LUND UNIVERSITY

Stylistic Fronting in corpora

Sigurðsson, Halldor Armann

Stylistic Fronting in corpora

Halldór Ármann Sigurðsson
Lund University

Stylistic Fronting (SF) fronts various types of non-subjects to the preverbal position in subjectless clauses. With the exception of Icelandic and Faroese, SF has disappeared from Scandinavian. It is commonly assumed that even in Icelandic it is formal and old fashioned, indicating that it might be on its way out. However, this assumption has not been supported by frequency surveys. This paper studies the distribution and frequency of Stylistic Fronting in two large language corpora, *Timarit.is* and the Internet. The results support the common assumption that SF is on the retreat. Nevertheless, the survey also highlights that both this change is proceeding slowly. The study also shows that Google Search can be used as a research tool in linguistics – no small advantage.

Keywords: expletive insertion, Extended Projection Principle, Google Search, impersonal clauses, Stylistic Fronting, relative clauses, *Timarit.is*, verb-initial adverbial clauses, word order frequencies

1. Introduction[*]

Icelandic *Stylistic Fronting*, **SF**, was first systematically (and influentially) studied in Maling 1980[1] and has been discussed in many works since, including two doctoral dissertations (Franco 2009, Angantýsson 2011).[2] Holmberg (2000:445) succinctly describes it as follows:

---

[*] This is my own (clumsy) formatting, with the same page numbers as in the published JB version. The copyright of the ideas and scientific results presented here is the "property" of mine (which I gladly share with all others on our rapidly shrinking globe).

[1] I am grateful to Anders Holmberg, Ásgrímur Angantýsson, Irene Franco, Jim Wood, Valéria Molnár, and Verner Egerland for comments and discussions and to the editors/reviewers of this volume for generous and valuable remarks and corrections. Thanks also to the Landsbókasafn – Háskólabókasafn staff for answering my questions about *Timarit.is*.

[2] See also, e.g., Rögnvaldsson & Thráinsson 1990, Jónsson 1991, Falk 1993, Kosmejer 1993, Holmberg & Platzack 1995, Holmberg 2000, Hrafnbjargarson 2004, Holmberg 2006, Thráinsson 2007, Ott 2009, Wood 2011, Thráinsson et al. 2015, Angantýsson 2017.

… stylistic fronting is an operation that moves a category, often but not always a single word, to what looks like the subject position in finite clauses where that position is empty, namely, in subject relatives, embedded subject questions, complement clauses with an extracted subject, and various impersonal constructions.

Some typical examples are given in (1).[3]

(1) a.   Eins og þeir vita   [sem **lesið** hafa *t* bókina   ] þá …
          as   they know who   read   have   book-the   then …
          ’As they who have read the book know, then …’

gthg.blog.is/blog/gthg/entry/202600/ – March 8, 2010

     b.   … ég fór   aftur til læknis [eins og **um**   var talað *t*] og …
          … I   went   again to doctor as   about was talked   and …
          ’(Anyway) I went to see the doctor again, as had been agreed upon, and …’

blogs.myspace.com/index.cfm?fuseaction=blog.view...blogId – March 8, 2010

     c.   **Sagt** er *t* [að fegurðin komi   að   innan ...]
          said   is   that beauty-the comes from   inside
          ‘It is said that the beauty comes from the inside …’

asarut.blogcentral.is/ – March 8, 2010

The central traits of SF are listed in (2) (see, e.g., Maling 1980, Jónsson 1991, Holmberg 2000, 2006, Thráinsson 2007:352ff., Angantýsson 2011).

(2) a.   The fronted element: SF fronts a non-subject, usually a small (one word) category
     b.   Precondition: SF can only apply in clauses with a "subject gap"[4]
     c.   Landing site: SF seemingly moves a category into the subject gap
     d.   Locality restriction: SF usually fronts the SF candidate that is structurally closest to the subject gap
     e.   Domain(s):   e1. SF applies in finite clauses only
                                 e2. SF is strictly clause-bounded
                                 e3. SF is common in (certain) subordinate clauses[5]

---

[3] The position where the stylistically fronted element has been moved from is indicated by *t* ("trace").

[4] But see Hrafnbjargarson 2004 for a different understanding of the subject gap requirement. For a different understanding of the landing site issue (2c), see Sigurðsson 2010.

[5] As seen in (1c), SF occurs in impersonal main clauses, but it does so much less frequently than in impersonal subordinate clauses. Of the first 50 examples in *Timarit.is* of *Farið/farið er að*

The categories moved by SF are heterogeneous: commonly adverbs, participles or particles. Maling 1980 (see also Jónsson 1991) analyzed fronting of all (non-subject) maximal categories as topicalization, even in clauses with a subject gap, while other studies (e.g. Holmberg 2000) take the subject gap to be the distinguishing factor, thus assuming that SF comprises movement of maximal categories as well as of smaller categories in the presence of a subject gap (see the overview in Thráinsson 2007:369). I will adopt this latter understanding here. Maling (1980) argued that SF is amenable to an *accessibility hierarchy*, movement of the negation *ekki* 'not' taking precedence over movement of a predicate adjective, which in turn takes precedence over movements of particles and past participles (*ekki* > predicate adjective > particle/participle). However, the "formulation of the hierarchy is controversial" (Holmberg (2006:537) and the relative accessibility of other SF categories remains to be scrutinized (various classes of adverbials, infinitives, and stranded prepositions in extraction domains).

Jónsson (1991) argued that the acceptability of SF is partly controlled by minimality, the moved category usually being closer the subject gap than any other potential SF candidates, and Holmberg (2000:463) developed and refined the relevant locality notion: Where A c-commands both B and C, B is structurally closer to A than is C if B asymmetrically c-commands C. Usually, the structurally closest candidate is also linearly closest to the subject gap. However, on Holmberg's understanding, a head and its complement are equally close to (equidistant from) the subject gap (there being a symmetric and not an asymmetric c-command relation between sister nodes). Given that, a participle and its complement should be equally amenable to SF, but, as we will see, that is not borne out, the applicability of SF being affected by the properties of both the potential "mover" and its "neighbors".

In his influential *Linguistic Inquiry* article on SF Holmberg (2000:446) argued that it is EPP-driven, like expletive insertion:[6] "the element moved by SF functions as a pure expletive in its derived position … it alternates with the special expletive *það* in some cases. The trigger of the movement is a version of the Extended Projection Principle (EPP)." However, SF does not seem to be a triggered movement in any obvious sense. Indeed, it is not clear whether or in what sense it is a single phenomenon. There are two rather different SF contexts, as sketched in (3) (which

---

(lit. 'begun is to' ≈ 'people/someone has begun to'), three are found in main clauses, 47 in subordinate clauses. I will set SF in main clauses aside.

[6] EPP = Extended Projection Principle, i.e., the requirement that the canonical subject position (Spec,TP) should be spelled out (see Holmberg 2000:447).

was the main reason why Sigurðsson 2010 claimed that that SF and insertion of expletive *það* 'there, it' are subject to different conditions).

(3) a. Clauses with a **subject trace** ^{ok}V1[7] ^{ok}SF   *$það$-V[8]
          (i.e., clauses relativized/extracted from)
     b. Clauses with a **non-trace subject gap** $^{??/ok}$V1 ^{ok}SF $^{ok}það$-V
          b1. Subjectless impersonal clauses
          b2. Clauses with a "late" subject

For examples, see (5)–(7) below. In addition, SF has a different stylistic value in different constructions. It has been suggested that SF in general has a formal flavor (e.g., Angantýsson 2009, 2011, 2017, Sigurðsson 2010, Wood 2011), but this does not apply to certain impersonal clause types, where SF is particularly frequent (see sections 5–6).

Claims that SF is formal and old fashioned, indicating that it might be on its way out of the language, have not been substantiated or supported by frequency surveys in large written language corpora, understandably so as such corpora have not been accessible until recently. This paper purports to "remedy" this by studying the distribution of SF across the different domains in (3a) and (3b1) in two corpora: *Timarit.is* and the World Wide Web. The main purpose of the study is to provide some reliable data indicating how frequent SF is in these domains (as compared to V1 and *það*-V), in everyday written Icelandic as found in newspapers and other media. As it turns out, the survey shows that SF has a strong foothold in potential SF contexts, even though the data suggest that it is presently losing ground against V1 in subject relatives and against *það*-V in impersonal clauses. The applicability of SF seems to be affected by a number of factors (in addition to the ones listed in (2)), including clause type (and/or complementizer type), the properties of the potentially fronted category, and the presence and properties of other SF "contenders" in the same clause.

2. *Timarit.is* and Google Search

Timarit.is (http://timarit.is/) is an open access digital library hosting newspapers and magazines published in Iceland (and the Faroe Islands and Greenland). It contains almost 4,900,000 photographed pages (July 22, 2015), easily searchable,

---

[7] V1 = non-application of SF or *það*-insertion, yielding a verb-initial order; $^{??/ok}$ indicates variable acceptance, depending on constructions, contexts, and individuals.

[8] This is a slight simplification. *Það*-insertion is more sharply ungrammatical when the extracted/relativized argument is a subject than when it is a non-subject.

from 972 different sources (newspapers, magazines of various sorts, pamphlets, brochures, etc.). *Timarit.is* is thus extensive, considering the size of the Icelandic linguistic society. Information on the number of words it contains is not available, but by searching for individual words one can get some idea about its size. Thus, searching for the negation *ekki* (July 1, 2015) yields almost 3,600,000 (3,6m) results.[9] The bulk of the photocopied texts come from the second half of the 20th century, containing almost 2,2m, ca 61%, of the occurrences of the negation in the entire corpus, but the earliest example found for the negation was from the year 1816.[10] On the negative side, *Timarit.is* is not lemmatized, it counts results in terms of the number of pages containing the search string and not in terms of the number of occurrences of the string (meaning that multiple occurrences of a string on one and the same page just count as one occurrence), and it counts repeated occurrences of the same text on different pages (e.g., advertisements) as separate independent occurrences.[11] This can obviously distort search results for individual words, but it has limited effects when one searches for strings that contain three or more words (as the search strings in the present study). In short, there is every reason to believe that search in *Timarit.is* gives a fairly reliable picture of word order pattern frequencies in the texts in the corpus. It is a useful tool for the purposes of the present study.

Google Search is a less reliable tool, with properties that limit its usefulness for linguistic research. "Googleology is bad science" is the title of Kilgarriff 2007, and that is certainly true if Google Search is carelessly used. The number of hits for any given search string is unreliable and varies greatly from time to time, even overnight (see Rayson et al. 2012, Gatto 2014); one of the reasons behind this is that pages that are low ranked by Google's (secret) algorithms disappear from the overt web down into the so-called deep web. Also, the number of hits is hugely overestimated

---

[9] For comparison, searching for the Swedish negation *inte* 'not' (July 1, 2015) in the extensive Språkbanken (http://spraakbanken.gu.se/swe) gives just about 11,7m results. The tagged corpus Mörkuð íslensk málheild (http://mim.hi.is/) contains 25m tokens, thereof 211,173 tokens of *ekki* (0,8%).

[10] The "temporal distribution" of *ekki* in the corpus (July 1, 2015):

– 1815:	0	
1816–1850:	4,402	0,1%
1851–1899:	77,780	2,2%
1900–1949:	692,900	19,4%
1950–1999:	2,185,460	61,3%
2000–2015:	607,363	17,0%
Total:	**3,567,905**	

Other frequent words, such as *og* 'and' and *að* 'that', show similar distribution patterns over time.

[11] Both these drawbacks are shared by Google Search.

as any string on a webpage is recounted whenever the page is updated, and many pages are updated on a daily basis or even many times a day. However, if one opts for googling within a given period (in the "search tools") the numbers become more stable and credible.[12] Thus, searching (July 6, 2015) for the V1 string *sem hafa verið* 'who/that have been' vs. the SF string *sem verið hafa* gave the results in Table 1. No hits were found prior to 1970.

Table 1. Google Search results (July 6, 2015) for different periods for *sem hafa verið* and *sem verið hafa* (in terms of number of pages).

	sem hafa verið (V1)	sem verið hafa (SF)
Unlimited	389,000	85,500
1970.01.01-2000.01.01	811	81
2000.01.01-2010.01.01	15,000	729
2010.01.01-2015.07.01	25,400	695
2000.01.01-2015.07.01	34,300	1,220
2005.07.01-2015.07.01	31,700	974
1970.01.01-2015.07.01	34,800	1,220

These numbers suggest that Google counts are biased such that the algorithms tend to 'skip' pages the more the farther back in time they were uploaded. Nevertheless, after repeated checks (2010, 2013, 2014, 2015), I can confirm that Google Search results within a given period are largely stable and seem also to be realistic in the sense that they come much closer to reflecting the actual number of independent occurrences of the searched strings on the Internet than does unlimited search.[13] The results in the present study indicate that Google are using some effective algorithms to filter out uploading repetitions of one and the same page when one searches within a specific period.

Google Search has obvious drawbacks as a research tool but it also has clear advantages. The size of the Web is enormous and searching it with Google yields fast results and costs nothing. These are no small advantages in an academic world that is constantly short of resources. In addition, Google Search is a superb tool to find out whether some particular word order is very rare or even non-existent in

---

[12] It took me a long time and many attempts to discover this (trial and error). In an earlier attempt to use Google to study the frequency of SF (Sigurðsson 2013) I used the number of pages made visible by Google (by browsing all the way to the last visible page), but that is only a good method for rare constructions.

[13] The searches in Table 1 were repeated on July 31, 2015, showing fluctuation within the limits of 10%, with the exception of the unlimited search for *sem hafa verið*, which yielded 606,000 hits.

published texts. All in all, it seems to me that the pros of carefully using the Web as a corpus in a study like the present one outweigh the potential cons by far.

The World Wide Web and *Timarit.is* are dissimilar corpora in many ways. The texts in *Timarit.is* are from newspapers and other edited sources; such texts are of course published on the Internet too, but it also contains large amounts of unedited texts (blogs, etc.). One can thus expect to find less formal texts on the Web than in *Timarit.is*. In addition, as already mentioned, the bulk of the *Timarit.is* texts are from the second half of the 20th century and thus older than most of the Internet texts. Table 2 shows the "temporal distribution" of *sem hafa verið* 'who/that have been' (searched on July 6, 2015) in both corpora.

Table 2. The distribution of *sem hafa verið* 'who/that have been' over time in Google and *Timarit.is.*

	Google		Timarit	
	#	%	#	%
Prior to 1900	0		493	1,0%
1900–1949	0		5,185	10,7%
1950–1999	811	2,3%	28,574	59,0%
2000–2015(01.07)	34,300	97,7%	14,160	29,2%

My purpose by searching both the Internet and *Timarit.is* is to study two corpora that are partly dissimilar and complementary but can nevertheless be characterized as reflecting "everyday written Modern Icelandic". Given the different nature of many of the texts in these corpora this characterization might seem questionable. However, both corpora contain large amounts of (mostly) non-fictive texts meant for everyday consumption for the general public, so in that perspective the characterization is warranted. Even so, it is clear that the texts in the corpora reflect many "realities", both across and within the corpora. An intriguing question is how these different "realities" relate to the "realities" reflected in informant studies, as in Angantýsson 2011, Thráinsson et al. 2015 and Angantýsson 2017. I will make some comparisons of the results of these studies and my survey.[14]

---

[14] The spoken language corpora (Talmál on http://corpus.arnastofnun.is/) studied by Wood (2011) are too small for my purposes (Wood managed to make use of them by searching for general patterns rather than for specific strings and by applying fine grained regression analyses). For example, they contain only 115 instances of the string *hafa verið* 'have been' (83 in Alþingisumræður, 21 in Ístal, 3 in Samtöl, 8 in Viðtöl) (one can only search for strings containing one or two words; of the 115 *hafa verið* occurrences only 16 were *sem hafa verið*). In comparison, *Timarit.is* contains 917,605 instances of this string (July 16, 2015) and searching for it on Google for the period July 1, 2005 to July 1, 2015 gave 170,000 hits. The string *verið hafa* gave zero hits in

3. Two different Stylistic Fronting contexts

As mentioned above, three word order types compete in potential SF domains, namely:

(4) a. V1 (verb-initial) order: neither SF nor insertion of expletive *það* takes place
     b. SF
     c. *Það*-insertion

However, as indicated in (3), these types are not equally available across the different SF contexts: (3a), clauses with a subject cap containing a trace, and, (3b), clauses with a subject gap that does not contain a trace. While SF is available in both contexts, *það* is excluded in the trace context.[15] The examples in (5)–(7) illustrate this (the underline indicates a subject gap of some sort).

(5) A. Clauses with a subject trace:
     a. … fyndnasta bók   [sem __   hefur verið skrifuð].
         funniest   book   that   has   been   written
         '… the funniest book that has (ever) been written.'
                                                   www.123.is/thorkell/blog/month/200711/ – March 11, 2010
     b. … fyndnasta bók   [sem **skrifuð** hefur verið *t*].
         '… the funniest book that has (ever) been written.'
                                                   www.thjodmal.is/index.php/page/30.html – March 9, 2010
     c. * … fyndnasta bók   [sem **það**   hefur verið skrifuð].
         funniest   book   that   there   has   been   written

(6) B. Clauses with a non-trace subject gap.
     B1. *Subjectless* impersonal clauses (here illustrated with impersonal passives):
     a. … þegar __   verður komið í …
         … when   will_be come   into
         '… when I/we/they will get into …'
                                                   sigurjonn.blog.is/blog/sigurjonn/?offset=10 – March 11, 2010

Talmál (vs. 22,369 in *Timarit.is* and 1,260 on Google, with the same premises as for *hafa verið*). Like the Talmál corpus, the tagged written language corpus Mörkuð íslensk málheild (http://mim.hi.is/) is a valuable tool for many purposes, but it is also too small for the purposes of my study (it contains 9,288 vs. 64 ocurrences of the strings *hafa verið* and *verið hafa*). For clarity:

    hafa verið: 917,605 in *Timarit.is*, 170,000 on Google, 9,288 in mim.hi.is, 115 in Talmál.

    verið hafa: 22,369 in *Timarit.is*, 1,260 on Google, 64 in mim.hi.is, 0 in Talmál.

[15] And V1 is sometimes degraded in the non-trace context.

b. … þegar **komið** verður *t* heim …

    when come will_be home

'… when I/we/they will get (back) home …'

c. … þegar **það** verður komið heim …

    … when there will_be come home

'… when I/we/they will get (back) home …'

(7) B. Clauses with a non-trace subject gap.

B2. Clauses with a *late subject*:

a. … þegar __ verða komnir <u>bjórkælar</u> við nammibarinn á …

    … when will_be come.PL beer_coolers at candybar.the at

'… when beer coolers will have been introduced at the candybar at …'

b. … þegar **komnir** verða *t* <u>hvolpar</u> …

    when come.PL will_be.3PL puppies

'… when puppies will have arrived/come into being …'

c. … þegar **það** verða komnir <u>hvolpar</u> …

    when there will_be.3PL come.PL puppies

'… when puppies will have arrived/come into being …'

I will study and discuss clauses with a subject trace (subject relatives) in section 4, turning to clauses with a non-trace subject gap in section 5. For practical reasons, the scope of both sections is limited to the most typical types of clauses with a subject trace vs. a non-trace subject gap, and thus the late subject type in (7B2) falls outside the scope of the study.

4. Clauses with a subject trace ("personal" clauses)

As we have seen, in clauses with a subject trace, SF competes with only V1, expletive *það* being excluded.[16] This is illustrated further in (8)–(10) (from Sigurðsson 2010:179–180).

(8) a. * Þetta er bók sem **það** hefur verið skrifuð um einmitt þetta.
        this is book that there has been written about exactly this
     b. Þetta er bók sem **skrifuð** hefur verið *t* um einmitt þetta.
     c. Þetta er bók sem __ hefur verið skrifuð um einmitt þetta.
        'This is a book that has been written about exactly this.'

(9) a. * Veit hún hver **það** hefur skrifað um þetta?
        knows she who that has written about this
     b. Veit hún hver **skrifað** hefur *t* um þetta?
     c. Veit hún hver __ hefur skrifað um þetta?
        'Does she know who has written about this?'

(10) a. * Hver heldur þú að **það** hafi skrifað um þetta?
        who think you that there has written about this
     b. Hver heldur þú að **skrifað** hafi *t* um þetta?
     c. Hver heldur þú að __ hafi skrifað um þetta?
        'Who do you think has written about this.'

In the following I will present a study of the frequency of SF and V1 in clauses with a subject trace. For practical reasons, the study is limited to relative clauses introduced by *sem* 'that, which, who', and where the potential SF element usually is a past participle. Many of the Google searches were conducted on September 25,

---

[16] Faroese differs from Icelandic in this respect, expletive *tað* being an option in, e.g., subject relatives (see Angantýsson 2011, chapter 5.3). Given the analysis in Sigurðsson 2010, this suggests that *tað* differs from *það* in not blocking a trace from matching abstract features in the C-domain (C/edge linkers in the sense of Sigurðsson 2011), perhaps via or in chain with the expletive. I will not discuss this here, though (as it would require too a leangthy explication of a technically detailed approach). Also, as discussed in e.g. Rögnvaldsson 1984, Magnússon 1990, and Rögnvaldsson & Thráinsson 1990, some factors other than just the operator–variable (i.e., the C/edge–trace) relation may affect the acceptability of expletive *það* in relatives. Thus, while *það* is impossible when the variable is a subject, it is commonly well-formed when the variable is a prepositional complement or an adverbial. I must put this aside here.

2014 searching for results within the date range from January 1, 2004 to January 1, 2014, while many of the *Timarit.is* search was conducted on September 3, 2014 and searched the whole corpus. In addition, I made a number of searches in July and August 2015 (as will be pointed out when clarification is needed).

A number of my examples with the finite auxiliary *hafa* 'have' plus a main verb participle are given in (11)–(13).[17]

(11) a.   sem __   hafa   verið
          that   have   been
     b.   sem **verið** hafa   *t*
(12) a.   sem __   hafa   farið
          that   have   gone
     b.   sem **farið** hafa   *t*
(13) a.   sem __   hafa   lesið
          that   have   read
     b.   sem **lesið** hafa   *t*

The results for these examples are shown in Table 3.[18]

The informant surveys of Angantýsson (see 2011:153; also 2017) and of Thráinsson et al. (2015:284ff.) show that young informants are generally more likely than older ones to question or reject SF in subject relatives (the acceptance rate nevertheless being roughly 40-65% for the youngest informants). It would thus seem that SF in subject relatives is losing ground in the present day language. As the Google texts in my survey are more recent than the bulk of the *Timarit.is* texts, the results in Table 3 seem to yield support to that conclusion. A good method to shed some light on this issue is to check the frequency of V1 vs. SF for whole paradigms

---

[17] The examples in (11) stand out, showing a much lower frequency of SF (see Table 3) than do any of the other searched relative clause strings. The reason is that most of the hits in question contain passive *verið*. As discussed in Jónsson 1991 (see also, e.g., Holmberg 2000, Thráinsson 2007, Angantýsson 2017), the passive auxiliary usually resists SF. As we will see, progressive *vera* 'be (doing)' behaves very differently from the passive auxiliary in this respect.

[18] The frequencies of V1 and SF in these and my other results in this section are only representative for the contexts searched for (three word strings with *sem*–verb–participle and *sem*–**participle**–verb). A quick check indicates that most other types of subject relatives do not apply SF of participles, instead being V1 or fronting other categories than participles, understandably so, as most clauses do not contain any participle. Searching (July 31, 2015) for simple *sem* __ *eru þar* 'who/that are there' and *sem þar eru* yielded 1,810 V1 vs. 19,574 SF hits in *Timarit.is* (91,5% SF). The corresponding numbers for Google (July 1, 2005 – July 1, 2015) were 1,020 V1 vs. 6,060 SF hits (85,6% SF). For *sem* __ *eru á Íslandi* 'who/that are in Iceland' vs. *sem á Íslandi eru* the *Timarit.is* numbers were 70 V1 vs. 100 SF (58,8%), whereas the Google numbers were 537 V1 vs. 55 SF (9,3%).

Table 3. Results (in September 2014) in Google (for the period January 1, 2004 to January 1, 2014) and *Timarit.is* (till September 3, 2014) for the examples in (11)–(13).

	Google		Timarit	
	#	%SF	#	%SF
V1: sem __ hafa verið	24,600		46,738	
SF: sem **verið** hafa	1,680	6,4%	14,101	27,7%
V1: sem __ hafa farið	2,220		4,268	
SF: sem **farið** hafa	2,170	49,4%	6,335	59,7%
V1: sem __ hafa lesið	284		1,444	
SF: sem **lesið** hafa	150	34,6%	2,433	62.8%
V1 totals	27,104		52,450	
SF totals	4,000	12,9%	22,869	30,4%

of verbs and participles. I checked this (in September 2014) for the indicative verb forms *er*, *var*, *hefur verið*, *hafði verið* 'is, was, has been, had been' plus the participle forms of *skrifa* 'write' in the singular neuter, feminine, and masculine (*skrifað*, *skrifuð*, *skrifaður*, respectively). The strings searched for were thus the ones in (14) (24 in number).

(14) a. sem __ er/var/hefur verið/hafði verið   skrifað/skrifuð/skrifaður   V1
          that   is/was/has been/had been   written.SG.NT/FEM/MASC
      b. sem **skrifað/skrifuð/skrifaður** er/var/hefur verið/hafði verið   *t*   SF

The results for the individual examples are given in (15).

			Google	*Timarit.is*
(15)	a1.	V1: sem __ er skrifað	233	429
	a2.	SF: sem **skrifað** er	418	1,993
	b1.	V1: sem __ var skrifað	110	294
	b2.	SF: sem **skrifað** var	261	1,393
	c1.	V1: sem __ hefur verið skrifað	229	185
	c2.	SF: sem **skrifað** hefur verið	154	922
	d1.	V1: sem __ hafði verið skrifað	5	21
	d2.	SF: sem **skrifað** hafði verið	22	118
	e1.	V1: sem __ er skrifuð	116	392
	e2.	SF: sem **skrifuð** er	182	830
	f1.	V1: sem __ var skrifuð	124	227
	f2.	SF: sem **skrifuð** var	228	617
	g1.	V1: sem __ hefur verið skrifuð	32	41
	g2.	SF: sem **skrifuð** hefur verið	73	623
	h1.	V1: sem __ hafði verið skrifuð	0	5
	h2.	SF: sem **skrifuð** hafði verið	2	14

i1.	V1: sem ___ er skrifaður	55	101
i2.	SF: sem **skrifaður** er	153	240
j1.	V1: sem ___ var skrifaður	19	44
j2.	SF: sem **skrifaður** var	38	85
k1.	V1: sem ___ hefur verið skrifaður	5	10
k2.	SF: sem **skrifaður** hefur verið	9	47
l1.	V1: sem ___ hafði verið skrifaður	1	7
l2.	SF: sem **skrifaður** hafði verið	1	7

These results are summarized in Table 4.

Table 4. Results for the strings in (14)/(15) in Google (January 1, 2004 to January 1, 2014; conducted September 25, 2014) and *Timarit.is* (till September 3, 2014).

	Google		Timarit	
	#	%SF	#	%SF
V1 totals	929		1,756	
SF totals	1,541	62,4%	6,889	79,7%

With the exception of (15c1) on Google and the insignificant (15l1/2), SF is more or even much more common than V1 in all cases, not only in *Timarit.is* but also and perhaps more surprisingly on the Internet. Nevertheless, as also in (11)–(13), the SF frequency is lower in my Internet results than in the *Timarit.is* results, raising the question of whether this difference arises because the Web texts are generally more recent or because they are commonly less formal than the *Timarit.is* texts. To shed some light on this issue I checked the frequency of V1 *sem er skrifað* 'that is written.NT.SG' vs. SF *sem skrifað er* over time in *Timarit.is*. The search was conducted in July 2015 (so the results are not exactly the same as in (15a1/2)). The results are presented in Table 5.

Table 5. *Timarit.is* results for *sem er skrifað* '*that is written*.NT.SG' vs. *sem skrifað er* in different periods (search conducted July 3, 2015).

	–1949		1950–1999		2000–2015	
	#	% SF	#	% SF	#	% SF
V1: sem ___ er skrifað	44		263		131	
SF: sem **skrifað** er	408	90,3%	1289	83,1%	333	71,8%

These results suggest that even within *Timarit.is* the frequency of SF in subject relatives is decreasing over time. Other combinations of auxiliaries and common participles yield similar results. This is exemplified and illustrated in Table 6.

Table 6. *Timarit.is* results for different periods (search conducted July 3, 2015) for V1 vs. SF strings: *sem er tekið* 'that is taken', vs. *sem tekið er*; *sem hefur tekið* 'that has taken', vs. *sem tekið hefur*; *sem er farið* 'that is gone' vs. *sem farið er*; *sem hefur farið* 'that has gone' vs. *sem farið hefur*.

	–1949		1950–1999		2000–2015	
	#	% SF	#	% SF	#	% SF
V1: sem __ er tekið	91		424		210	
SF: sem **tekið** er	1,627	94,7%	4,833	91,9%	1,900	90,0%
V1: sem __ er farið	119		456		249	
SF: sem **farið** er	1,882	94,1%	6,914	93,8%	3,397	93,2%
V1: sem __ hefur tekið	155		2,669		1,364	
SF: sem **tekið** hefur	289	65,1%	3,781	58,6%	979	41,8%
V1: sem __ hefur farið	80		2,575		1,617	
SF: sem **farið** hefur	376	82,5%	5,784	69,2%	1,440	47,1%
V1 totals	445		6,124		2,440	
SF totals	**4,174**	**90,3%**	**21,312**	**77,7%**	**7,616**	**75,7%**

Interestingly, the selection of finite auxiliary, *er* 'is' vs. *hefur* 'has', markedly affects the SF frequency: SF of the participles in Table 6 is more frequent with *er* than with *hefur*. The same effect of auxiliary selection is clearly seen for e.g. the disyllabic participles *byrjað* 'begun', *búið* 'done, finished; lived', *talið* 'considered, reckoned, counted', and the monosyllabic *gert* 'done' and *sagt* 'said'. That is: *sem byrjað/búið/talið/gert/sagt **er*** are all more frequent (in relation to V1, pairwise) than are *sem byrjað/búið/talið/gert/sagt **hefur***.[19] I have no obvious account of this curious fact. It might relate to prosody (the monosyllabic vs. the disyllabic structure of *er* vs. *hefur*, cf. Wood 2011), but the results are too opaque and diffuse to allow any conclusion or claim to that effect, as far as I can judge.

The examples we have looked at so far are simple, with the relative complementizer *sem* 'that, who, which', a finite auxiliary and a main verb past participle. In examples of this sort, the participle is the only potential SF "candidate". If the clause also contains an object DP, an adverbial, particle or an adjectival predicate, more contenders come into play. Some cases of this sort, with an adverbial complement of the participle, are exemplified in (16) and (17).

(16) a.   sem __   hafa   búið   þar …
          that   have   lived   there
      b.   sem **búið** hafa   *t*   þar ...
      c.   sem **þar**   hafa   búið   *t* ...

---

[19] The SF ratios for the former in *Timarit.is* (in July 2015) were between 87% and 97%, for the latter between 58% and 81%.

(17) a. sem __   hafa búið í Danmörku …
        that   have lived in Denmark
    b. sem **búið**   hafa *t*   í Danmörku …
    c. sem **í Danmörku** hafa búið *t* ...

My search results for these examples are presented in Table 7.

Table 7. Search results for the examples in (16) and (17). The Google search was conducted on September 25, 2014 and it searched for results within the date range from January 1, 2004 to January 1, 2014. The *Timarit.is* search was unlimited, conducted on September 3, 2014.

	Google		Timarit	
	#	%	#	%
V1: sem __ hafa búið þar	10	29%	23	10%
SF: sem **búið** hafa þar	4	12%	22	9%
SF: sem **þar** hafa búið	20	59%	196	81%
V1: sem __ hafa búið í Danmörku	1		8	42%
SF: sem **búið** hafa í Danmörku	2		11	58%
SF: sem **í Danmörku** hafa búið	1		0	

Despite the low numbers for the *búið* fronting in (16b), there is nothing "wrong" with *búið* as an SF candidate, as such. This is illustrated by the results for *búið* fronting in (17b) and also by the results in Table 8 for the simple strings *sem __ hafa búið* 'who/that have lived' and *sem búið hafa*; these results include the types in (16a–b) and (17a–b), in addition to other types (e.g., with *búið* as a particle verb).

Table 8. Results for Google and *Timarit.is* searches for *sem hafa búið* vs. *sem búið hafa* on July 4 2015. The Google search was limited to the period July 1 2005 to July 1 2015, whereas the *Timarit.is* search was unlimited.

	Google		Timarit	
	#	%SF	#	%SF
V1: sem __ hafa búið	420		1,459	
SF: sem **búið** hafa	243	36,7%	1,690	54,2%

The effect of the presence of *þar* 'there' in (16) is striking and so is the fact that the prepositional phrase *í Danmörku* 'in Denmark' has no such effect.[20] That is:

---

[20] The same applies to other locative PPs that are complements of the participle *búið*. I checked this in September 2014 for the strings *sem í X hafa búið*, where X = *New York, London, París, Stokkhólmi, Berlín, Moskvu, Róm, Kaupmannahöfn, Madríd, Lissabon, Aþenu, Peking/Beijing, Tókýó, Japan, Þýskalandi, Frakklandi, Grikklandi*. These searches gave zero hits in both corpora.

í Danmörku is clearly not a "serious SF contender" in (17) whereas *þar* is in (16), only the latter outcompeting the participle *búið* as an SF candidate. Both *þar* and *í Danmörku* are complements of *búið*, and should thus, contrary to fact, be equally amenable to SF under Holmberg's (2000, 2006) understanding of equidistance and structural closeness. Either Holmberg's definition of structural closeness must be revised or the properties of the potentially moved category (and its "neighbors") interfere with locality, thus affecting the applicability of Stylistic Fronting (see also the discussion in Ott 2009:149ff., Wood 2011). I assume that the latter is the case.

Fronting of full DP objects is generally rare in subject relatives regardless of the presence or absence of a participle. Thus (on July, 6 2015), *sem **bækurnar** lásu* 'who the books read' and *sem **bækurnar** hafa lesið* each gave a single hit in *Timarit.is*. The V1 "competitors", *sem lásu bækurnar* and *sem hafa lesið bækurnar*, yielded 6 and 18 hits respectively. On the other hand, *sem **þær** lásu* and *sem **þær** hafa lesið*, with the feminine plural pronoun *þær* 'them' (as an object), yielded 4 and 12 hits, respectively, whereas their V1 competitors *sem lásu þær* and *sem hafa lesið þær* gave 11 and 20 hits respectively. Searching for other examples of this sort yielded similar results.

Personal pronouns and adverbs like *þar* (as in (16)) and *hér* 'here' are indexical or deictic elements, with their reference depending on properties of the speech event (see Sigurðsson 2014 and the references there). That is: the interpretation of such elements depends on who is talking to whom, where and when. DPs and PPs/AdvPs that contain deictic elements seem to front more readily than do other DPs and PPs/AdvPs. Thus, searching *Timarit.is* (July 6, 2015) for *sem **við mig** hafa talað* 'who with me have spoken' gave 47 hits, whereas its "competitors", *sem hafa talað við mig* and *sem **talað** hafa við mig*, yielded 56 and 24 hits respectively.[21] Comparable results for *sem **á hann** hafa hlustað* 'who to him have listened' and its competitors *sem hafa hlustað á hann* and *sem **hlustað** hafa á hann* gave 11, 8 and 8 hits, respectively. For clarity, these results are stated in Table 9.

Evidently, the frequency or applicability of SF in subject relatives is affected by a number of factors other than just the "X-bar form" of the potential "mover" and its closeness to the subject gap. The presence of other SF contenders is obviously an important factor and indexicality seems to play a role too. Other factors are more moot and difficult to isolate and estimate. Thus, it has been observed that SF is sometimes accompanied by focus or accentuation (Hrafnbjargarson 2004, Molnár 2010), but focus/accentuation is not a triggering or favoring factor, at least not a general one.[22]

---

[21] Many thanks to a very sharp reviewer for pointing these examples out to me.

[22] Accentuation may for instance apply in rare cases of clear contrasts, as in *sem GERT hafa eitthvað en ekki bara TALAÐ* lit. 'who DONE have something and not just TALKED' (Sigurðsson 1997), but comparable examples without a contrast or accentuation are fine too (*sem gert hafa ýmislegt fyrir byggðarlagið*, 'who done have various things for the district', etc.).

In my judgment SF is in fact typical of generic clauses with a flat intonation and information contour (cf. Egerland 2013; but see shortly on *víst* and *vissulega* in (18)).

Table 9. A few results in *Timarit.is*, July 6, 2015.

	#	SF
V1: sem __ lásu bækurnar	6	
SF: sem **bækurnar** lásu	1	14%
V1: sem __ hafa lesið bækurnar	18	
SF: sem **bækurnar** hafa lesið	1	5%
V1: sem __ lásu þær	11	
SF: sem **þær** lásu	4	27%
V1: sem __ hafa lesið þær	20	
SF: sem **þær** hafa lesið	12	38%

	#	%
V1: sem __ hafa talað við mig	56	44%
SF: sem **talað** hafa við mig	24	19%
SF: sem **við mig** hafa talað	47	37%
V1: sem __ hafa hlustað á hann	8	30%
SF: sem **hlustað** hafa á hann	8	30%
SF: sem **á hann** hafa hlustað	11	40%

Actually, "lightness" rather than focus/accentuation seems to favor SF. Wood presents evidence from spoken language corpora that "constituents with 1 syllable highly favor fronting, those with 2 syllables weakly disfavor fronting, and those with 3–5 strongly disfavor fronting" (2011:45). Deictic elements are also "light" in another sense: they are presupposed in a given speech event and thus "informationally light". As many indexicals are monosyllabic and often deaccentuated, informational lightness and phonetic lightness commonly overlap, and it is not always easy to tell these factors apart. However, when they can be teased apart, there is some evidence that mere phonetic lightness is not a strongly promoting or favoring factor. Consider the examples in (18) and the search results for these in Table 10.

(18) a.　sem hefur víst / sem víst hefur
　　　　that has sure / that sure has
　　　　'that/who allegedly has; that/who for sure has'
　　b.　sem hefur vissulega / sem vissulega hefur
　　　　that has certainly / that certainly has
　　　　'that/who certainly has; that/who I grant you has'

These figures are striking, showing a very strong *negative* correlation between the frequency of SF and the phonetic lightness of the potential "mover". However, it seems likely to me that the behavior of *víst* and *vissulega* is somewhat special. Both

have multiple meanings, their interpretation relating to evidentiality and other modality and discourse factors that are not easy to pin down. I have the intuition (at least for subject relatives) that fronting of these elements is commonly accompanied by accentuation, otherwise atypical of SF (in Icelandic as opposed to Sardinian, see Egerland 2013), and that their reading is often affected by fronting and/or accentuation.

Table 10. Results for Google and *Timarit.is* searches for the examples in (18) in July 2015. The Google search was limited to the period July 1 2005 to July 1 2015, whereas the *Timarit.is* search was unlimited.

	Google			*Timarit.is*		
	#V1	#SF	%SF	#V1	#SF	%SF
18a: víst	34	0	0%	83	24	22,4%
18b: vissulega	24	52	68,4%	65	365	84.9%

I also searched for examples with the roughly synonymous but variably light adverbials *því* 'thus, therefore', *þess vegna* 'therefore' (lit. 'that because (of)'), and *þar af leiðandi* 'therefore' (lit. 'there of leading'). The examples are given in (19) and the search results are shown in Table 11.

(19) a. sem hefur því / sem **því** hefur
           that has thus / that thus has
           'that/who has thus/therefore'
      b. sem hefur þess vegna / sem **þess vegna** hefur
           that has that-because / that that-because has
           'that/who has thus/therefore'
      c. sem hefur þar af leiðandi / sem **þar af leiðandi** hefur
           that has there-of-leading / that there-of-leading has
           'that/who has thus/therefore'

Table 11. Results for Google and *Timarit.is* searches for the examples in (19) in July 2015. The Google search was limited to the period July 1 2005 to July 1 2015, whereas the *Timarit.is* search was unlimited.

	Google			*Timarit.is*		
	#V1	#SF	%SF	#V1	#SF	%SF
19a: því	1,280	620	32,6%	273	940	77,5%
19b: þess vegna	2	0	0%	10	8	44,4%
19c: þar af leiðandi	5	4	44,4%	13	12	48,0%

Again, there is a negative correlation between SF and the phonetic lightness of the potential "mover" in the Google data, whereas the opposite holds of the *Timarit.is* data.

Thus, while the figures in Tables 7 and 9 indicate that there might by a strong positive correlation between (at least informational) lightness of the potential "mover" and the frequency of SF, the figures in Tables 10 and 11 indicate the opposite, with the exception of the *Timarit.is* figures in Table 11. Notice also that SF of the trisyllabic *skrifaður* 'written' in (15i–j) above is about as frequent as SF of the bisyllabic *skrifað* and *skrifuð* in (15a–b) and (15e–f).[23] Probably, lightness is a more prominent factor in spoken than in written language, but as the bulk of the corpora studied by Wood contain (often written) speeches in Alþingi, the Icelandic parliament, it is unclear whether they are much closer to everyday spoken Icelandic than the texts I have searched on Google. In any event, we can conclude that the frequency of SF is affected by a complex interplay of a number of factors. Thus, if we replace *hefur* in (19) by *er*, the results show a much stronger correlation with phonetic lightness, thus being more in line with Wood's findings, but if we do the same in (18), we still get a negative correlation with lightness (*vissulega* fronting more readily than *víst*). I leave this discussion of the effects of lightness on the frequency of SF in subject relatives in this inconclusive and rather unsatisfactory state. More research on this issue, with more powerful tools, is clearly needed.

The statistics presented in this section confirm that SF in subject relatives is robust in everyday written Icelandic. Nevertheless they show, first, that SF is markedly less frequent on the World Wide Web than in *Timarit.is*, and, second, that the frequency of SF in *Timarit.is* declines over time (see Tables 3–6 above). Other things being equal, these results would thus seem to corroborate the results of recent informant surveys, reported in work by Angantýsson (2009, 2011, 2017) and Thráinsson et al. (2015), showing that young informants are somewhat more likely than older ones to reject or question SF. If so, my results would indicate a change in real time, whereas the informant surveys indicate a change in apparent time. However, it is not clear that the methods of these different types of studies of different data are comparable or bear on the "same reality" in some sense. In addition, the trend seen in my data for SF frequency in subject relatives to decline over time might not be the result of an ongoing historical change but a side effect of increased written language informality, not only on the Internet but also in the texts in *Timarit.is*. Nevertheless, it seems that SF in subject relatives is gradually losing ground against V1 in everyday written Icelandic, even though this domain loss is happening slowly.

---

[23] The ratios SF/V1+SF (referred to as %SF in my tables) for *skrifað* were 66,4% (Google) and 82,4% (Timarit), and 63,1% (Google) and 70,0% (Timarit) for *skrifuð*. For *skrifaður* they were 72,1% (Google) and 69,1% (Timarit).

5. Clauses with a non-trace subject gap (impersonal clauses)

In this section, I study clauses with a non-trace subject gap (impersonal clauses), where SF competes with both V1 and insertion of the expletive *það* 'it, there'. The most central result of this study is that SF has a strong foothold in impersonal clauses in written Icelandic, even though there are clear indications in the data that expletive insertion is gaining ground there.

For practical reasons the survey was limited to clauses with participles as potential SF-candidates (mostly in the impersonal passive). Data were collected for the clause types listed in (20):[24]

(20) a. Declarative *að* 'that' clauses (in the subjunctive)
      b. Interrogative *hvort* 'whether, if' clauses (in the indicative)
      c. Conditional *ef* 'if' clauses (in the indicative)
      d. Comparative *eins og* 'as (if)' clauses (in the indicative)
      e. Temporal *þegar* 'when' and *áður en* 'before' clauses (in the indicative)

The examples are shown in (21)–(26).

(21) *Declarative* að *clauses* (in the subjunctive):

   a. að   __   hefði   átt
       that   had   ought
       'that one/people should have'
   b. að **átt** hefði *t*
   c. að **það** hefði átt

(22) *Interrogatives*:

   a. hvort   __   verður   farið
       whether   will-be   gone/begun
   b. hvort   **farið**   verður   *t*
   c. hvort   **það**   verður   farið

(23) *Conditionals*:

   a. ef __ er farið
       if   is   gone/begun
   b. ef **farið** er *t*
   c. ef **það** er farið

---

[24] It is difficult to search mechanically for indicative declarative *að* 'that' clauses as there are many more indicative *að* clause types than just declaratives. The subjunctive strings I opted for searching, in (21a–c), are unlikely to be anything but declarative. For the other clause types I searched separately for both indicatives and subjunctives (the latter being much fewer in all cases). As I could not discern any significant relations of the moods with word order type differences I only account for my results for the indicatives for these other clause types. On the other hand, as we will see in section 6, the subjunctive seems to be a strongly favoring factor for SF in *að* clauses.

(24) *Comparatives*:

    a.   eins og __ var   gert

           as   was done/made

    b.   eins og **gert** var *t*

    c.   eins og **það** var gert

(25) *Temporals A*:

    a.   þegar __ er   gengið

           when   is   walked

    b.   þegar **gengið** er *t*

    c.   þegar **það** er gengið

(26) *Temporals B*:

    a.   áður en __ er   komið

           before   is   arrived/come

    c.   áður en **komið** er *t*

    b.   áður en **það** er komið

The results are presented in Table 12. Seaching for *það* 'there, it, that' in this context will necessarily turn up many referential *það*'s and such examples are obviously irrelevant for our purposes. In an effort to remedy this the first 20 (or up to 20) *það*-examples were manually checked in each case. If at least 50% of these first instances of *það* turned out to be referential, the figure in Table 12 is marked with a strikethrough.[25]

Expletive *það* 'it, there' was largely absent in early Icelandic but it has been gradually gaining ground since at least around 1500 (Rögnvaldsson 2002:21ff.). Like many other historical changes in Ielandic this change has been proceeding very slowely. Informant surveys would seem to indicate that the use of the expletive is still spreading – informants over the age of 40 accepting it somewhat more reluctantly than younger speakers (see Thráinsson et al. 2015:285). Again, however, it is unclear whether this (not very strong) correlation with age is due to an ongoing historical change or to variation in style and formality. The expletive is commonly considered too informal for written style and fought against by teachers and language planners (see Rögnvaldsson 2002:27 and the references there) and this might affect informant judgments.

Regardless of informant judgments and the different status of the expletive in written and spoken Icelandic my results indicate that *það* is gaining ground at the expence of SF in at least some impersonal sentence types in everyday written Icelandic. Thus, many of the relatively numerous *ef það er farið* (lit. 'if it/there is gone/begun') examples in (23c) do contain an expletive where only V1 or SF would have

---

[25] Again, the frequency of SF, V1 and *það*-V is only representative of the types of strings searched for (mostly only a complementizer plus a finite verb, a participle and potentially *það* in impersonal contexts).

Table 12. Search results for the examples in (21)–(26). The Google search was conducted on September 25, 2014 and it searched for results within the date range from January 1, 2004 to January 1, 2014. The *Timarit.is* search was unlimited, conducted on September 3, 2014. "ÞA" = *það* and the strikethrough indicates that at least 50% of the first 20 instances of *það* were referential. The corresponding ratios are given within parentheses.

	Google		Timarit	
	#	%	#	%
21a. V1: að __ hefði átt	16	7,4%	326	23,5%
21b. SF: að **átt** hefði	10	4,6%	231	16,7%
21c. ÞA: að **það** hefði átt	190	88,0%	831	59,9%
22a. V1: hvort __ verður farið	1	2,1%	10	2,8%
22b. SF: hvort **farið** verður	44	93,6%	349	95,4%
22c. ÞA: hvort **það** verður farið	2	4,3%	7	1,9%
23a. V1: ef __ er farið	4	0,2%	2	0,05%
23b. SF: ef **farið** er	1,610	66,3%	4,002	98,8%
23c. ÞA: ef **það** er farið	791	32,9%	47	1,2%
24a. V1: eins og __ var gert	153	13,3%	166	2,3%
24b. SF: eins og **gert** var	993	86,0%	7,047	97,3%
24c. ÞA: eins og **það** var gert	8	0,7%	~~28~~	(0,4%)
25a. V1: þegar __ er gengið	3	0,2%	29	0,9%
25b. SF: þegar **gengið** er	1,470	99,7%	3,041	98,7%
25c. ÞA: þegar **það** er gengið	2	0,1%	~~12~~	(0,4%)
26a. V1: áður en __ er komið	3	0,2%	4	0,3%
26b. SF: áður en **komið** er	1,010	75,5%	1,396	95,1%
26c. ÞA: áður en **það** er komið	~~307~~	(23,3%)	~~68~~	(4,6%)
21a–26a. V1:	180	2,7%	537	3,1%
21b–26b. SF:	5,137	77,6%	16,066	91,3%
21c–26c. ÞA:	1,302	(19,9%)	993	(5,6%)

been possible at earlier historical stages of the language. While the sharp contrast between my Google and *Timarit.is* results (32,9% vs. 1,2%) might be partly due to style and genre differences it seems likely to me that it largely reflects an ongoing expansion of the domain of *það* in the written language. Thus, 51% (24) of the 47 *Timarit.is ef það er farið* examples in (23c), are found in texts published in the year 2000 or later (the comparable figure in Table 2 for the string *sem hafa verið* is 29,2%).

The overwhelmingly most common type of *það* in the declartives in (21c) is *það* that anticipates a postposed infinitival or clausal subject.[26] Anticipating *það*

---

[26] 82,1% of the *að það hefði átt* examples in both corpora (exactly the same ratio) contained *að* 'that, to' directly after *átt*. In the remaining examples *það* is almost exclusively referential (*átt* there being a main verb meaning 'own' and not a modal meaning 'should, ought').

is found already in Old Icelandic (Rögnvaldsson 2002), so the results for (21) in table 11 (88% and 59,9% with *það*) do not necessarily suggest that more modern expletive types are gaining ground, but they indicate that at least anticipating *það* is spreading in impersonal declaratives, at the expense of SF (but less clearly so at the expense of V1).[27] That this is probably the case gains some credibility from the fact that the frequency of *að það hefði átt* in *Timarit.is* markedly increases over time, as seen by the results in Table 13.

Table 13. Results (July 14, 2015) for different periods in *Timarit.is* for the strings in (21).

	–1949		1950–1999		2000–2015	
	#	%	#	%	#	%
V1: að __ hefði átt	83	27,6%	197	22,7%	54	21,7%
SF: að **átt** hefði	62	20,6%	146	16,8%	26	10,4%
ÞA: að **það** hefði átt	156	51,8%	524	60,4%	169	67,8%

We will see further evidence in the next section suggesting even more decisively that *það* is gaining ground at the expense of SF, in particular in *að* clauses but also to some extent in other clause types.

The three points in (27) summarize the most central results and conclusions of this section on impersonal clauses.

(27) a. V1 is the least common of the three word orders and it is unevenly spread across clause types, but it is far from being non-existent and it does not seem to be generally losing ground in the written language. We will return to subordinate V1.

  b. Expletive *það* is on the increase in the written language, but, with the exception of declarative *að* clauses, this is a slow process and it is also unevenly spread across clause types.

  c. SF is still the most common of the three competing word order types in impersonal clauses in the written language, much more common than V1 and *það*-V together in all the clause types checked, with declarative *að* clauses as an exception.

These conclusions will be further tested in the next subsection, where I also check whether there is a tendency for SF of participles to get frozen in idiomatic expressions – which, if true, might indicate that it is becoming marginal in the language. As we will see, this does not (generally) seem to be the case.

---

[27] For an extensive discussion of different types of *það* in Icelandic, see Thráinsson 1979:176ff. See also Thráinsson 2007:309ff.

6. Idiomatization?

As Angantýsson (2009, 2011:158ff.) points out there are certain impersonal constructions where SF has been idiomaticized in the sense that it is the only or at least the most salient option by far, both V1 and *það*-insertion being either awkward or outright unacceptable. (28) is a case in point (my judgements).

(28) a. Ef **grannt** er skoðað *t* er ljóst að …
   if closely is looked-at is clear that
   'On scrutiny, it is clear that …'
  b. ?*Ef __ er skoðað grannt er ljóst að …
  c. * Ef það er skoðað grannt er ljóst að …[28]

However, none of Angantýsson's examples of idiomatization contain a fronted past participle (instead containing fronted particles, adverbs, adjectives, etc.), and I have not discerned any idiomatization tendency for participles. To throw some light on this issue I checked the frequency of V1, SF and *það*-initial orders in impersonal adverbial clauses with present tense *er* 'is' in combination with 10 participles and 3 connectives, as stated in (29).

(29) a. The present tense *er* 'is' (3 person singular).
  b. The connectives *áður en* 'before', *ef* 'if' (conditional), *eins og* 'as (if)'.
  c. The participles *byrjað* 'begun', *farið* 'gone, begun', *gengið* 'walked', *gert* 'done, made', *lesið* 'read', *sagt* 'said', *spurt* 'asked', *talið* 'believed, counted', *talað um* 'talked about', *verið* 'been'.

The strings checked were thus 90 in number (3 connectives x 10 participles x 3 word orders). In a sense, the results of these checks were negative. That is, the data showed no clear correlations between individual participles and the frequency of SF, thus no indications of idiomatization of SF. Also, none of the fronted participles gets an idiomatic reading in any of the SF strings, and both V1 and *það*-insertion are acceptable in all the examples (at least in my grammar). However, some correlations with V1 and *það*-V (hence indirectly with SF frequencies) can be discerned, as I will discuss in the following.

 First, it should be noted that *það* is very commonly referential in combination with *er* + *gert/lesið/sagt/spurt/talið*, the searched strings then usually meaning 'it/that is done/read/told/asked/counted' (rather than impersonal 'there is something unspecified being done/read/told/asked/counted by somebody', as it were). The

---

[28] Rare but possible if *það* is referential.

overall results for the strings with *er + gert/lesið/sagt/spurt/talið* are summarized in Table 14. As before the strikethroughs indicate that at least 50% of the (up to) first 20 instances of *það* were referential, hence irrelevant (but in some of the cases expletives nevertheless constitute a substantial portion of the *það*s).

Table 14. Results for V1, SF and *það*-V strings (ÞA) in examples with *gert, lesið, sagt, spurt, talið* on Google and in *Timarit.is*. The Google search was conducted on September 25, 2014 and it searched for results within the date range from January 1, 2004 to January 1, 2014. The *Timarit.is* search was unlimited, conducted on September 3, 2014. The strikethrough indicates as before that at least 50% of the (up to) first 20 instances of *það* were referential.

	Google		Timarit	
	#	%	#	%
V1	395	4,5%	128	0,5%
SF	4,776	54,2%	23,391	91,3%
ÞA	~~3,647~~	(41,4%)	~~2,108~~	(8,2%)
Totals	8,818		25,627	

As seen, the frequency of (referential and expletive) *það* was about five times higher in the Goolge search than in *Timarit.is*. V1 is also markedly more frequent in the Google results than in *Timarit.is*. No clear correlation was found for any of the word order types with individual participles, whereas there is a strong correlation between V1 and the connective *eins og* 'as (if)'. Of the 395 V1 Google hits, 393 were found in *eins og* clauses (8,6% of the 4,548 *eins og* Google clauses), two in *ef* 'if' clauses, none in *áður en* 'before' clauses. Of the 128 V1 *Timarit.is* hits, 124 were found in *eins og* clauses, two in *ef* clauses, two in *áður en* clauses.

As stated in (29c), the other five participles checked were *byrjað, farið, gengið, talað um, verið*. More than 50% of the (up to) first 20 instances of *það* in examples with these were expletive. The results are summarized in Table 15.[29]

---

[29] Most of the Google searches were conducted on September 25, 2014 searching for results within the date range from January 1, 2004 to January 1, 2014, and most of the *Timarit.is* searches were conducted on September 3, 2014 and searched the whole corpus (till then). However, strings with the progressive participle *verið* 'been' were not included in these 2014 searches, so they were specifically searched for in July 2015 (for July 1, 2005 to July 1, 2015 in the Google search and in the whole *Timarit.is* corpus). The effects of these temporal differences are marginal.

Table 15. Results for V1, SF and *það*-V strings (ÞA) in examples with *byrjað*, *farið*, *gengið*, *talað um*, *verið* on Google and in *Timarit.is*. The Google search was conducted on September 25, 2014 and it searched for results within the date range from January 1, 2004 to January 1, 2014. The *Timarit.is* search was unlimited, conducted on September 3, 2014.

	Google		Timarit	
	#	%	#	%
V1	468	3,7%	43	0,3%
SF	8,285	66,4%	15,557	98,5%
ÞA	3,719	29,8%	193	1,2%
Totals	12,472		15,793	

As seen, there is little variation in the *Timarit.is* data, SF being ca 66 times more common than V1 and *það*-V together. The Google results are more varied and also more interesting. They are broken down for the different connectives in Table 16.

Table 16. The Google results in Table 15 broken down for the three different connectives.

	Google	
	#	%
V1: áður en ___ er X	29	0,6%
SF: áður en **X** er	4,539	94,1%
ÞA: áður en **það** er X	255	5,3%
V1: ef ___ er X	14	0,2%
SF: ef **X** er	3,256	48,4%
ÞA: ef **það** er X	3,456	51,4%
V1: eins og ___ er X	425	46,1%
SF: eins og **X** er	490	53,1%
ÞA: eins og **það** er X	7	0,8%
V1 totals	468	3,7%
SF totals	8,285	66,4%
ÞA totals	3,719	29,8%

We see clear correlations with the connectives here. First, V1 is very common in the *eins og* clauses. Second, *það* is roughly 10 times more common in *ef* clauses than in *áður en* clauses and 64 times more common than in *eins og* clauses. Presumably, these facts are to some extent interrelated, but, in view of the uncertainty of how the Google algorithms work, these deviant data must be cautiously interpreted. They are largely due to clauses with the participle *verið* 'been'. The Google results for the *ef* and *eins og* clauses are further broken down in Table 17.

Table 17. The Google results for the *ef* and *eins og* clauses in Table 15 further broken down (singling out clauses with *verið*).

	X = byrjað, farið, gengið, talað um		X = verið	
	#	%	#	%
V1: ef ___ er X	5	0,2%	9	0,2%
SF: ef **X** er	2,313	74,0%	951	26,3%
ÞA: ef **það** er X	806	25,8%	2,650	73,4%
V1: eins og ___ er X	18	6,8%	407	61,8%
SF: eins og **X** er	240	91,3%	250	37,9%
ÞA: eins og **það** er X	5	1,9%	2	0,3%

As seen, expletive *það* is exceptionally frequent in Google *ef* conditionals with *verið* 'been'. However, the conditional examples with *verið* almost exclusively contain progressive *vera* 'be (doing)'. The examples in (30) are typical.[30]

(30) a.   ef   **það er verið** að nota símann

         if   there is been   to   use   phone-the

         'if the phone is being used'

         https://barn.is/boern-og-unglingar/spurt-og-svarad-safn/2015/04/ma-kennari-taka-og-geyma-sima/ – July 17, 2015

     b.   ef   **það er verið** að gróðursetja í sólskini

         if   there is been   to plant   in sunshine

         'if there is planting of something in the sunshine'

         http://www.plantan.is/index.php/fraedhsla/avaxtatre – July 17, 2015

The frequency of V1 *eins og __ er verið* 'as is been' is also extraordinary. The example in (31) is typical; interestingly, and curiously, the introducing temporal clause contains an example of *það er verið* 'it is been', underlining the coexistence of V1 and *það*-V.

(31) [Á meðan **það   er verið** að skera niður]

      in-while   there is been to cut   down

      eins og **er verið** að gera núna

      as   is   been to do   now

      'While the budged is being cut, as is being executed for the time being'

      https://www.betrireykjavik.is/ideas/183-sundlaug-i-fossvogsdal August 2, 2015

The different behavior of passive and progressive *verið* in potential SF contexts (previously discussed by Jónsson 1991 and others) shows, once again, that many

---

[30] See Sigurðsson 1989, chapter 3.2.2, for a discussion of aspectual verbs in Icelandic. On the progressive in particular, see Jóhannsdóttir 2011.

factors affect the applicability of SF other than just the form of the potential "mover" and its distance from the subject gap.

With the curious exception of *ef* 'if' clauses with the participle *verið*, SF is the prevailing option in impersonal adverbial clauses, even in other clause types with *verið* (I checked this in a Google search in July 2015 for *verið* clauses introduced by a number of connectives). Nevertheless, the results above strongly indicate that *það* is gaining ground. This tendency is seen even more clearly in clauses introduced by *að* 'that'. I checked this (in July 2015) for the five participles in Table 14 (*byrjað*, *farið*, *gengið*, *talað um*, *verið*), with both third person singular indicative *er* 'is' and subjunctive *sé* 'is, be' (without trying to distinguish between the many functions of clauses introduced by *að*). The Google data showed that indicative *að það er farið/talað um/verið* are more or much more frequent than their V1 and SF competitors (while the data for the *byrjað* and *gengið* clauses were less clear). Interestingly, the opposite holds for the subjunctive clauses. The results for the *verið* clauses are presented in Table 18.

Table 18. Google results (in July 2015) for indicative and subjunctive *að* clauses with *verið* 'been' (for July 1 2005 to July 1 2015).

	Indicative (*er*)		Subjunctive (*sé*)	
	#	%	#	%
V1: að __ er verið / að __ sé verið	276	2,8%	174	2,3%
SF: að **verið** er / að **verið** sé	1,740	18,6%	5,170	68,4%
ÞA: að **það** er verið / að **það** sé verið	7,620	79,1%	2,220	29,3%

The corresponding results for *að* clauses in *Timarit.is* were rather different, showing much higher frequencies for SF than for *það*-insertion for all five participles (*byrjað*, *farið*, *gengið*, *talað um*, *verið*), in both indicative and (especially) subjunctive clauses (nevertheless showing slowly rising frequencies for *það* over time). For subjunctive *að* clauses with *verið* in the *Timarit.is* corpus the SF ratio (SF/V1+SF+ÞA) was 87,7%.

It seems to me, not surprisingly, that the Google results show a much closer affinity with common spoken Modern Icelandic (as I know it) than do the *Timarit.is* results. However, neither corpora show any clear signs of idiomatization of SF of the past participles searched for.

7. And when "nothing" happens?

Some researchers (e.g., Kosmejer 1993, Holmberg & Platzack 1995, Holmberg 2000) have assumed that V1 is ungrammatical in Icelandic subordinate clauses with the exception of subject relatives and other clauses with a subject trace gap. However, in the absence of a participle or some other "relatively good" SF candidate, V1 is easily found in impersonal clauses with a non-trace subject gap. A few such examples were searched for (in September 2014), with the connectives *áður en* 'before' and *þegar* 'when' and the predicates *(það) fer/fór að rigna* '(it) begins/began to rain'. The results are presented in Table 19.

Table 19. Results (in September 2014) for V1 vs. *það*-V in (present and past) *áður en* and *þegar* clauses without a "good SF candidate".

	Google		Timarit	
	#	%V1	#	%V1
V1: áður en __ fer/fór að rigna	9	56,2%	36	85,7%
ÞA: áður en **það** fer/fór að rigna	7		6	
V1: þegar __ fer/fór að rigna	17	68,0%	132	87,4%
ÞA: þegar **það** fer/fór að rigna	8		19	

The figures are low and the relatively low frequency of *það* in *Timarit.is* is probably due to it commonly being "weeded" out in written style, and this "weeding" obviously also affects the Google statistics, albeit to a lesser extent. Nevertheless it is remarkable that V1 is more common than *það*-V in all four cases (and also in all eight cases, if one splits up the results for past and present tense).

I complemented this little study in August 2015 by searching for V1 and *það*-V orders on Google (for July 1,2015–July 1, 2015) in the context of *þegar* 'when' in combination with the third person singular present indicative forms *birtir* 'gets brighter', *byrjar* 'begins', *dimmir* 'darkens', *hlýnar* 'gets warmer', and *hættir* 'stops', getting altogether **1,199** V1 hits and **294** *það*-V hits, respectively, V1 thus being ca 4 times more common in these contexts than *það*-V. An informant survey reported in Angantýsson (2011:155; see also Þráinsson et al. 2015:280) shows that young speakers accept the expletive more readily in *þegar það fer að snjóa* 'when it begins snowing' than do older informants (85% vs 68%), but it also shows that V1 (*þegar __ fer að rigna*) is widely accepted by both age groups (65% vs 91%). There is no question that V1 is "alive and relatively well" in some impersonal adverbial clauses.

8. Conclusion

This paper studies the distribution and frequency of Stylistic Fronting (SF) and the competing V1 and *það*-V orders on the World Wide Web and in *Timarit.is* across two distinct domains: (i), subject relatives, and, (ii), subjectless impersonal clauses. The survey shows that SF is robust in potential SF contexts in everyday written Icelandic, even though the data strongly suggest that it is presently losing ground against V1 in subject relatives and against *það*-V in impersonal clauses. Simultaneously, the availability of V1 in certain subordinate impersonal constructions shows that Icelandic (like so many other languages) does not obey a strict syntactic Extended Projection Principle. Nevertheless, the frequency of SF (plus *það*-insertion) in impersonal constructions suggests that filling the left edge of CP is a "target" in Icelandic grammar, but it seems to be an externalization or performance target – a commonly desirable PF goal, as it were.[31] SF is sensitive to syntactic conditions (being clause bounded, confined to finite clauses, etc.), but it would seem that it nevertheless involves some kind of an adjustment in PF, the externalization component. That tallies with the standard generative assumption that PF is an interpretative interface, "interpreting" syntax (phonologically), among other things by regulating word order. It has been repeatedly argued (for example in the work of Sigurðsson, see, e.g., 2010, 2014 and the references there) that much of what is traditionally referred to as "syntax" is actually part of PF – and that claim would seem to gain support from the results of the present study.[32]

An encouraging extra result of the study, a methodological byproduct, as it were, is the conclusion that Google Search, if carefully used, is a much more valuable research tool in linguistics than commonly assumed. Repeated checks in the years 2010-2015 have shown that Google searches within a given period, as opposed to unlimited searches, yield reasonably stable results. Also, comparison of the Google results with the *Timarit.is* results reveals fairly consistent statistical correlations between the corpora.

References

Angantýsson, Ásgrímur. 2009. Stylistic Fronting and expletive insertion: Some empirical observations. Paper presented at the Joan Maling Seminar, Reykjavík.

Angantýsson, Ásgrímur. 2011. *The Syntax of Embedded Clauses in Icelandic and Related Languages*. Reykjavík: Hugvísindastofnun.

---

[31] When leaving Spec,CP empty does not serve some specific "purpose", as, e.g., in topic drop and narrative inversion (see Sigurðsson 2010).

[32] This is partly similar to and partly rather different from Holmberg's approach (2000, 2006), where SF is taken to be a syntactic process that nevertheless moves only the phonetic matrix of the fronted category.

Angantýsson, Ásgrímur. 2017. Stylistic Fronting and related constructions in the Insular Scandinavian languages. In this volume, 277–306.

Egerland, Verner. 2013. Fronting, bagckground, focus: A comparative study of Sardininan and Icelandic. *Lingua* 136:63–76.

Falk, Cecilia. 1993. Non-referential subjects in the history of Swedish. Doctoral dissertation, Lund University.

Franco, Irene. 2009. Verbs, subjects and Stylistic Fronting. Doctoral dissertation, University of Siena.

Gatto, Maristella. 2014. *Web as Corpus: Theory and Practice*. New York: Bloomsbury.

Holmberg, Anders. 2000. Scandinavian Stylistic Fronting: How any category can become an expletive. *Linguistic Inquiry* 31:445–483.

Holmberg, Anders. 2006. Stylistic fronting. In *The Blackwell Companion to Syntax*, edited by Martin Everaert & Henk van Riemsdijk, 532–565. Oxford: Blackwell.

Holmberg, Anders & Christer Platzack. 1995. *The Role of Inflection in Scandinavian Syntax*. Oxford: Oxford University Press.

Hrafnbjargarson, Gunnar Hrafn. 2004. Stylistic Fronting. *Studia Linguistica* 58:88–134.

Jóhannsdóttir, Kristín M. 2011. Aspects of the progressive in English and Icelandic. Doctoral dissertation, The University of British Colombia.

Jónsson, Jóhannes Gísli. 1991. Stylistic Fronting in Icelandic. *Working Papers in Scandinavian Syntax* 48:1–43.

Kilgarriff, Adam. 2007. Googleology is bad science. *Computational Linguistics* 33:147–151.

Kosmeijer, Wim. 1993. Barriers and licensing. Doctoral disseratation, University of Groningen.

Magnússon, Friðrik. 1990. *Kjarnafærsla og* það-*innskot í aukasetningum í íslensku* [Topicalization and *það*-insertion in subordinate clauses in Icelandic]. Reykjavík: Institute of Linguistics.

Maling, Joan. 1980. Inversion in embedded clauses in Icelandic. *Íslenskt mál og almenn málfræði* 2:175–193 [republished 1990 in *Modern Icelandic Syntax*, ed, by Joan Maling & Annie Zaenen, 71–91. San Diego: Academic Press].

Molnár, Valéria. 2010. Stylistic Fronting and discourse. In *Tampa Papers in Linguistics*, Vol. 1, ed. by Stefan Huber & Sonia Ramírez Wohlmuth, 30–61. Department of World Languages, University of South Florida.

Ott, Dennis. 2009. Stylistic Fronting as remnant movement. *Working Papers in Scandinavian Syntax* 83:141–178.

Rayson, Paul, Oliver Charles & Ian Auty. 2012. Can Google count? Estimating search engine result consistency. In *Proceedings of the seventh Web as Corpus Workshop*, ed. by Adam Kilgarriff & Serge Sharoff, 24–31. At https://sigwac.org.uk/wiki/WAC7.

Rögnvaldsson, Eiríkur. 1984. Icelandic Word Order and *það*-insertion. *Working Papers in Scandinavian Syntax* 8.

Rögnvaldsson, Eiríkur. 2002. *ÞAÐ* í fornu máli – og síðar [*Það* in Old Norse – and later]. *Íslenskt mál og almenn málfræði* 24:7–30.

Rögnvaldsson, Eiríkur & Höskuldur Thráinsson. 1990. On Icelandic word order once more. In *Modern Icelandic Syntax*, ed. by Joan Maling and Annie Zaenen, 3–40. San Diego: Academic Press.

Sigurðsson, Halldór Ármann. 1989. *Verbal Syntax and Case in Icelandic*. Lund [republished 1992 in Reykjavík: Institute of Linguistics].

Sigurðsson, Halldór Ármann. 1997. Stylistic Fronting. Ms. University of Iceland [presented at Subjects, Expletives, and the EPP, Tromsø].

Sigurðsson, Halldór Ármann. 2010. On EPP effects. *Studia Linguistica* 64:159–189.

Sigurðsson, Halldór Ármann. 2011. Conditions on argument drop. *Linguistic Inquiry* 42:267–304.

Sigurðsson, Halldór Ármann. 2013. On Stylistic Fronting. Ms. Lund University [accessible on http://lingbuzz.auf.net/lingbuzz/001847].

Sigurðsson, Halldór Ármann. 2014. About pronouns. *Working Papers in Scandinavian Syntax* 92:65–98.

Thráinsson, Höskuldur. 1979. *On Complementation in Icelandic*. New York: Garland.

Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.

Thráinsson, Höskuldur, Ásgrímur Angantýsson & Einar Freyr Sigurðsson. 2015. *Tilbrigði í íslenskri setningagerð II. Helstu niðurstöður. Tölfræðilegt yfirlit með skýringum* [Variation in Icelandic Syntax. Main results. Statistical Overview with Explanations]. Reykjavík: Málvísindastofnun Háskóla Íslands.

Wood, Jim. 2011. Stylistic Fronting in spoken Icelandic relatives. *Nordic Journal of Linguistics* 34:29–60.