

On the Development of *because*: A Corpus-based Study

Ji Won Lee, Seoul National University

This paper follows the rise and development of the causal connective *because* in Middle and Early Modern English. The causal marker of native origin, *forþam þe*, used since the beginning of English, decreased in number and ultimately became obsolete by the end of Middle English period. Causality being a basic quotidian concept, it is logical to assume that the decline of the traditional connective is tied to the introduction of the new form, *because*. In this light, the paper addresses the rise and decline of markers of reason. By aid of data from the Helsinki Corpus, the study investigates the formation and stabilisation of *because* to its present form in a diachronic perspective. In the process, grammaticalisation of *because* is discussed.

1. Introduction

This paper addresses the rise and development of the causal connective *because*. In order to express causative relationship, we use *because* on a daily basis. But this expression came into English only in the 14th century. Furthermore, it even looked slightly different (*by cause that*) when it was first introduced.

Old English had a native connective *forþam þe*¹, which has approximately the same meaning and function as *because*. *Forþam þe* existed until the 15th century but its occurrences sharply decreased entering the second half of 13th century (see section 2 below). In its place, a causal marker of Anglo-Norman origin, *because*, was adopted in the 14th century. Study revealed that *because* underwent transformation before it was realised into current form (PDE *because* <ModE *be cause (that)* < ME *by cause (of, that)* < Fr *par cause de, à cause que*).

In the following sections the study follows the formation and stabilisation of *because* as an English vocabulary. In doing so, a process called grammaticalisation is discussed. The Helsinki Corpus of English Texts (HC) is used for diachronic research, along with the Word Smith Concordance program. Oxford English Dictionary (OED) also served as a valuable reference. We begin with the native causal connective *forþam þe*.

2. The native causal connective *forþam þe*

Forþam þe was the causal connective used during the Old and the Early Middle English period. Its internal construction can be broken down to *for* followed by various forms of dative or instrumental demonstrative pronoun *ðan, þan, þi, þam* and subordinating particle *þe/ðat*. As spelling was not yet fixed in the earlier times,

¹ *Forþam þe* is used here as a representative form of all the variations preposition *for* + dative and instrumental demonstrative pronoun (*ðan, þam, ðam, ði* etc.) + subordinating particle (*þe, þat*) construction takes to express causality.

Ji Won Lee

as many as 74 different written forms² were found in the HC. Its primary meaning in the sentence is the same as today's *because* in that it marks the relationship of reason and consequence. Here are examples from the M1 (1150-1250) sub-period³ of the HC:

- (1) Þerefter wæx suythe micel uerre betuyx þe king & Randolf eorl of Cæstre: noht *forþi ðat* he ne iaf him al ðat he cuthe axen him, also he dide alle other. (After this waxed a very great war betwixt the king and Randolph, Earl of Chester; not *because* he did not give him all that he could ask him, as he did to all others.) (*Peterborough Chronicle* AD 1140)
- (2) Mi feader & Mi moder *for-þi þt* ich nule þe forsaken; habbe forsake me. (My father and my mother, *because* I haven't forsaken you, have forsaken me.) (*ST Juliana* M2/HC)

Table1 below shows the occurrences of *forþam þe* in all periods of the HC. The number drops noticeably in the M2 (1250-1350) period and no instance is found entering the Early Modern English period:

Table1: Occurrences of *forþam þe* in the HC (per 100,000 words)

O1	O2	O3	O4	M1	M2	M3	M4
46	149	122	154	76	2	2	3

The new form *because*, on the other hand, gains frequency starting from the M3 (1350-1420) period. Table2 below is where the occurrences of *because* is summarised:

Table2: Occurrences of *because* in HC (per 100,000 words)

M3	M4	E1	E2	E3
19	31	98	121	72

It is interesting to note that just as there were many forms of *forþam þe*, *because* did not emerge fully fixed in form⁴. In the next section, such morphological diversity and the significance of it will be discussed in terms of grammaticalisation.

² Refer to Appendix A at the end of the paper for a comprehensive list.

³ Time division here follows the convention of the Helsinki Corpus: O1 (-850), O2 (850-950), O3 (950-1050), O4 (1050-1150), M1 (1150-1250), M2 (1250-1350), M3 (1350-1420), M4 (1420-1500), E1 (1500-1570), E2 (1570-1640), E3 (1640-1710)

⁴ Refer to Appendix B at the end of the paper for a comprehensive list.

3. The new causal connective *because*

3.1 The emergence of *because*

According to OED, *because* was originally a phrase consisting of preposition *by* and substantive *cause*. The dictionary reads that *because* was followed by the preposition *of*, or subordinator *that* or *why* to express cause or purpose. *For cause* was equivalent to *because* thus *for because* is said to be found as an interim form. Results obtained from searching the ME and EME section of the HC is provided in 3.1.1, starting with the internal construction of *because*.

3.1.1 *Because: its form*

Data from the HC rendered ample variations of *because* which reflected its change in form over the course of time. Evidence for two points is provided in this subsection. First, since *because* was initially a phrase (*by + cause*), two word construction is frequent in the earlier period but statistics lean toward dominance in merged form (*because*). Second, *because* with preposition *by* is prevalent at the incipient developmental stage but ones with *be-* is much more frequently found in the later period.

Table3 shows frequencies of merged (*because*) and separate (*by cause*) construction of *because*. The hyphenated form *be-cause* that can be regarded as an interim form on the way to the merged one. In the M3 (1350-1420) period, separate construction prevails at 91%. The percentage drops to 45 in the M4 sub-period and in the EME period, *because* is mostly one word.

Table3: Occurrences of merged and separate construction of *because*

	preposition separated	preposition hyphenated	preposition merged	total
M3	28(91%)	1	5	34
M4	30(45%)	17(26%)	19(29%)	66
E1	10	0	177(95%)	187
E2	8	0	222(97%)	230
E3	0	0	124(100%)	124

Data analysis from another angle reveals that prepositional elements are headed towards unity as well. The original preposition *by* prevails in the earlier periods but increasingly *be-* becomes dominant. Table4 shows the change:

Table4: Spelling variations of the prepositional elements

	<i>be-</i>	<i>by-</i>	<i>bi-</i>	total
M3	9	22(65%)	3	34
M4	28(42%)	36(54%)	2	66
E1	114(61%)	41(22%)	32	187
E2	222(97%)	7	1	230
E3	124			124

In Table4 above, *because* in combination with the preposition *by* accounts for 65% of the total make up in the M3 (1350-1420) period. But in the subsequent period, the percentage drops to 54. When it enters the EME period, *be-*, presumably the phonetically reduced form of *by*, occupies majority of the occurrences. Finally in the E3 period, the spelling of *because* is unified to current usage. Although the knowledge of the arbitrariness of medieval spellings may cast doubts on the validity of this study, general trend is strong enough to conclude that the transformation of *because* into current form is reflected in the data.

3.1.2 *Because: because that and because of*

In PDE, *because of* is a prepositional phrase followed by a noun attributed as the cause. The ratio of prepositional phrasal use (*because of*) to subordinating conjunctive use of *because* in the HC data is shown in Table5. The proportion of the prepositional phrase decreases with time resting at 7% ($8/124=0.065$) in the E3 period. The earlier form *because that* shows the most remarkable change in frequency in that it occupies 56% ($19/34=0.558$) of the *because*-related marker slot in the M3 (1350-1420) period but occurring only once in the E3 (1640-1700) period.

Table5: Occurrences of *because*-related grammatical markers

	<i>because that</i>	<i>because</i>	<i>because of</i>	total
M3	19	1	14	34
M4	10	41	15	66
E1	16	156	15	187
E2	7	207	16	230
E3	1	115	8	124

On the Development of Because: A Corpus-based Study

History of the English language tells us that subordinating particle *that* (*þe*, *þat*) came after the connective to mark its function in a sentence, one prominent example being *forþam þe*, the predecessor of *because*. Same rule seems to have applied to *because* since it came with the subordinator in its early forms (19 compared to 1 in the M3 period). But frequency drops at a considerable rate and by the E3 period, the situation is completely reversed (115 appearing without as opposed to 1 with the subordinator).

3.1.3 The interim form *for because*

As mentioned in 3.1, the interim form *for because* was found in the HC. Examples (3) and (4) are from the M4 (1420-1500) and E2 (1570-1640) period respectively. And (5) is the OED's example of the interim construction from the M4 period.

- (3) Þen **for bycause þat** Sonday ys no day of fastyng, þerfor 3e schull begyn your fast at Aske-Wanysday. (Then *because* Sunday is no day of fasting, therefore you shall begin your fast at Ash Wednesday.)(Mirk's *Festial* P82 M4/HC)
- (4) And the Angel of the LORD called vnto Abraham out of heauen the second time, And said, By my selfe haue I sworne, saith the LORD, **for because** thou hast done this thing, and hast not withheld thy sonne, thine onely sonne, That in blessing I will blesse thee, and in multiplying, I will multiply thy seed as the starres of the heauen, and as the sand which is vpon the sea shore. (The angel of the LORD called to Abraham from heaven a second time and said, "I swear by myself, declares the LORD, that *because* you have done this and have not withheld your son, your only son, I will surely bless you and make your descendants as numerous as the stars in the sky and as the sand on a seashore.)(*The Old Testament* Authorized Version E2/HC)
- (5) **For because that** Saturne is of late so sterynge. (*Because* Saturn is lately so strange)(1400/*Mandeville*/OED)

For because construction was rare, appearing 4 times in the M4 (1420-1500) period, 7 times in the E1 (1500-1570) period and once in the E2 (1570-1640) period before it was no longer found. Relative rarity of the interim construction compared to *forþam þe* and *because* may indicate that the transition was met without resistance from the public and that the new causal marker caught on quickly. We discuss the grammaticalisation of *because* in the next section.

3.2 Grammaticalisation of *because*

Fischer et al. (2000: 286) defines grammaticalisation as "a process whereby a lexical item, with full referential meaning, developments into a grammatical marker". According to Hopper and Traugott (2003: 1), "grammaticalisation refers to that part of study of language change that is concerned with such questions as how lexical items and constructions come in certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions".

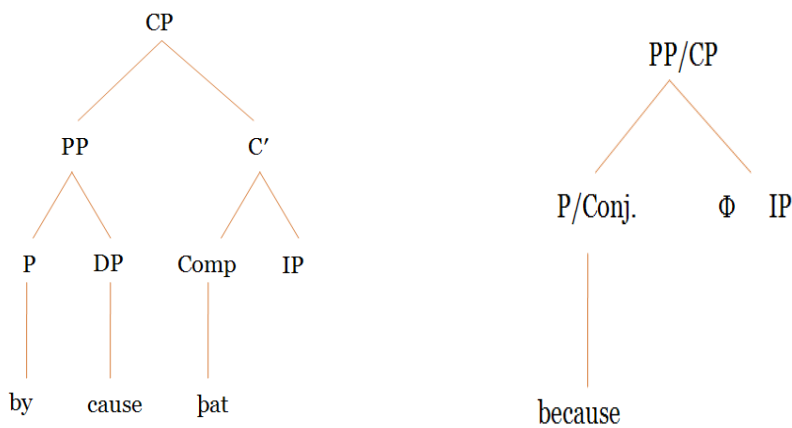
The case we have here is a grammatical item (PP *by cause*) becoming a more grammatical item (Conjunction *because*). In 3.2.1 some trees are shown to exem-

plify this development. And in 3.2.2, mechanisms of grammaticalisation are discussed in full. We commence with the tree structure analogy.

3.2.1 Tree structure analogy

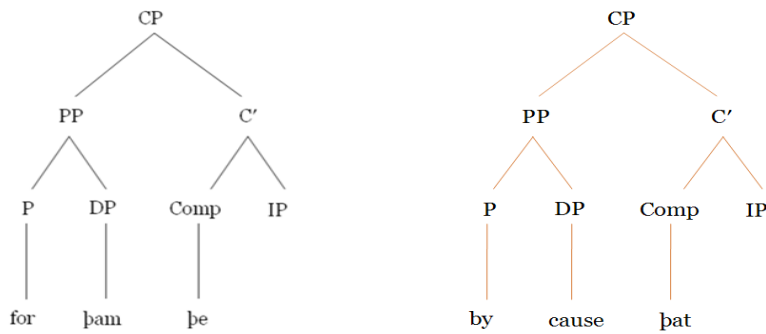
In the introduction, *because* is said to have developed from a phrase, *by + cause + (that)*. Tree on the left below is the internal make of *because* when it was first adopted. The appended *that* functions as a marker of subordination (see 3.1.2 above). Tree on the right below shows the grammaticalised structure of *because*, in which case it appears both as a PP (for example, *because of*) and as a conjunction. Look at the trees⁵ below:

Figure1: Changes in the tree structure of *because (that)*



The tree structure of *by cause þat* is analogous to that of its predecessor, *forþam þe*. Consider the following:

Figure2: The tree structure analogy between *forþam þe* and *because that*



⁵ DP in the trees refer broadly to both NP and determiner phrase.

On the Development of Because: A Corpus-based Study

Both *forþam þe* and *because* are in the form of a preposition followed by a substantive and a subordinating particle. With the exception of the subordinating particle *þe* and *þat*, trees bear striking resemblance. But knowing that even *forþam þe* had more occurrences of *þat* than *þe* in the ME period, it is fair to assume that *because* developed imitating the structure of *forþam þe*. For reference, occurrences of *þe* and *þat* with *forþam þe* are given below in Table6. The subordinating particle is varied in appearance but can be grouped around few characteristic patterns. *þe* (413 and *ðe* 156) is by far the dominant marker of subordination. But its numerousness is accounted for in the OE period only. Out of the 569 cases with the use of *þe* (and *ðe*) for the *forþam þe* constructions, the ME portion in the total frequency is only 48 instances (8.4%). But *that* and its equivalents (*ðat*, *þatt*, *þt*) are better choice for the ME turnouts: 48 instances out of 96 (50%) employ *that* as subordinating particle in the ME period. Table6 below summarises the finding:

Table 6: Occurrences of the subordinating particle with *forþam þe* (absolute figures)

Type	Number of Tokens		
	Total	OE	ME
<i>þe</i>	413	378	35
<i>ðe</i>	156	143	13
<i>þatt</i>	21	-	21
<i>þæt</i>	14	14	-
<i>þat</i>	11	-	11
<i>þt</i>	11	-	11
<i>ðæt</i>	10	9	1
<i>ðætte</i>	4	4	-
<i>ðat</i>	3	-	3
<i>þet</i>	1	1	-
<i>þætte</i>	1	1	-
<i>that</i>	1	-	1

One objection to the tree structure analogy is the seemingly reversed analytic to synthetic development of *because*. Against the gradual typological change of English towards analyticity in the Middle English period, *because* here is seen to become even more synthetic (*by cause* → *because*). But it may be that the loan word status of *because* has overridden the general development trend. This should be an interesting topic for further study. For now, we move on to features of grammaticalisation noted in the formation of *because*.

3.2.2 Mechanisms of Grammaticalisation

In Heine (2003: 578-9), grammaticalisation is said to involve four interrelated mechanisms:

- i. desemanticisation: loss in meaning content
- ii. extension: use in new contexts
- iii. decategorialisation: loss in morphosyntactic properties characteristic of the source forms, including the loss of independent word status
- iv. erosion: loss in phonetic substance

In the following, the four mechanisms of grammaticalisation are tested against the development of *because*.

A feature previously covered is (iv) erosion, loss in phonetic substance: phonetically, current causal connective *because* starts with *be-*, the reduced version of *by*. Table 4 in 3.1.1 above demonstrated that *by-* accounted for 65% of the total prepositional element in the M3 (1350-1420) period but later, more grammaticalised version employed *be-* 97% of the time in the E2 (1570-1640) period. This explains the gradual phonetic erosion of the preposition in the course of grammaticalisation.

Decategorialisation (iii) refers both to loss in morphosyntactic properties of the source form and loss of independent word status. First, in terms of loss in morphosyntactic property, look at the following examples, which exemplify the use of *because* with the noun *cause* still strongly exhibiting its nominal property, appearing with a determiner:

- (6) Thus I sawe howe cryste has compassyon of vs *for the cause of* synne, and ryght as I was before with the passyon of cryste fulfilled with payne and compassion. (M3/HC)
- (7) *Be þe cause* Pat þei scholde rise Erly, Vnto her reste went þei att nihte. (1386/OED)
- (8) Thei had in hem no shame nor drede *by the cause* thei wer so used. (1450/OED)

In the examples above, the definite article is placed before *cause*. Examples such as (6), (7) and (8) are found only until the earlier phase of grammaticalisation but are practically non-existent in the EME period (1500-1700). This means *cause* in *because* lost its nominal characteristics and the cohesion between *be-* and *cause* has become stronger. In addition, prepositional element *be-* indicate that the original preposition *by* has lost its word status and is attached to *cause* thereby forming *because*, which testifies another trait of decategorialisation.

Desemanticisation (i) and (ii) extension seem to go hand in hand with the development of *because*. In order to explain the loss of meaning content in *cause* and the use of *because* in a context where the dictionary meaning of *cause* is not the norm, direct reason relationship given by Quirk et al. (1985: 1103) is provided as evidence. In a monumental grammar reference book of Quirk and his col-

On the Development of Because: A Corpus-based Study

leagues, causal relationship marked by *because* can be divided into three basic categories:

Direct reason relationships

- Cause and effect: inherent objective connection in the real world
- Reason and consequence: speaker's inference of a connection
- Motivation and result: the intention of an animate being that has a subsequent result

Cause-effect, reason-consequence pairs are easier to accept as having directly stemmed from the root word *cause* in *because* construction. But the third category motivation-result makes a person think a little deeper for the connection. OED in fact lists this third sense as *because's* obsolete meaning (*for the purpose that, in order that*). Here is the example:

(9) They axed him *because* they might acuse hym. (1526 Tindale *Matt. xii. 10*/OED)

This extended use of *because* to account for motivation-result relationship can be attributed to desemanticisation (the semantically blurred sense of the literal phrase *by (the) cause*) and extension (the expanded meaning of the noun *cause*).

Another principle of grammaticalisation applicable to explaining the development of *because* is layering, which according to Hopper (1991:22) is “more than one form available in a language to serve similar or even identical functions”. One definite evidence of this principle is the fact that *forþam þe* and *because* existed in parallel over a certain period of time:

(10) But *forþi þat* þou schalt not erre in þis worching, and wene þat it be oþerwise þen it is, I schal telle þee a lityl more þer-of, as me þinkeþ. (M3/HC)

(11) In Egypt þere ben but fewe Forcelettes or castelles *be cause þat* the contree is so strong of himself. (M3/HC)

And a rather developed stage of layering is exemplified by the occurrences of the interim construction *for because that* (see 3.1.3 (3), (4) and (5) above), all be it fewer in number.

Fischer et al. (1992: 288) listed as parameters of grammaticalisation, ‘increase in cohesion’ and ‘decrease in variability’. In terms of cohesion, the readers are referred to Table 3 (Occurrences of merged and separate construction of *because*). As for the decrease in variability, Table 4 (Spelling variations of the prepositional elements) explains the facts well.

In this section, the development of *because* was accounted for in terms of grammaticalisation. It was interesting to observe that *because* largely followed the general mechanism and featured principles of grammaticalisation. Empirical data from the HC served as a valuable source to elucidate the process. The frequency

alone cannot explain much but data findings combined with the analysis helped answer questions on the development of a connective.

4. Conclusion

This paper attempted to address the development of the causal connective *because*. Studies to date have mainly focused on the syntactic structure of the connected clause or semantics of causality but not the connective itself. The value of this paper is on the effort to explain the internal construction of *because* per se.

In a historical perspective, the paper followed the steps *because* took from its initial appearance to stabilization. Findings revealed that the development of *because* reflected features such as increase in cohesion, decrease in variability and other mechanism of grammaticalisation. The tree structure analogy of *forþam þe* and *because* showed that *because* syntactically imitated its predecessor *forþam þe*. The unique status of *because* as a loan word justified its seemingly backward evolution.

The study of one causal connective provided an opportunity for close individual analysis. The value of this study will be augmented if the explanatory adequacy gained here can be expanded to answer for other grammatical markers and broader linguistic phenomenon as well. And in doing so, evidence from the computerised corpora will render results to be more convincing.

References

- Fischer, Olga – Ans van Kemenade – Willem Koopman – Wim van der Wurff
2001 *The Syntax of Early English*. Cambridge: Cambridge University Press.
- Fisiak, Jacek – M. Krygier (edd.)
1998 *Advances in English Historical Linguistics*. Berlin: Mouton de Gruyter.
- Heine, Bernd
2003 "Grammaticalization" in: *The Handbook of Historical Linguistics*. Oxford: Blackwell.
- Hopper, Paul
1991 "On Some Principles of Grammaticalization", in: Traugott and Heine (edd.), 17-35
- Hopper, Paul – Elizabeth Closs Traugott
2003 *Grammaticalization*. Cambridge: Cambridge University Press.
- Kitis, Eliza
2006 "Causality and Subjectivity: The causal connectives of Modern Greek", in: Hanna Pishwa (ed.), 223-267.

On the Development of Because: A Corpus-based Study

- Pishwa, Hanna
2006 *Language and Memory: Aspects of Knowledge Representation*. Berlin– New York: Mouton de Gruyter.
- Quirk, Randolph – Sidney Greenbaum – Geoffrey Leech – Jan Svartvik
1985 *A Comprehensive Grammar of the English Language*. London-New York: Longman.
- Rissanen, Matti
1998 "Towards an integrated view of the development of English: Notes on causal linking", in: Jacek Fisiak and M. Krygier (edd.), 389-406.
- Sweetser, Eve E.
1990 *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs – Bernd Heine (edd.)
1991 *Approaches to Grammaticalization Vol I: Focus on Theoretical and Methodological Issues*. Amsterdam-Philadelphia: John Benjamins.
- Van Geldederen, Ellen
2007 "Accelerated Grammaticalization in the Peterborough Chronicle", in: Alex Bergs and Janne Skaffari (eds.), 93-110.

Appendix A: Occurrences of *forþam þe* in the Helsinki Corpus

No.	Forms/Period	O1	O2	O3	O4	M1	M2	M3	M4
1	for ðam ðe			4					
2	for ðam þe				1				
3	for ðy ðe	1							
4	for ðæm þæt		1						
5	for ðæm þe		2	5					
6	for ða ðe		1	1					
7	for ðan ðe		1	40	5	1			
8	for ðan þæt					1			
9	for ðan þe			9	7	1			
10	for þam ðe				4				
11	for ði þæt			3					
12	for ði ðat					3			

Ji Won Lee

No.	Forms/Period	O1	O2	O3	O4	M1	M2	M3	M4
13	for ði ðe					1			
14	for ðon þe			1					
15	for ðy þe		1						
16	for þa þe					2			
17	for þæm þe		7	1					
18	for þæm ðæt		1						
19	for þam ðe					3			
20	for þan ðe			2					
21	for þam þe				1	6			
22	for þan þe		1	12	11	21			
23	for þon þe		10	5					
24	for þon ðe		1						
25	for þi þe				1				
26	for þi þt					4			
27	for þy þt					1			
28	forðæm þæt		3						
29	forðæm þe		6	3					
30	forðæm ðæt		7						
31	forðæm ðætte		4						
32	forðæm ðe		2						
33	forðæmðe		38						
34	forðamðe		2						
35	forðam þæt				1				
36	forðam þe			26	11				
37	Forðæmþe		10						
38	Forðamþe				1				
39	forðan ðe			19		2			
40	forðan þe			9	4				
41	forðan þet				1				
42	forðanþe			5					
43	forðon ðe		2	5					

On the Development of Because: A Corpus-based Study

No.	Forms/Period	O1	O2	O3	O4	M1	M2	M3	M4
44	forðon þætte			1					
45	forðon þe		8	11	1				
46	forðonðe		7	3					
47	forðonþe			2					
48	forðy ðæt		1						
49	forðy þe		2	1					
50	forðyðe		2						
51	forþæmþe		2						
52	forþam þe		1	42	12	2			
53	forþam ðe		1			3			
54	forþamþe		1		1				
55	forþan ðæt					1			
56	forþan ðe		1			3			
57	forþan þe		1	9	7	2			
58	forþanþe			23	11				
59	forþi þæt				2				
60	forþi þat					1			
61	forþi þe				1				
62	forþi þe					1			
63	forþi þt					6			
64	forrþi þatt					21			
65	forþi þat						1	4	
66	forþi þat								3
67	forþon þæt			4	1				
68	forþon þe		8	51	20				
69	forþonþe			9					
70	forþy ðe		1						
71	forþy þe		1				1		
72	forþi that								1
73	forþy þat								1
74	forþy þe								1

Ji Won Lee

total	O1	O2	O3	O4	M1	M2	M3	M4
absolute figure	1	137	306	104	86	2	4	6
corpus size	2,190	92,050	251,630	67,380	113,010	97,480	184,230	213,850
per 100,000 words	46	149	122	154	76	2	2	3

Appendix B: Occurrences of *because* in the Helsinki Corpus

No.	Forms/Period	M3	M4	E1	E2	E3
1	by cause that	13	3			
2	be cause þat	2				
3	by-cause þat		1			
4	by-cause that		1			
5	be-cause þat		2			
6	be-cause that	1				
7	bicause that			2		
8	because that		3	7	7	1
9	bycause that			3		
10	because þat	3		3		
11	becawse that			1		
12	be caus				5	
13	be cause		2			
14	by cause	1	16	9		
15	by caus				1	
16	by-cawse		1			
17	by-cause		1			
18	be-cawse		2			
19	be-cause		5			
20	becavsse		1			
21	becawis		1			
22	becawse			2	2	
23	becase				1	
24	because		7	91	191	115

On the Development of Because: A Corpus-based Study

No.	Forms/Period	M3	M4	E1	E2	E3
25	bycause		3	26	6	
26	bicause		2	28	1	
27	be caus of				2	
28	by cause of	8	6	1		
29	bi cause of	1				
30	be cause of	3	3			
31	by-cause of		2			
32	be-cawse of		1			
33	be-cause of		1			
34	becawse of			1		
35	bycause of		2	2		
36	bicause of	2		2		
37	because of			9	14	8
	absolute figure	34	66	187	230	124
	corpus size	184,230	213,850	190,160	189,800	171,040
	per 100,000 words	19	31	98	121	72