Stylistic Fronting in corpora

Sigurðsson, Halldor Armann

Published in:
Syntactic Variation in Insular Scandinavian

Published: 2017-01-01

Document Version
Peer reviewed version

Link to publication

Citation for published version (APA):
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 1980¹ 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:

---

¹ 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).

… 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).

(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


(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”

   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-bound
      e3. SF is common in (certain) subordinate clauses

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: “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**  
(i.e., clauses relativized/extracted from)  
\[ \text{okV1}^7 \quad \text{okSF} \quad \ast \text{það-V}^8 \]

b. Clauses with a **non-trace subject gap**  
\[ \text{??/okV1} \quad \text{okSF} \quad \text{okþað-V} \]

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. 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. 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. 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://sprarakbanken.gu.se/swe) gives just about 11.7m results. The tagged corpus Mörkuð íslensk máflheild (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):

<table>
<thead>
<tr>
<th>Year Range</th>
<th>Total Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1815</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>1816–1850</td>
<td>4,402</td>
<td>0.1%</td>
</tr>
<tr>
<td>1851–1899</td>
<td>77,780</td>
<td>2.2%</td>
</tr>
<tr>
<td>1900–1949</td>
<td>692,900</td>
<td>19.4%</td>
</tr>
<tr>
<td>1950–1999</td>
<td>2,185,460</td>
<td>61.3%</td>
</tr>
<tr>
<td>2000–2015</td>
<td>607,363</td>
<td>17.0%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,567,905</strong></td>
<td></td>
</tr>
</tbody>
</table>

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. 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).

<table>
<thead>
<tr>
<th>Period</th>
<th>sem hafa verið (V1)</th>
<th>sem verið hafa (SF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlimited</td>
<td>389,000</td>
<td>85,500</td>
</tr>
<tr>
<td>1970.01.01-2000.01.01</td>
<td>811</td>
<td>81</td>
</tr>
<tr>
<td>2000.01.01-2010.01.01</td>
<td>15,000</td>
<td>729</td>
</tr>
<tr>
<td>2010.01.01-2015.07.01</td>
<td>25,400</td>
<td>695</td>
</tr>
<tr>
<td>2000.01.01-2015.07.01</td>
<td>34,300</td>
<td>1,220</td>
</tr>
<tr>
<td>2005.07.01-2015.07.01</td>
<td>31,700</td>
<td>974</td>
</tr>
<tr>
<td>1970.01.01-2015.07.01</td>
<td>34,800</td>
<td>1,220</td>
</tr>
</tbody>
</table>

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. 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*.

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Google</th>
<th></th>
<th>%</th>
<th>Timarit</th>
<th></th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to 1900</td>
<td>0</td>
<td></td>
<td></td>
<td>493</td>
<td></td>
<td>1.0%</td>
</tr>
<tr>
<td>1900–1949</td>
<td>0</td>
<td></td>
<td></td>
<td>5,185</td>
<td></td>
<td>10.7%</td>
</tr>
<tr>
<td>1950–1999</td>
<td>811</td>
<td>2.3%</td>
<td></td>
<td>28,574</td>
<td>59.0%</td>
<td></td>
</tr>
<tr>
<td>2000–2015 (01.07)</td>
<td>34,300</td>
<td>97.7%</td>
<td></td>
<td>14,160</td>
<td>29.2%</td>
<td></td>
</tr>
</tbody>
</table>

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-fictional 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.\(^\text{14}\)

---

\(^{14}\) 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 Alþingisraður, 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 occurrences 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.
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 skrifð um einmitt þetta.
   this is book that there has been written about exactly this
b. Þetta er bók sem skrifð hefur verið t um einmitt þetta.
   ‘This is a book that has been written about exactly this.’
c. Þetta er bók sem __ hefur verið skrifð 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?
   ‘Does she know who has written about this?’
c. Veit hún hver __ hefur skrifað um þetta?
   ‘Who do you think has written about this?’

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

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>%SF</th>
<th>Timarit</th>
<th>%SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1: sem __ hafa verið</td>
<td>24,600</td>
<td></td>
<td>46,738</td>
<td></td>
</tr>
<tr>
<td>SF: sem verið hafa</td>
<td>1,680</td>
<td>6.4%</td>
<td>14,101</td>
<td>27.7%</td>
</tr>
<tr>
<td>V1: sem __ hafa farið</td>
<td>2,220</td>
<td></td>
<td>4,268</td>
<td></td>
</tr>
<tr>
<td>SF: sem farið hafa</td>
<td>2,170</td>
<td>49.4%</td>
<td>6,335</td>
<td>59.7%</td>
</tr>
<tr>
<td>V1: sem __ hafa lesið</td>
<td>284</td>
<td></td>
<td>1,444</td>
<td></td>
</tr>
<tr>
<td>SF: sem lesið hafa</td>
<td>150</td>
<td>34.6%</td>
<td>2,433</td>
<td>62.8%</td>
</tr>
<tr>
<td>V1 totals</td>
<td>27,104</td>
<td>12.9%</td>
<td>52,450</td>
<td></td>
</tr>
<tr>
<td>SF totals</td>
<td>4,000</td>
<td></td>
<td>22,869</td>
<td>30.4%</td>
</tr>
</tbody>
</table>

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).

(15) a1. V1: sem __ er skrifað

a2. SF: sem skrifað er

b1. V1: sem __ var skrifað

b2. SF: sem skrifað var

c1. V1: sem __ hefur verið skrifað

c2. SF: sem skrifað hefur verið

d1. V1: sem __ hafði verið skrifað

d2. SF: sem skrifað hafði verið

e1. V1: sem __ er skrifuð

e2. SF: sem skrifuð er

f1. V1: sem __ var skrifuð

f2. SF: sem skrifuð var

g1. V1: sem __ hefur verið skrifuð

g2. SF: sem skrifuð hefur verið

h1. V1: sem __ hafði verið skrifuð

h2. SF: sem skrifuð hafði verið
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).

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th></th>
<th></th>
<th>Timarit</th>
<th>#</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>%SF</td>
<td></td>
<td>%SF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1 totals</td>
<td>929</td>
<td></td>
<td>1,756</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF totals</td>
<td>1,541</td>
<td>62,4%</td>
<td>6,889</td>
<td>79,7%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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’ 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’ vs. *sem skrifað er* in different periods (search conducted July 3, 2015).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>V1: <em>sem er skrifað</em></td>
<td>44</td>
<td>90,3%</td>
<td>263</td>
<td>83,1%</td>
<td>131</td>
<td>71,8%</td>
</tr>
<tr>
<td>SF: <em>sem skrifað er</em></td>
<td>408</td>
<td></td>
<td>1289</td>
<td></td>
<td>333</td>
<td></td>
</tr>
</tbody>
</table>

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.
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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>V1: sem __ hafa búið þar</td>
<td>10</td>
<td>29%</td>
</tr>
<tr>
<td>SF: sem búið hafa þar</td>
<td>4</td>
<td>12%</td>
</tr>
<tr>
<td>SF: sem þar hafa búið</td>
<td>20</td>
<td>59%</td>
</tr>
<tr>
<td>V1: sem __ hafa búið í Danmörku</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SF: sem búið hafa í Danmörku</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SF: sem í Danmörku hafa búið</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%SF</td>
</tr>
<tr>
<td>V1: sem __ hafa búið</td>
<td>420</td>
<td>36,7%</td>
</tr>
<tr>
<td>SF: sem búið hafa</td>
<td>243</td>
<td></td>
</tr>
</tbody>
</table>

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, Paris, Stokkhólm, Berlin, Moskva, Róm, Kaupmannahöfn, Madrid, Lissabon, Ähen, Peking/Beijing, Tókýó, Japan, Pýskalandi, 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 búð as an SF candidate. Both þar and í Danmörku are complements of búð, 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 þar lásu* and *sem þar hafa lesið*, with the feminine plural pronoun þar ‘them’ (as an object), yielded 4 and 12 hits, respectively, whereas their V1 competitors *sem lásu þar* and *sem hafa lesið þar* 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 eithvuað 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 ýnislegt 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.

<table>
<thead>
<tr>
<th></th>
<th>V1: sem __ lásu bækurnar</th>
<th>SF: sem bækurnar lásu</th>
<th>SF: sem __ hafa lesið bækurnar</th>
<th>V1: sem __ hafa lesið</th>
<th>SF: sem bækurnar hafa lesið</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>14%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>18%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9 continued...

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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit.is</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#V1 SF %SF</td>
<td>#V1 SF %SF</td>
</tr>
<tr>
<td>18a: víst</td>
<td>34 0 0%</td>
<td>83 24 22.4%</td>
</tr>
<tr>
<td>18b: vissulega</td>
<td>24 52 68.4%</td>
<td>65 365 84.9%</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit.is</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#V1 SF %SF</td>
<td>#V1 SF %SF</td>
</tr>
<tr>
<td>19a: því</td>
<td>1,280 620 32.6%</td>
<td>273 940 77.5%</td>
</tr>
<tr>
<td>19b: þess vegna</td>
<td>2 0 0%</td>
<td>10 8 44.4%</td>
</tr>
<tr>
<td>19c: þar af leiðandi</td>
<td>5 4 44.4%</td>
<td>13 12 48.0%</td>
</tr>
</tbody>
</table>
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 skrífadur ‘written’ in (15i–j) above is about as frequent as SF of the bisyllabic skrífad and skrífud 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 skrífad were 66,4% (Google) and 82,4% (Timarit), and 63,1% (Google) and 70,0% (Timarit) for skrífud. For skrífadur 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):  

(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. Searching 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 Icelandic this change has been proceeding very slowly. 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 expense 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).
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 farð 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 declaratives in (21c) is það that anticipates a postposed infinitival or clausal subject.\textsuperscript{26} Anticipating það

\textsuperscript{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).\textsuperscript{27} That this is probably the case gains some credibility from the fact that the frequency of að það hefði átt in \textit{Timarit.is} markedly increases over time, as seen by the results in Table 13.

Table 13. Results (July 14, 2015) for different periods in \textit{Timarit.is} for the strings in (21).

<table>
<thead>
<tr>
<th></th>
<th>1949</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>V1: að __ hefði átt</td>
<td>83</td>
<td>27.6%</td>
<td>197</td>
<td>22.7%</td>
</tr>
<tr>
<td>SF: að átt hefði</td>
<td>62</td>
<td>20.6%</td>
<td>146</td>
<td>16.8%</td>
</tr>
<tr>
<td>PA: að það hefði átt</td>
<td>156</td>
<td>51.8%</td>
<td>524</td>
<td>60.4%</td>
</tr>
</tbody>
</table>

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.

\textsuperscript{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 idiomatized 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 grannnt 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)’.

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 þaðs).

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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>V1</td>
<td>395</td>
<td>4,5%</td>
</tr>
<tr>
<td>SF</td>
<td>4,776</td>
<td>54,2%</td>
</tr>
<tr>
<td>ÞA</td>
<td>3,647</td>
<td>(41,4%)</td>
</tr>
<tr>
<td>Totals</td>
<td>8,818</td>
<td></td>
</tr>
</tbody>
</table>

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, talð um, verîð 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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Timarit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>V1</td>
<td>468</td>
<td>3,7%</td>
</tr>
<tr>
<td>SF</td>
<td>8,285</td>
<td>66,4%</td>
</tr>
<tr>
<td>ÞA</td>
<td>3,719</td>
<td>29,8%</td>
</tr>
<tr>
<td>Totals</td>
<td>12,472</td>
<td>15,793</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td></td>
<td>V1: áður en __ er X</td>
<td>29</td>
<td>0,6%</td>
<td></td>
<td>V1: ef __ er X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SF: áður en X er</td>
<td>4,539</td>
<td>94,1%</td>
<td></td>
<td>SF: ef X er</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ÞA: áður en það er X</td>
<td>255</td>
<td>5,3%</td>
<td></td>
<td>ÞA: ef það er X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V1: eins og __ er X</td>
<td>425</td>
<td>46,1%</td>
<td></td>
<td>V1 totals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SF: eins og X er</td>
<td>490</td>
<td>53,1%</td>
<td></td>
<td>SF totals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ÞA: eins og það er X</td>
<td>7</td>
<td>0,8%</td>
<td></td>
<td>ÞA totals</td>
</tr>
</tbody>
</table>

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 verîð ‘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*).

<table>
<thead>
<tr>
<th></th>
<th>X = byrjað, farið, gengið, talað um</th>
<th>X = verið</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>V1: ef ___ er X</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0,2%</td>
<td>0,2%</td>
</tr>
<tr>
<td>SF: ef X er</td>
<td>2,313</td>
<td>951</td>
</tr>
<tr>
<td></td>
<td>74,0%</td>
<td>26,3%</td>
</tr>
<tr>
<td>PA: ef það er X</td>
<td>806</td>
<td>2,650</td>
</tr>
<tr>
<td></td>
<td>25,8%</td>
<td>73,4%</td>
</tr>
<tr>
<td>V1: eins og ___ er X</td>
<td>18</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>6,8%</td>
<td>61,8%</td>
</tr>
<tr>
<td>SF: eins og X er</td>
<td>240</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>91,3%</td>
<td>37,9%</td>
</tr>
<tr>
<td>PA: eins og það er X</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1,9%</td>
<td>0,3%</td>
</tr>
</tbody>
</table>

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’  

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’  

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 (*byrjado, 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 *byrjado* 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.

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 (*byrjado, 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.

<table>
<thead>
<tr>
<th></th>
<th>Indicative (<em>er</em>)</th>
<th>Subjunctive (<em>sé</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td><strong>V1:</strong> að _ <em>er</em> <em>verið</em> / að _ <em>sé</em> <em>verið</em></td>
<td>276</td>
<td>2.8%</td>
</tr>
<tr>
<td><strong>SF:</strong> að <em>verið</em> <em>er</em> / að <em>verið</em> <em>sé</em></td>
<td>1,740</td>
<td>18.6%</td>
</tr>
<tr>
<td><strong>þA:</strong> að <em>það</em> <em>er</em> <em>verið</em> / að <em>það</em> <em>sé</em> <em>verið</em></td>
<td>7,620</td>
<td>79.1%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Google</th>
<th>%V1</th>
<th>%V1</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1: áður en ___ ferfór að rigna</td>
<td>9</td>
<td>56,2%</td>
<td>36</td>
<td>85,7%</td>
</tr>
<tr>
<td>ÞA: áður en það ferfór að rigna</td>
<td>7</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>V1: þegar ___ ferfór að rigna</td>
<td>17</td>
<td>68,0%</td>
<td>132</td>
<td>87,4%</td>
</tr>
<tr>
<td>ÞA: þegar það ferfór að rigna</td>
<td>8</td>
<td></td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

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 Thráinsson 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 \(pad\)-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 \(pad\)-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 \(pad\)-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

\(^{31}\) 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.


