in each and every utterance whether this adjective expresses a shortcoming ('typical of sb from a low social class and not having good manners') or a frequency ('happening often, existing in large numbers or in many places = commonplace') and reject all the hits with the latter meaning? Theoretically, this is what should be done, but would it be feasible?

Table 5 Qualities with *HAVE* in the BNC Baby in frequency order

Noun	Number	%
Knowledge	6	18
Experience	5	15
Confidence	3	9
Character	2	6
Faith	2	6
Mercy	2	6
Power	2	6
Appeal	1	3
Breeding	1	3
Capacity	1	3
Commitment, (consultation and collaboration)	1	3
Courage	1	3
Determination and drive	1	3
Experience (mountain ~)	1	3
Faith and confidence	1	3
Influence	1	3
Leverage	1	3
Style	1	3
Taste	1	3
Total	34	100

3.4.3.2 List of qualities/(shortcomings) with *HAVE*

Nine percent of the *HAVE* + noun hits express qualities, a figure which does not seem overwhelming. Table 5 shows the eighteen quality nouns that emerge from those hits, in frequency order.

It is noteworthy that no shortcomings collocate with *have*, which corroborates both the strong and weak versions of the hypothesis. Among the eleven qualities mentioned by P. Cotte, we find only *character, courage*, and *style*. Moreover, these three qualities are not the most frequent: *character* represents 6% of occurrences, and *courage* and *style* only 3%, respectively. They are well below the two most frequent qualities: *knowledge* and *experience*, which represent 18% of occurrences each, if we add together *experience* and *mountain experience*.

It is interesting to note that two of these qualities (*character* and *style*) are acquired through learning; they are less inherent than *confidence* or *faith*, for example. It could be argued that most qualities that co-occur with *HAVE* are by and large acquired through experience: *knowledge, experience, character, power, breeding, influence, leverage, style*, and *taste*.

This would corroborate widely known analyses of the perfect, especially the *present perfect*.[9]

3.4.3.3 List of qualities/shortcomings with *BE*

How many qualities and how many shortcomings collocate with BE?

Nearly 40.5% of the *BE* + adjective hits express qualities or shortcomings. If we compare this figure with the 9% of hits expressing qualities with *HAVE*, we have to conclude that, although *HAVE* is considered as a state verb, it is not as widely used as *BE* for qualifying purposes. This is not very surprising, since *BE* is usually divided into three main uses: existence, identity, and qualification.

For the three different forms of *BE* examined in the BNC Baby we find 194 different qualities or shortcomings, more than ten times the number of qualities that collocate with *HAVE*. One has to keep in mind, however, that the verb *BE* is more frequent than the verb *HAVE*: in the BNC, out of a total of 4.8 million hits for both verbs, *HAVE* represents only 24.2% of occurrences. This proportion seems to imply that we could expect forty-seven different qualities to appear with *HAVE*, more than double the number that we actually found. We could conclude that the range of qualities and shortcomings expressed is much more narrow with *HAVE* than with *BE*. But such an inference would not take into account the fact that no shortcomings collocate with *have*. And it would run counter to the variance we calculated for *HAVE* + noun, this noun not being restricted to a quality or a shortcoming.

I had difficulty classifying some adjectives with multiple meaning. For example, an adjective like *candid*, which is defined as 'saying what you think openly and honestly; not hiding your thoughts' (OALD, 2000, p. 170), could be considered as a quality or a shortcoming depending on the situation. When there was only one or very few occurrences of such an adjective, I could use the context to classify it.

Finally, I sorted out 110 different quality adjectives immediately following *BE* in the BNC Baby and 84 different shortcoming adjectives. We can conclude that this verb also seems to favour the expression of qualities, more so from a numerical

Table 6 List of the top twenty quality adjectives immediately following *BE* in frequency order

Adjective	Number	%
Able	140	22.9
Important	47	7.7
Nice	37	6.0
Careful	31	5.1
True	24	3.9
Good	21	3.4
Quiet	18	2.9
Honest	13	2.1
Interesting	13	2.1
Relevant	11	1.8
Strong	11	1.8
Useful	11	1.8
Bold	10	1.6
Independent	9	1.5
Responsible	9	1.5
Capable	8	1.3
Helpful	8	1.3
Open	8	1.3
Confident	7	1.1
Quick	7	1.1

point of view (74.5% of all hits for *BE* + adjective) than from a variety point of view (56.5% of all adjectives). This confirms the weak version of the hypothesis, whereby *BE* can be used to express both qualities and shortcomings.

From these results for *HAVE* and *BE*, we can see that there is a general tendency to express qualities rather than shortcomings (since all the clauses examined are affirmative). This could be accounted for sociolinguistically or sociologically, if one considers that it is more appropriate to praise people and to ignore their shortcomings.

How many adjectives exemplify the qualities quoted by the theoretical linguist?

There are only three adjectives (*beautiful*—3 hits, *charming*—1 hit, *patient*—6 hits) that exemplify the qualities quoted by P. Cotte. One finds no occurrences of *courageous* (but there is one occurrence of *brave*), *intelligent* (but there are two of its antonym *stupid*) or *stylish* (which might appear only in attributive position).

To parallel the list of eighteen qualities expressed with the *HAVE* + noun construction (Table 5), Table 6 shows the twenty most frequent quality adjectives following *BE*. It is quite remarkable that none of the quality adjectives which we could expect

from the list established by the theoretical linguist (and especially *beautiful, brave, charming, patient*) appear in this list.

Do quality nouns with HAVE *have a corresponding quality adjective occurring with* BE *in the BNC Baby?*

To stay within the field of corpus linguistics, it would be interesting to see if the qualities expressed with the *HAVE* + noun construction in the BNC Baby have a corresponding adjective following *BE* in the same corpus (Table 7).

It appears that only eight nouns with *HAVE* have corresponding adjectives following *BE* in the BNC Baby: *appeal, capacity, character, confidence, courage, faith, influence*, and *leverage*. This is not the case for ten other quality nouns: *breeding, commitment, determination, experience, knowledge, mercy, power, style*, and *taste*. There does not seem to be a common denominator to either group of nouns that could explain why some qualities are also expressed through the *BE* + adjective construction and some are not.

It is noteworthy that the idea of capacity, which is certainly the most frequently expressed with the *BE* + adjective construction (*able, capable*), is not so well ranked among the *HAVE* + noun constructions, near the top of the second half. But I would not venture to say that there is a reverse order in the qualities associated with *BE* and *HAVE*.

Do the quality adjectives with BE *in the BNC baby have a corresponding quality noun with* HAVE *in the BNC?*

A complementary line of research is to see which quality adjectives with *BE* in the BNC Baby have a corresponding quality noun with *HAVE*. Such a query has to be put to the BNC since we already know the answers from the BNC Baby (Table 5). This will give us another opportunity to increase the list of qualities that can be expressed with *HAVE*.

Two preliminary remarks have to be made. The first is that all those quality adjectives have a corresponding noun, some difficult cases being solved with the *-ness* suffix only.[10] The second remark is that some adjectives have a corresponding noun with a different usage. For example, *agreeable* means: 'pleasant and easy to like' (OALD, 2000, p. 25). But *She is agreeable* would not have for equivalent: * *She*

Table 7 Adjectives corresponding to *HAVE* + noun constructions in the BNC Baby

Rank with *HAVE*	Noun (with *HAVE*)	Adjective (with *BE*)	Rank with *BE*
1	Knowledge	Knowledgeable = 0 hit	
2	Experience	Experienced = 0 hit	
3	Confidence	Confident = 7 hits (1.1%)	19
4	Character	Strong-willed; courageous; firm = 2 hits (0.3%); resolute.	41
5	Faith	Faithful = 1 hit (0.2%)	52
6	Mercy	Merciful = 0 hit	
7	Power	Powerful = 0 hit	
8	Appeal	Attractive = 2 hits (0.3%);	41
		Interesting = 13 hits (2.1%)	8
9	Breeding	Well-bred = 0 hit	
10	Capacity	Capable = 8 hits (1.3%);	14
		Able = 140 hits (22.9%)	1
11	Commitment	Committed = 0 hit	
12	Courage	Courageous; brave = 1 hit (0.2%)	52
13a	Determination	Determined; resolute = 0 hit	
13b	Drive	Energetic; dynamic = 0 hit	
14	Influence	Influential = 1 hit (0.2%)	52
15	Leverage	Influential = 1 hit (0.2%)	52
16	Style	Stylish = 0 hit	
17	Taste	Elegant, refined = 0 hit	

has pleasure and easiness (to like). *Agreeable* is a description, an evaluation (cf. the category 'evaluative adjectives'), usually by somebody different from the subject referent (*I am agreeable* sounds strange). It is not a quality that somebody has in him/herself, and from which that same person benefits, a feature which is typical of *HAVE*. It is a quality that is appreciated by somebody else because that somebody else benefits from this quality.

Similarly, *amazing* means: 'very surprising, especially in a way that makes you feel pleasure or admiration' (OALD, 2000, p. 35). *-Ing* adjectives seem to imply that somebody different from the subject referent feels something for his/her own benefit or to his/her detriment. If one could say: * *He is amazing, so he has amazement*, *amazement* would be a feeling that would have to be felt by the subject referent. The same is true for *surprising* versus *surprise*.

Desirable seems very similar. This *-able* adjective means: '(of a person) causing other people to feel sexual desire' (OALD, 2000, pp. 340–341). A feeling or reaction is caused by the subject referent but is not felt by this same subject referent: *she is desirable* does not mean *She has a strong desire for X*.[11] The same is true for *attractive* versus *attractiveness*.

In Table 8, we notice only four examples (*beauty, charm, intelligence, patience*) out of the eleven examples given by P. Cotte, these four examples totalling 35 hits (6.8% of all hits). It must be remembered that some adjectives (*courageous, intelligent, stylish*) do not immediately follow *BE* in the BNC Baby. The most frequent example (*charm*) ranks sixth in the list. Most importantly, we can add twenty-seven quality nouns to this list. Although some linguists say it is limited by its corpora, it is clear that corpus linguistics widens the range of research.

These qualities do not seem to share a common denominator apart from the syntactic property of entering the *HAVE* + noun construction and the semantic property of applying to humans. Yet, if we take a close look at the examples, we notice that a certain proportion of subjects do not refer to humans. To give only two examples:

It is not Venice but **it has warmth**, colour, and views such as could stretch the most infertile imagination. (hit text = 'CN4' *n* = '804')

It is the grid for a scene of battle, whether it be the Rout of San Romano by Uccello or Guernica by Picasso. Whereas both vertical

Table 8 *Have* + quality nouns in the BNC corresponding to the *BE* + quality adjectives in the BNC Baby

Quality adjective	Quality noun	Number	%
Able	Ability	8	1.6
Active	(Physical) Activity	1	0.2
Autonomous	Autonomy	5	1.0
Beautiful	**Beauty**	2	0.4
Brilliant	Brilliance	0	
	Intelligence	5	1.0
Charming	**Charm**	16	3.1
Confident	Confidence	99	19.2
Consistent	Consistency	1	0.2
Famous	Fame	2	0.4
Good	Goodness	1	0.2
Honest	Honesty	1	0.2
Honourable	Honour	2	0.4
Important	Importance	8	1.6
	Authority[12]	93	18.1
Independent	Independence	2	0.4
Influential	Influence	19	3.7
Logical	Logic	1	0.2
Nice	nicenEss;	1	0.2
Open	Openness	2	0.4
Patient	**Patience**	12	2.3
Pure	Purity	2	0.4
Relevant	Relevance	37	7.2
Responsible	Responsibility	154	29.8
Sensible	Sense	9	1.7
Sensitive	Sensitivity	2	0.4
Stable	Stability	4	0.8
Strong	Strength	14	2.7
Successful	Success	5	1.0
Tough and wise-cracking	Toughness	1	0.2
Traditional	Tradition	2	0.4
Warm	Warmth	4	0.8
Wise	Wisdom	2	0.4
	Total	516	100.0

Bold typeface indicates the four examples given by P. Cotte that appear in the BNC.

and horizontal **lines have stability**, sloping lines are in a state of change from one condition to the other. (hit text = 'ADX' n = '556')

Although it cannot be denied that these qualities generally apply to humans, this tendency admits exceptions which the use of corpora can highlight. Admittedly, our results are only provisional, as we need to study all the other forms of the verbs *BE* and *HAVE*, starting with the remaining forms of the present indicative of *BE*. But one already has to qualify the claim that a zero article precedes the noun in the *HAVE* + quality noun construction because the subject referent is human and every human, including the speaker, would like to share in this quality.

4 Conclusion

I hope it has become evident that corpus linguistics does not yield results with a simple click of a mouse. One has to check all the hits, which can be quite numerous.

This study has also shown that corpus linguistics enables the linguist to check (and correct) his/her hypotheses. I proved that the first hypothesis (when a speaker wants to express a person's quality/ qualities, (s)he uses the *HAVE* + noun construction) was overstated, and that a weaker hypothesis seems to hold: the verb *HAVE* is not used to express shortcomings; the verb *BE* can be used to express either qualities or shortcomings.

It has finally been shown that corpus linguistics widens the scope of research and promotes further research: new qualities and shortcomings have been found and can be studied more thoroughly. Using the BNC and SARA, we could further calculate and compare the frequencies for the qualities that we had found and that can be expressed with *BE* and *HAVE*. Another possible study would be to check that there are no nouns with *HAVE* expressing the shortcomings found with *BE*. We could also verify that *BE* collocates more frequently with inanimate subjects, although we have seen that *HAVE* can collocate with such subjects.

Thus, corpus linguistics can change the way French linguists do linguistics on English or any other language. Some linguists are happy with a few examples and work out a whole theory out of them. Others will obviously offer counterexamples and think they prove the first theory wrong. This is very much a matter of seeing linguistic phenomena in black and white, whereas a corpus linguist generally has a more balanced attitude. (S)he provides more examples to check a theory and highlights trends which (s)he or the theoretical linguist still has to explain.

Acknowledgements

References

Aston, G. and Burnard, L. (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Biber, D., Johanson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman Pearson Education Ltd.

Boisson, C. (1987). *Anglais HAVE, français AVOIR et l'empathie* in La Transitivité. Université de Saint Etienne: C.I.E.R.E.C. Travaux 52, pp. 155–81.

Cotte, P. (1996). *L'explication Grammaticale de Textes Anglais*. Perspectives Anglo-Saxonnes. Paris: Presses Universitaires de France Coll. Perspectives anglo-saxonnes.

Cotte, P. (1997). *Grammaire linguistique*. Paris: Didier Erudition – CNED.

Cotte, P. (1998). *Have* n'est pas un verbe d'action: l'hypothèse de la réélaboration. In Rousseau, A. (ed.), *La transitivité* Villeneuve d'Ascq: Presses Universitaires du Septentrion Coll. Travaux et Recherches, pp. 415–39.

Freud, S. (1988). *Sur le Rêve*. Paris: Gallimard.

Hornby, A. S. (2000). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.

Huddleston, R. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

OALD (2000) see Hornby, A. S. (2000).

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Harlow: Longman Group UK Ltd.

Notes

1 Throughout this article, the term 'quality' will mean 'excellence of disposition; good natural gifts' for animates, and 'the degree or grade of excellence, etc., possessed by a thing' for inanimates, as explained in the Articles 1c and 8 of the OED. This concept is studied in Section 3.4.3.1.

2 This remark, and the following first hypothesis, were put forward during a seminar in October 2004. For further discussion of the verb *have*, see Cotte (1996) (especially pp. 72–4 and pp. 93–105), Cotte (1997) especially pp. 43–65, 116–30, and pp. 213–48), and Cotte (1998).

3 See the analysis of the generic *man* in P. Cotte, 1996: 218.

4 P. Cotte quotes eleven qualities altogether. *Backbone, character, charm, grit, personality, spirit*, and *style* would have to be added to the four qualities studied here.

5 The meaning of the part-of-speech tags can be found in Aston and Burnard (1998, pp. 230–3). Concerning the verb BE, VBB is the present tense forms, except for *is* or *'s*; VBD is the past tense forms; VBG is the *-ing* form; VBI is the infinitive form; VBN is the past participle form, and VBZ is the *-s* form. As for the verb HAVE, VHB is the finite base form; VHD is the past tense form; VHG is the *-ing* form; VHI is the infinitive form; VHN is the past participle form, and VHZ is the *-s* form.

6 In the BNC, there are five classes of written media, apart from 'unclassified' (0.5%): books, periodicals, miscellaneous published, miscellaneous unpublished, and written-to-be-spoken. In the BNC Baby, 'written fiction' seems to be a mix of <medium = books> and <domain = imaginative>. Therefore, books seem to be limited to one domain, imaginative. The category 'academic prose' does not seem to have a well-defined medium: books, periodicals, miscellaneous published, or miscellaneous unpublished. 'Periodicals' should probably be excluded since the media classification of the BNC Baby provides for the category 'newspaper texts'. Quite rightly, the domain of this category is not specified, as it ranges over arts, belief and thought, commerce and finance, leisure, natural and pure science, applied science, social science, and world affairs in the BNC.

7 The tags <kw></kw> should indicate the query focus; in this case, they should be positioned between the verb form *is* and an adjective, e.g. *He is* <kw></kw>*able and energetic*.

8 F. Antoine, personal communication.

9 See Cotte (1996, p. 73) or (1997, p. 232) among many possible references.

10 F. Antoine remarks (personal communication) that the fact that many nouns ending in *-ness* are not very common could account for the relatively high

proportion of *BE* + adjective constructions to express qualities. The proportion of these nouns would have to be checked first to confirm this interesting hypothesis.

11 Thus the difference between *desirability* and *desire*.

12 *Important* is also glossed as *influent*. The corresponding noun (*influence*) is examined when dealing with *influential*.