

Cataphora, expletives and impersonal constructions in the history of Icelandic

Hannah Booth

2019

NB: This is a post-print version reflecting changes made in the peer review and editing process, but not the publisher’s PDF. This article appeared in *Nordic Journal of Linguistics* 42(2), pages 139–164. When citing, please use the page numbers given there. The publisher should be contacted for permission to re-use or reprint the material in any form. For the published version, see <https://www.cambridge.org/core/journals/nordic-journal-of-linguistics/article/abs/cataphora-expletives-and-impersonal-constructions-in-the-history-of-icelandic/3ACEBCE5235D23B274AB7578>

Abstract

This paper presents a diachronic account for the emergence of the expletive *það* in Icelandic impersonal constructions. Using data from the IcePaHC corpus (Wallenberg et al., 2011), I show that a cataphorically referential *það* functions as a topic position placeholder in Old Icelandic (1150-1350) in impersonals with a clausal argument. The corpus findings indicate that *það* spread from this early cataphoric context to impersonals which lack the clausal argument, with ‘say-type predicates’ acting as a bridging context. Strikingly, this coincides with another change whereby cataphoric *það* becomes increasingly restricted to the topic position in constructions with a clausal subject. I interpret this as evidence that cataphoric *það* is losing its subject status in such contexts and becoming a topic position placeholder, in line with its function in impersonals. This sheds light on the mixed status of cataphoric *það* in modern Icelandic and challenges the ‘Prefinite First Hypothesis’ for the diachrony of Germanic expletives (e.g. Faarlund, 1990).

Keywords: Cataphora, expletives, historical corpora, Icelandic, impersonals, syntactic change

1 INTRODUCTION

Present-day Icelandic has one main form which functions as an expletive: *það*, formally identical to the 3SG.NT pronoun.¹ It is well known that expletive *það* is positionally restricted to the clause-initial prefinite position (e.g. Sells, 2005, Sigurðsson, 2007, Thráinsson, 2007, 312), as illustrated in (1).²

- (1) a. **Það** var dansað í gær.
EXPL be.PST dance.PASS.PTCP yesterday
‘It was danced yesterday.’
b. Í gær var (***það**) dansað.
yesterday be.PST (*EXPL) dance.PASS.PTCP
‘Yesterday it was danced.’

¹*Hann*, formally identical to the 3SG.MASC pronoun, is an alternative to *það* in constructions with weather predicates in present-day Icelandic (e.g. Eythórrsson & Sigurðardóttir, 2016).

²Glossing throughout follows the Leipzig Glossing Rules: <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>. I use CATPH to gloss a cataphorically referential *það* and EXPL to gloss an expletive (i.e. non-referential) *það*.

The restricted positional distribution of expletive *það* is standardly interpreted as evidence that it does not qualify as a subject, since subjects in a Germanic verb-second language like Icelandic occur in the immediately postfinite position in contexts like (1-b) (see Thráinsson, 1979, 480–481, Platzack, 1983, Sigurðsson, 2007). Compare the expletive *det* in Swedish, which does behave like a subject, e.g. (2) (for details on the distribution of the Swedish expletive, see Falk, 1993 and Håkansson, 2017).

- (2) a. **Det** dansades i går.
EXPL dance.PST.PASS yesterday
'It was danced yesterday.'
- b. I går dansades **det**.
yesterday dance.PST.PASS EXPL
'Yesterday it was danced.'

The clause-initial prefinite position in Icelandic is standardly assumed to be an information-structurally privileged position which is associated with topical constituents, henceforth TOPIC POSITION (e.g. Rögnvaldsson & Thráinsson, 1990). As such, it has been claimed that the clause-initial expletive in examples like (1-a) signals a topicless verb-second (V2) sentence (Zaenen, 1983, Rögnvaldsson & Thráinsson, 1990, Sells, 2005). *Það* in such contexts is a TOPIC POSITION PLACEHOLDER.³

When one examines topicless impersonal sentences in Old Icelandic (1150-1350), expletive *það* is typically absent, rendering verb-initial (V1) structures, e.g. (3).

- (3) a. Var þá hleypt suður til Reykja...
be.PST then gallop.PASS.PTCP south to Reykir
'It was galloped south to Reykir...' (1250, Sturlunga.406.624)⁴
- b. Skal af því víst á daginn berjast...
shall therefore certainly on day.DEF fight.INF
'One shall therefore certainly fight on this day.' (1300, Alexander.1358)

Previous studies claim that the expletive did not emerge in such contexts until c.1500 (Rögnvaldsson, 2002) and underwent a dramatic increase in frequency in the 19th century (Hróarsdóttir, 1998).

In other constructions – namely those with a clausal argument – *það* is however robustly attested in Old Icelandic ('extraposition', Faarlund, 1990, Rögnvaldsson, 2002). This applies both to constructions where the clausal argument is the subject of the matrix clause predicate, e.g. (4-a), as well as those where the clausal argument is the object of the matrix clause predicate, e.g. (4-b).

- (4) a. **Það**_i var rétt [að spakur engill boðaði Guð borinn spökum
CATPH be.PST right COMP wise angel proclaim.PST God born wise.DAT
Gyðingum]_i.
Jews.DAT
'It was right that a wise angel proclaimed God born to wise Jews.'
(1150, Homiliubok.1319)

³Faarlund (1990, 64–65) refers to the positionally restricted clause-initial expletive in (1) as an 'expletive topic' but I prefer the term 'topic position placeholder'. Since the expletive lacks semantics, it cannot be considered a topic in any information-structural sense. Rather, it is merely a structural filler for the topic position.

⁴Unless otherwise stated, all examples are from IcePaHC and are referenced in the form: Year, Text.CorporusID. All IcePaHC texts are normalised to modern orthography, regardless of their date. I follow this normalisation throughout.

- b. ...og **það**_i sér Finnbogi [að hann var hlífarlaus]_i...
 and CATPH see.PRS Finnbogi.NOM COMP he.NOM be.PST uncovered
 ‘..and Finnbogi sees that he was uncovered...’ (1350, Finnbogi.657.1902)

In contexts like (4), I assume that *það* has cataphoric reference to the clausal argument (henceforth CATAPHORIC *það*) and is distinct from EXPLETIVE *það*, which is neither cataphorically nor anaphorically referential.⁵ Another property which distinguishes expletive *það* from cataphoric *það* is the fact that the expletive is indeclinable, whereas cataphoric *það* inflects for case. The example in (5) features the predicate *geta* ‘guess’, which takes a genitive argument as its object; cataphoric *það* has genitive case marking (*þess*).

- (5) En **þess** get ég fyrst [að þú sért maður íslenskur]...
 but CATPH.GEN guess.PRS I.NOM first COMP you.NOM be.PRS.SBJV man Icelandic
 ‘But first I guess that you are an Icelandic man...’ (1275, Morkin.1564)

In this paper, I present an account for the diachronic development whereby *það* spread from cataphoric contexts like (4) to contexts like (3) which lack a clausal argument. In this account, the early attested cataphoric *það* plays a role in the later emergence of expletive (non-cataphoric) *það*. The structure of the paper is as follows: Section 2 outlines the methodology of the corpus study. Section 3 presents findings for the status of cataphoric *það* in various constructions types in the history of Icelandic and discusses a structurally ambiguous construction, using Lexical-Functional Grammar (LFG) for the formal analysis. Section 4 presents an account for the development whereby expletive *það* emerged on the model of earlier cataphoric contexts and compares this with previous findings for the rise of *það* in presentational constructions (Booth, 2018). Section 5 concludes the paper.

2 METHODOLOGY

The basis for this study is data from the Icelandic Parsed Historical Corpus (‘IcePaHC’, Walenberg et al., 2011). IcePaHC contains approximately 1,000,000 words, from 61 text extracts spanning 10 centuries (1150-2008), thereby covering all attested stages of Icelandic. IcePaHC thus allows one to examine change across the centuries which many studies focusing on data from specific periods do not capture (e.g. Hróarsdóttir, 1998, Rögnvaldsson, 2002).

The IcePaHC annotation follows the Penn treebank format established for historical English (e.g. Kroch & Taylor, 2000, Santorini, 2010) and is compatible with the CorpusSearch query language (Randall, 2005). All content is lemmatised, part-of-speech tagged and annotated for constituent structure, with additional tagging for certain grammatical functions (e.g. subject, object). For further information on IcePaHC, see Rögnvaldsson et al. (2012). The corpus does have some limitations: the texts included represent a very small sample of attested historical Icelandic, and certain genres are over-represented and others under-represented. Moreover, the ‘narrative’ genre which is dominant in the corpus comprises mainly saga-style texts up to the 19th century but modern novels after that, and thus cannot be considered a homogenous category across time. Some texts are also based on source texts in other languages. Nevertheless, the advantages offered by the syntactic annotation outweigh these issues, and as long as one keeps these limitations in mind, IcePaHC is a valuable source of data.

Both overt and ‘null’ expletives – constructions where expletive *það* could be expected but is not attested, e.g. (3) – are distinctively tagged in IcePaHC. For sake of time, I do not look beyond the tagged instances for overt or ‘null’ expletives which are not tagged as such in the corpus. However, a parallel manual investigation of expletives in Old Icelandic presented in

⁵Thráinsson (1979) claims that prefinite *það* in contexts like (4) can be either referential or expletive, while postfinite *það* is always referential. Instead, I assume that a *það* in any position which co-occurs with a clausal argument is cataphoric.

Booth (2018) yielded comparable results to the IcePaHC findings for early texts, suggesting the IcePaHC tagging of expletives is reasonably reliable. Combining the overt and ‘null’ expletives gives a dataset which represents the total contexts in which the expletive could potentially occur (henceforth ‘expletive contexts’). This allows me to go beyond previous studies which only count the number of sentences where expletive *það* is present, without taking into account the overall number of contexts in which the expletive could plausibly occur (Hróarsdóttir, 1998, Rögnvaldsson, 2002). Impersonal constructions which qualify as expletive contexts were isolated via CorpusSearch queries (see Booth, 2018 for details). To make the investigation manageable, I restrict the study to matrix clauses.

The IcePaHC treatment of constructions with cataphoric *það* and a clausal argument is not consistent. Sometimes *það* is tagged as an expletive (‘ES’), sometimes as a referential pronoun (‘PRO’), depending on the subject properties of the clause. A variety of CorpusSearch queries are therefore required in order to capture all examples of cataphoric *það* (see again Booth, 2018). ‘Null’ instances of cataphoric *það* – contexts where it could be expected but is not attested – are also tagged. Combining the overt and ‘null’ instances thus gives a total dataset of potential cataphoric contexts. As with the expletive contexts, I only include matrix clauses in the cataphoric dataset.

Once the data was collected, I manually examined each example to check whether the tagging was correct. Once erroneous examples were removed, I tagged the remaining examples for additional properties relevant to the investigation: position of *það* if present; verb position; predicate type. The two datasets which form the basis of the study are shown in Table 1. I show the number of contexts for each dataset as a proportion of the total number of matrix clauses to give an idea of relative frequency across time stages.

Time period	Expletive contexts		Cataphoric contexts		All matrix clauses <i>n</i>
	<i>n</i>	% of total	<i>n</i>	% of total	
1150-1350	441	2.2%	378	1.9%	19,771
1351-1550	273	1.7%	163	1.0%	15,785
1551-1750	192	1.6%	183	1.5%	12,090
1751-1900	220	1.8%	172	1.4%	12,357
1901-2008	250	2.1%	227	1.9%	11,703
All periods	1376	1.9%	1123	1.6%	71,706

Table 1: Expletive and cataphoric contexts in IcePaHC.

As shown in Table 1, I split the IcePaHC diachrony into five periods. Periodising the data in this way has the advantage of abstracting away from the fact that the individual texts in IcePaHC are not evenly distributed across time. Throughout the study, I compare the proportion of examples with *það* to those without, over the five periods. This allows me to assess the strength of preference for *það* in the various contexts historically.

3 CATAPHORIC *ÞAÐ* IN DIACHRONY

As already shown, Icelandic exhibits constructions where a cataphoric *það* co-occurs with a clausal argument, see (4). In this paper, I adopt an LFG analysis for such constructions which has been proposed by Berman et al. (1998) for parallel constructions in German (*es* plus clausal argument). I return to the architecture of LFG in section 3.3; for a general account, see Bresnan et al. (2015). This analysis assumes that the cataphoric element and the clausal argument contribute to the same argument slot of the matrix predicate; the information contributed by each constituent unifies under the same grammatical function at LFG’s f-structure (the functional

dimension).⁶ For constructions like (4-a), I assume that both *það* and the clausal argument map to the SUBJ(ect) function. For constructions like (4-b), I assume that *það* and the clausal argument map to the OBJ(ect) function.⁷ By contrast, expletive *það* in constructions like (1-a) is a topic position placeholder and is exclusively motivated for structural reasons; it does not map to a grammatical function. In LFG terms, expletive *það* is thus an element which is present at c-structure (the constituent structure dimension) but has no representation at f-structure (the functional dimension) (see Lødrup, 2011).

3.1 Constructions with a clausal subject

In earlier Icelandic (1150-1750), cataphoric *það* in constructions like (4-a) qualifies as a subject. Evidence for this comes from its positional distribution in such contexts. Firstly, in sentences with TOPICALIZATION – where a non-subject constituent occupies the sentence-initial position – the dominant pattern is for cataphoric *það* to be present in the immediately postfinite position, e.g. (6). This is typical subject behaviour in a Germanic V2 language like Icelandic.

- (6) Satt er **það**_i [að mikið afgangur er Grettir annarra manna]_i...
 true be.PRS CATPH COMP great paragon be.PRS Grettir other.GEN men.GEN
 ‘It is true that Grettir is a great paragon of other men...’ (1310, Grettir.1695)

The IcePaHC data shows that a postfinite *það* is overwhelmingly present in topicalization contexts in the data for 1150-1750 (81.8%), see Table 2.⁸ The frequency of postfinite *það* is roughly comparable with the frequency at which *það* occurs in the clause-initial prefinite position in sentences without topicalization (86.4%), see Table 3.

Time period	Postfinite <i>það</i>	No <i>það</i> (V2)	Total	% Postfinite <i>það</i>
1150-1750	27	6	33	81.8%

Table 2: Frequency of postfinite *það* in constructions with a clausal subject and topicalization in IcePaHC, 1150-1750.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1750	89	14	103	86.4%

Table 3: Frequency of prefinite *það* in constructions with a clausal subject in IcePaHC, 1150-1750.

The second piece of evidence which supports the subject status of cataphoric *það* in this context is the fact that it occurs in the immediately postfinite position in yes/no-interrogatives, another key positional property of subjects in a Germanic V2 language. An early example is shown in (7).

⁶More specifically, Berman et al. (1998) assume that both the pronoun and the clausal argument contribute information towards the PRED feature of the same grammatical function: *es* introduces a variable in the form of a canonical PRED value [PRED ‘pro’]; the clausal argument introduces a semantic restriction on this value [RESTR [PRED ‘...’]]. This allows a grammatical function to be realised by independent c-structure constituents without violating the functional uniqueness condition on f-structure, which rules out two distinct appearances of the same grammatical function within a single f-structure (see Bresnan et al., 2015, 45). For comparable unification analyses in LFG, see Andrews (1990) on clitic doubling in Spanish and Kuhn (1999, 2001) on split noun phrase constructions in German.

⁷I follow Thráinsson (1979) in assuming that complement clauses in Icelandic are assigned the OBJ(ect) function, not the COMP(lement) function; see also Dalrymple & Lødrup (2000).

⁸In this part of the study I collapse the three periods prior to 1750 into one, since *það* shows similar behaviour in all three periods and the amount of data for each individual period is small.

- (7) Er **það**_i satt, Halli, [að þú hefir eigi hefnt föður þíns]_i?
 be.PRS CATPH true Halli COMP you.NOM have.PRS NEG avenge.PST father you.GEN
 ‘Is it true, Halli, that you have not avenged your father?’ (1275, Morkin.1203)

Yes/no-interrogatives which feature clausal subjects are very rare in IcePaHC; I have only found the single example in (7) in which a postfinite *það* is present. Similar examples can however be found in other Old Icelandic texts, e.g. (8).

- (8) a. Er **þat**_i satt [at Hrólfr kraki ok berserkir hans flýja hvárki eld né járn]_i?
 be.PRS EXPL true COMP Hrólfr kraki and berserks he.GEN flee.PRS neither fire nor iron
 ‘Is it true that Hrólfr kraki and his berserks flee neither fire nor iron?’
 (Snorra Edda: Skáldskaparmál, 59)
- b. Er **þat**_i satt [at annat líf er en þetta]_i?
 be.PRS EXPL true COMP other life be.PRS than DEM
 ‘Is it true that there is another life than this one?’ (Ceciliu Saga Meyjar, 6)

Of course, the fact that there is no example in IcePaHC which lacks a postfinite *það* does not rule out the possibility that such structures could occur, but (7) and (8) are at least clear evidence that a postfinite cataphoric *það* is possible in yes/no-interrogatives.

The third piece of positional evidence which supports the subject status of cataphoric *það* comes from sentences like (9), which are V1 declaratives where *það* occurs in the immediately postfinite position.

- (9) Og er **það**_i mitt ráð [að þér farið upp á húsin]_i.
 and be.PRS CATPH my advice COMP you.NOM go.PRS up on house.DEF
 ‘And it is my advice that you go up on the house.’ (1250, Sturlunga.415.836)

In the IcePaHC data for 1150-1750, a postfinite cataphoric *það* is present at a frequency of 78.1% in such contexts, see Table 4.

Time period	Postfinite <i>það</i>	No <i>það</i> (V1)	Total	% Postfinite <i>það</i>
1150-1750	50	14	64	78.1%

Table 4: Frequency of postfinite *það* in V1 declaratives with a clausal subject in IcePaHC, 1150-1750.

V1 declaratives are a marked construction in Icelandic; they are still a feature of the present-day language but occur more frequently in earlier stages (e.g. Butt et al., 2014). A particular type of V1 declarative is the NARRATIVE INVERSION construction (Platzack, 1985, Sigurðsson, 1990) which has a topical subject – prototypically a pronominal – in the immediately postfinite position, as in the second sentence in (10).

- (10) Auðun tekur nú að auka sína ferð slíkt er hann má. Þorir
 Auðun begin.PRS now to speed-up.INF his-own journey such as he.NOM may dare.PRS
hann þá eigi að stefna til gatnanna.
 he.NOM then NEG to go.INF to paths.DEF
 ‘Auðun now beings to speed up his journey such as he may. He then dares not make for the paths.’ (1250, Sturlunga.445.2015)

The fact that the immediately postfinite position is available to prototypical subjects in the narrative inversion construction suggests that immediately postfinite cataphoric *það* in V1 declaratives

like (9) can be analysed as a subject.

While the occurrence of cataphoric *það* in constructions with a clausal subject is already known for earlier Icelandic (Faarlund, 1990, Rögnvaldsson, 2002), the diachronic behaviour of this element has not been previously examined. The IcePaHC findings show a striking change which is visible in the data as of 1751 whereby *það* becomes increasingly restricted to the clause-initial prefinite position.⁹ Firstly, there is a decrease in the frequency at which *það* occurs in the immediately postfinite position in sentences with topicalization, see Table 5. By the modern period (1901-2008), *það* is only present in 38.9% of instances.

Time period	Postfinite <i>það</i>	No <i>það</i> (V2)	Total	% Postfinite <i>það</i>
1150-1750	27	6	33	81.8%
1751-1900	5	4	9	55.6%
1901-2008	7	11	18	38.9%
All periods	39	21	60	

Table 5: Frequency of postfinite *það* in constructions with a clausal subject and topicalization in IcePaHC, 1150-2008.

Secondly, in V1 declaratives there is also a decrease in the frequency at which *það* occurs in the immediately postfinite position in the periods after 1751, see Table 6.

Time period	Postfinite <i>það</i>	No <i>það</i> (V1)	Total	% Postfinite <i>það</i>
1150-1750	50	14	64	78.1%
1751-1900	6	8	14	42.9%
1901-2008	1	13	14	7.1%
All periods	57	35	92	

Table 6: Frequency of postfinite *það* in V1 declaratives with a clausal subject in IcePaHC, 1150-2008.

By contrast, clause-initial prefinite cataphoric *það* – which is ambiguously either a subject or a topic position placeholder – does not undergo a decrease, see Table 7.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1750	89	14	103	86.4%
1751-1900	41	8	49	83.7%
1901-2008	72	13	85	84.7%
All periods	202	35	237	

Table 7: Frequency of prefinite *það* in constructions with a clausal subject in IcePaHC, 1150-2008.

I interpret the approximately simultaneous decrease in *það* in two contexts in which it is unambiguously a subject – together with the stable status of prefinite *það* – as indication that cataphoric *það* in constructions with a clausal subject is losing its subject status and transitioning towards becoming a placeholder for the topic position, i.e. the same function that expletive *það* serves in impersonal constructions, see (1) above.

This diachronic account is interesting for two reasons. Firstly, it offers an explanation for the synchronic status of cataphoric *það* in this context in modern Icelandic. It has been observed

⁹Of course, it is impossible to connect such a change to a single year (e.g. 1751) but this is the point at which the change at least starts to become visible in the IcePaHC data.

that cataphoric *það* in constructions with a clausal subject in modern Icelandic exhibits a mixed positional distribution, whereby its occurrence in a position in which it is unambiguously a subject (the immediately postfinite position) is possible, but dispreferred (see Rögnvaldsson, 2002; Thráinsson, 1979; Thráinsson, 2007, 312–313). Rögnvaldsson (2002, 12) highlights the acceptability of all three variants in (11). *Það* is typically present in the clause-initial position in sentences without topicalization, e.g. (11-a); in topicalization contexts, it is both acceptable for *það* to be absent as in (11-b), or present in postfinite position as in (11-c). According to Rögnvaldsson, the variant without *það* (11-b) is more common than the ‘exceptional’ type with a postfinite *það* in (11-c).

- (11) a. **Það** er ótrúlegt [að enginn hafi tekið eftir þessu].
 CATPH be.PRS unbelievable COMP no-one have.PRS.SBJV notice.PST.PTCP
 DEM.DAT
- b. Ótrúlegt er [að enginn hafi tekið eftir þessu].
 unbelievable be.PRS COMP no-one have.PRS.SBJV notice.PST.PTCP DEM.DAT
- c. Ótrúlegt er **það** [að enginn hafi tekið eftir þessu].
 unbelievable be.PRS CATPH COMP no-one have.PRS.SBJV notice.PST.PTCP
 DEM.DAT
 ‘It is unbelievable that no one has noticed this.’

Standard accounts assume that postfinite *það* in contexts like (11-c) is a referential pronoun, while prefinite *það* can be either an expletive or a referential pronoun (as first proposed by Thráinsson, 1979; see also Thráinsson, 2007, 312–313). Furthermore, Thráinsson (1979, 481) interprets the synchronic observation whereby it is possible for some speakers to ‘raise’ cataphoric *það* as evidence for the fact that it is developing from a placeholder for the clause-initial prefinite position (‘surface adjustment particle’) into a subject, e.g. (12).

- (12) Jón telur **það** vera líklegt að María sé fífl.
 Jón believe.PRS CATPH be.INF likely COMP María be.PRS.SBJV fool
 ‘Jón believes it to be likely that María is a fool.’ (Thráinsson, 1979, 481)

The diachronic account I have proposed here offers an alternative explanation, claiming that cataphoric *það* optionally exhibits some subject properties in the modern language as a reflex of its older status as a subject. This account assumes a change for cataphoric *það* in the opposite direction to that suggested by Thráinsson (1979, 481): from subject to clause-initial prefinite position placeholder.

Secondly, from a broader cross-Germanic perspective, my diachronic account for ‘subject’ cataphoric *það* goes against standard accounts for the emergence of expletives in Germanic, which can be summarized as the PREFINITE FIRST HYPOTHESIS, see (13) (e.g. Breivik, 1990 and Ingham, 2001 on English, Lenerz, 1985 on German, Falk, 1993 on Swedish, Faarlund, 1990 on Norwegian; for more general overviews, see Haiman, 1974, Richards & Biberauer, 2005, Silva-Villar, 1996.)

- (13) *The Prefinite First Hypothesis*
 Prefinite expletive > subject expletive

The Prefinite First Hypothesis states that the general pathway whereby expletives emerge in Germanic is that positionally restricted expletives first appear in the clause-initial prefinite position, and only later generalise to all positions and thereby reach ‘subject expletive’ status. In line with this assumed trend, it has been claimed that the initial rise of expletives in Germanic is conditioned by structural considerations concerning V2 (i.e. as a strategy to occupy the

clause-initial prefinite position, e.g. Richards & Biberauer, 2005).

The account which I have proposed here for the development of cataphoric *það* in Icelandic challenges the Prefinite First Hypothesis. I have shown that cataphoric *það* had subject status prior to 1750, and only relatively late in the diachrony began to lose its subject status, transitioning to becoming a structural placeholder for the clause-initial prefinite position (i.e. a prefinite expletive). The proposed change thus operates in the opposite direction to that predicted by the Prefinite First Hypothesis: a prefinite expletive emerges from a subject element. Moreover, the Icelandic findings are in line with studies on Old High German (Axel, 2007) and Middle Norwegian (Kinn, 2016) which have similarly challenged the Prefinite First Hypothesis; comparative investigations in this area would be interesting to pursue in future.

3.2 Constructions with a clausal object

In this part of the study, I investigate the diachronic status of cataphoric *það* in constructions with a clausal object, e.g. (4-b) above. Such constructions are defined as those whose matrix clause predicate has active morphology and where the clausal argument would map to object for that predicate’s argument structure (I discuss mapping relations in section 3.3). Specifically, I examine a subset of such constructions which lack an overt subject in the matrix clause (IMPERSONAL CONSTRUCTIONS), e.g. (14).

- (14) En **það**_i er vitanda, [að þá göfgum vér réttlega postula
 but CATPH be.PRS know.PRS.PTCP COMP then honour.PRS we.NOM rightly apostles
 guðs alla]_i...
 god.GEN all...
 ‘But one should know that we then rightly honour all of God’s apostles...’
 (1150, Homiliubok.304)

Assuming the unification analysis outlined at the beginning of Section 3 – whereby the cataphoric element and the clausal argument contribute to the same argument slot of the matrix predicate – in such contexts *það* ought to qualify as an object. However, the IcePaHC findings regarding the positional distribution of *það* in such contexts challenge this assumption. The first relevant observation is that prefinite cataphoric *það* in constructions with a clausal object is optional and shows no clear change throughout the diachrony, see Table 8.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1350	12	11	23	52.2%
1351-1550	5	6	11	45.5%
1551-1750	5	3	8	62.5%
1751-1900	7	2	9	77.8%
1901-2008	5	4	9	55.6%
All periods	34	26	60	

Table 8: Frequency of prefinite *það* in constructions with a clausal object in IcePaHC, 1150-2008.

The second relevant observation is that *það* is virtually restricted to the prefinite position in these contexts; it is attested only very infrequently in the postfinite position in topicalization contexts, see Table 9.

Time period	Postfinite <i>það</i>	No <i>það</i> (V2)	Total	% Postfinite <i>það</i>
1150-1350	1	50	51	2.0%
1351-1550	2	21	23	8.7%
1551-1750	2	31	33	6.1%
1751-1900	1	5	6	16.7%
1901-2008	0	15	17	0.0%
All periods	6	122	128	

Table 9: Frequency of postfinite *það* in constructions with a clausal object and topicalization in IcePaHC, 1150-2008.

As an illustration of this positional restriction, compare the example with *það* in (15-a) with the example without in (15-b), which is taken from the same text.

- (15) a. **Það_i** er nú næst að segja frá Alexandro konungi, [að hann
 CATPH be.PRS now next to say.INF from Alexander king COMP he.NOM
 fer yfir ána Euphraten með allt lið sitt]_i.
 go.PRS over river.DEF Euphrates with all company his-own
 ‘Next one can now say of King Alexander that he travels over the river Euphrates
 with all of his company.’
 (1300, Alexander.1102)
- b. Nú er að segja frá Alexandro, [að, hvar sem hann fer, þá...].
 now be.PRS to say.INF from Alexander COMP wherever he.NOM go.PRS RSMP
 ‘Now one can say of Alexander that, wherever he goes, then....’
 (1300, Alexander.396)

Due to its positional restriction, it does not seem reasonable to analyse cataphoric *það* in these contexts as an object. Rather, the IcePaHC findings challenge the unification analysis outlined above and suggest that *það* instead qualifies as a structural placeholder for the clause-initial prefinite position. Moreover, an examination of the information-structural properties of the examples with clause-initial *það* reveals that such examples typically initiate a new discourse and exclusively express discourse-new information, i.e. lack a topic. An example was already shown in (15-a). Further examples are shown in (16).

- (16) a. **Það_i** er af Helga að segja [að hann fékk fæð mikla]_i.
 CATPH be.PRS of Helgi to say.INF COMP he.NOM get.PST melancholy great
 ‘One can say of Helgi that he grew greatly melancholy.’ (1400, Gunnar.876)
- b. **Það_i** er skjótast að segja frá för Alexandri, [að hann sækir
 EXPL be.PRS quickest to say.INF from journey Alexander COMP he.NOM seek.PRS
 fund Darii]_i...
 meeting Darius
 ‘One can say quickest of Alexander’s journey that he seeks a meeting with Darius...’
 (1300, Alexander.1113)

As such, cataphoric *það* appears to be a structural placeholder for the topic position when there is no appropriate constituent to occupy that position. This is the same function as that served by expletive *það* in impersonal constructions with no clausal object in modern Icelandic, see (1) above.

I have thus shown that the topic position placeholder *það* – generally assumed to be a relatively recent phenomenon in Icelandic (see Section 1) – in fact has a long history in the language and is already solidly attested in topicless impersonal constructions with a clausal

object in Old Icelandic (1150-1350). In Section 4, I claim that this status of cataphoric *það* in early Icelandic serves as a model from which *það* emerges as a topic position placeholder in a wide range of impersonal constructions in later stages of the language.

3.3 Structurally ambiguous constructions

3.3.1 Mapping Theory within LFG

So far, I have shown that two types of cataphoric *það* were already established at an early stage of Icelandic:

1. Cataphoric *það* in constructions with a clausal subject, which – at least in earlier Icelandic (pre-1750) – qualifies as a subject.
2. Cataphoric *það* in constructions with a clausal object, which functions as a topic position placeholder.

I now introduce a third, structurally ambiguous type of construction where cataphoric *það* and a clausal argument co-occur, e.g. (17). The predicate in this type is a passive transitive which can take a clausal complement, e.g. *segja* ‘say’. I henceforth refer to such predicates as SAY-TYPE PREDICATES.

- (17) **Það**_i er sagt [að Bárður bóndi átti sætur]_j.
 CATPH be.PRS say.PASS.PTCP COMP Bárður farmer own.PST mountain-pastures
 ‘It is said that Bárður the farmer owned mountain pastures.’
 ‘There is said that Bárður the farmer owned mountain pastures.’
 (1350, Finnbogi.636.641)

I argue that constructions like (17) are structurally ambiguous with respect to the grammatical function which is assigned to the clausal argument and allow for at least two analyses:¹⁰

1. As a promotional passive construction, in which the clausal argument maps to the SUBJ(ect) function.
2. As a non-promotional passive construction, in which the clausal argument maps to the OBJ(ect) function.

The parallel architecture of LFG offers many advantages for modelling this distinction between promotional and non-promotional passives, since different dimensions of linguistic information are handled at independent, interacting levels of structure (see e.g. Asudeh & Toivonen, 2009, Bresnan et al., 2015). There are three such levels which are relevant to this paper:

1. c-structure, which captures information about category and constituency.
2. f-structure, which captures information about grammatical functions, grammatical features and grammaticalised discourse functions.
3. a-structure, which captures information about the arguments required by a predicate and their thematic roles.

¹⁰Another possibility is that (17) is a construction which has passive morphology but is a syntactically active impersonal with a suppressed subject (see Booth, 2018 for discussion). There is a precedent for such constructions in Icelandic: other authors who have argued that certain morphologically passive constructions allow for an analysis as syntactically active impersonals are Sigurðsson & Egerland (2009) for impersonal passives and Maling & Sigurjónsdóttir (2002) for the New Impersonal/Passive, although the latter analysis remains disputed (Eythórsson, 2008, Jónsson, 2009). To show this conclusively would require data which IcePaHC cannot provide, and so I do not discuss this further and leave the issue open for future research.

I adopt the the view of a-structure assumed by Kibort (2007, 2008, 2014) and developed as part of her revision of (Lexical) Mapping Theory, which is a tool for mapping between grammatical functions (at f-structure) and arguments (at a-structure), as I show below. The Kibortian view of a-structure is shown in (18). It consists of two levels of information: an ‘argument position’ level and a ‘semantic participant’ level.

$$(18) \quad \begin{array}{ccccccc} & & x & & y & & \text{(semantic participants)} \\ & & | & & | & & \text{(participant-argument mapping)} \\ \text{predicate} & < & \text{arg}_1 & & \text{arg}_2 & & \dots & > & \text{(argument positions)} \\ & & [-o/-r] & & [-r] & & \dots & & \text{(syntactic pre-specification)} \end{array}$$

The argument position level is the core component of a-structure. At this level, any predicate has access to the universal subcategorization frame in (19), from which it selects certain arguments (Kibort & Maling, 2015, 152). In this paper, I only deal with arg_1 and arg_2 .

$$(19) \quad < \begin{array}{ccccccc} \text{arg}_1 & \text{arg}_2 & \text{arg}_3 & \text{arg}_4 & \dots & \text{arg}_n & \\ [-o/-r] & [-r] & [+o] & [-o] & & [-o] & \end{array} >$$

Each argument position in (19) is associated with an intrinsic syntactic classification expressed in terms of the binary features $[\pm r(\text{restricted})]$ and $[\pm o(\text{bjective})]$. This serves to specify what type of grammatical function each argument position can in principle map to, as summarised in Table 10 (see Bresnan & Kanerva, 1989). OBJ_θ is an abbreviation for secondary objects; OBL_θ abbreviates multiple oblique functions.

	[-r]	[+r]
[-o]	SUBJ	OBL_θ
[+o]	OBJ	OBJ_θ

Table 10: Decomposition of grammatical functions into features.

arg_1 is specified as either $[-o]$ or $[-r]$, depending on the predicate type: for transitive and unergative verbs, arg_1 is associated with $[-o]$; for unaccusative verbs, arg_1 is associated with $[-r]$. In this paper I only deal with mappings for transitive predicates, so arg_1 will always be $[-o]$.

A transitive predicate like *segja* ‘say’ in (17) takes two arguments: arg_1 , Agent and arg_2 , Theme. The default argument-function mapping for a say-type predicate is as in (20).

$$(20) \quad \begin{array}{ccccccc} & & \text{Agent} & & \text{Theme} & & \\ & & | & & | & & \\ \text{say (default)} & < & \text{arg}_1 & & \text{arg}_2 & & > \\ & & [-o] & & [-r] & & \\ & & | & & | & & \\ & & \text{SUBJ} & & \text{OBJ} & & \end{array}$$

Passivisation is an operation which results in a change in the default mapping between arguments and grammatical functions. In Mapping Theory, passivisation is understood as comprising two components: demotion and (potentially) promotion. When a transitive predicate like *segja* ‘say’ is passivised, the highest argument (already specified as $[-o]$) receives an additional specification that it must map onto a grammatical function which is also $[+r]$; the combination $[-o, +r]$ results in specifying an oblique (see Table 10). The result is the mapping in (21). The second argument (arg_2) – which by default maps to OBJ – remains specified as $[-r]$ and can be promoted to SUBJ (see Table 10).

(21)		Agent	Theme	
say (promotional passive)	<	arg ₁	arg ₂	>
		[-o]	[-r]	
		[+r]		
		OBL _θ	SUBJ	

Mapping Theory also allows for non-promotional transitive passives, which have the mapping in (22) (Kibort & Maling, 2015, 156). In a non-promotional passive, the first argument is still demoted to an OBL_θ, as in the promotional passive in (21). The point of difference is that in the non-promotional passive, the second argument is additionally specified as [+o] and is thus ‘blocked’ from being promoted to SUBJ; it is forced to remain an OBJ (see again Table 10).

(22)		Agent	Theme	
say (non-promotional passive)	<	arg ₁	arg ₂	>
		[-o]	[-r]	
		[+r]	[+o]	
		OBL _θ	OBJ	

For constructions like (17), the two analyses thus differ in terms of the grammatical function which is assigned to the clausal argument: in the promotional passive analysis, the clausal argument maps to subject; in the non-promotional analysis, it maps to object.

3.3.2 Comparable analyses for Germanic

The structural ambiguity of such constructions has been previously acknowledged for other Germanic languages. Specifically, the possibility that they are open to either a promotional passive or non-promotional passive analysis is captured by Berman (2003, 162–164) for parallel constructions in German and by Bennis (1986) for Dutch. Berman (2003, 162–164) presents an LFG analysis of parallel constructions in German with the predicate *gesagt*, the passive of *sagen* ‘say’, which also takes two arguments – an Agent and a Theme – which by default map to subject and object respectively. Berman argues that, in such constructions, the Theme straightforwardly maps to subject if it is a DP. If the Theme is clausal – as in the Icelandic construction in (17) – then it can map to either subject or object (Berman, 2003, 162). When the clause maps to subject, the construction qualifies as a promotional passive; when the clause maps to object, it qualifies as a non-promotional passive on the terms outlined above.

In a different framework, Bennis (1986, 108) points out that in Dutch constructions with passive transitives which take a clausal complement, both *het* (*it*-type expletive) and *er* (*there*-type expletive) are possible, see (23) (examples are taken from the discussion of Bennis, 1986, 108 in Vikner, 1995, 229).¹¹

¹¹Vikner (1995, 233–234, 242–246) discusses constructions with passive verbs which take a clausal complement in Danish, noting that both *det* (*it*-type expletive) and *der* (*there*-type expletive) are possible:

- (i) **Det** blev sagt [at du ville komme].
it was said that you would come
(Danish, Vikner, 1995, 244)
- (ii) **Der** blev sagt [at du ville komme].
there was said that you would come
(Danish, Vikner, 1995, 244)

- (23) a. **Het** wordt gezegd [dat Jan ziek is].
 it is said that Jan ill is
 (Dutch)
- b. **Er** wordt gezegd [dat Jan ziek is].
 there is said that Jan ill is
 (Dutch)

In examples with *het* like (23-a), Bennis assumes that *het* is base-generated in the object position (i.e. within the VP) and is moved into the subject position. The embedded clause is assumed to be in an adjoined position, cf. (24-a) (as presented in Vikner, 1995, 229). The sentence is thus analysed as a promotional passive, which in this particular framework involves movement of the constituent base-generated in object position (*het*) into subject position. In examples with *er* like (23-b), Bennis assumes that the embedded clause occupies the object position and that *er* is base-generated in subject position (SpecIP), cf. (24-b) (Vikner, 1995, 229). There is no movement of a base-generated object into subject position, and so the sentence essentially qualifies as a non-promotional passive.

- (24) a. [_{CP} Het_i wordt [_{IP} t_i [_{VP} t_i gezegd [_{CP} dat Jan ziek is]]]]
 b. [_{CP} Er_i wordt [_{IP} t_i [_{VP} gezegd [_{CP} dat Jan ziek is]]]]

This difference is in turn related to the claim that *het* has argument status (i.e. is a non-referential argument, a so-called ‘quasi-argument’, Chomsky, 1981, 324–327), and can thus be base-generated in a theta-marked position (e.g. the object position of a passive). *Er* does not qualify as an argument (i.e. is a non-referential non-argument, a ‘true expletive’), and so it can be base-generated in the subject position of a passive, which is a non-theta-marked position. What is relevant to my analysis is the fact that the passive of a say-type predicate with a clausal argument is in this view open to both a promotional and non-promotional passive analysis. Moreover, since Dutch has both an *it*-type expletive and *there*-type expletive, this difference shows up on the clause-initial element (*het* versus *er*). Since Icelandic has only the *it*-type (*það*) and not the *there*-type (*þar*), the two possibilities do not show up in the same way. The distinction can however be recovered by the positional distribution of *það*, as I now show.

3.3.3 Findings for Icelandic

So far, I have claimed that constructions like (17) are structurally ambiguous and allow for the clausal argument to be analysed as a subject (promotional passive) or an object (non-promotional passive). In Section 3.1, I showed that cataphoric *það* in constructions with an unambiguous clausal subject frequently occurs in both pre- and postfinite position, and thus qualifies as a subject – at least in earlier stages. By contrast, in Section 3.2 I showed that, in constructions with an unambiguous clausal object, cataphoric *það* is virtually restricted to the prefinite position and thus behaves as a placeholder for the topic position.

With these findings in mind, I now examine the positional distribution of cataphoric *það* in the ambiguous construction type in (17), which is attested in IcePaHC, albeit relatively infrequently. The IcePaHC findings support the availability of the two analyses I have proposed. The results indicate that cataphoric *það* in this context occupies an intermediate position between the findings for constructions with an unambiguous clausal subject and the findings for constructions with an unambiguous clausal object with respect to its frequency in postfinite position. Overall, cataphoric *það* in the ambiguous type occurs in the postfinite position not as frequently as in the unambiguous clausal subject type (Table 5) but more frequently than in the unambiguous clausal object type *það* (Table 9).

I interpret this intermediate result as supporting my proposal that, in these ambiguous contexts, the clausal argument can be analysed as a subject (in the promotional passive analysis) – in

Time period	Postfinite <i>það</i>	No <i>það</i> (V2)	Total	% Postfinite <i>það</i>
1150-1350	9	16	25	36.0%
1351-1550	2	2	4	50.0%
1551-1750	3	4	7	42.9%
1751-1900	1	3	4	25.0%
1901-2008	0	6	6	0.0%
All periods	15	31	46	

Table 11: Frequency of postfinite *það* in structurally ambiguous constructions with a clausal argument and topicalization in IcePaHC, 1150-2008.

which case *það* is preferred in postfinite position – or as an object (in the non-promotional passive analysis) – in which case it is dispreferred in postfinite position, in line with the general positional restriction observed above for cataphoric *það* in constructions with an unambiguous clausal object.

The corpus findings also reveal that cataphoric *það* occurs very frequently in the prefinite position in the ambiguous construction type, see Table 12.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1350	33	5	38	86.8%
1351-1550	5	4	9	55.6%
1551-1750	6	3	9	66.7%
1751-1900	5	0	5	100.0%
1901-2008	1	0	1	100.0%
All periods	50	12	62	

Table 12: Frequency of prefinite *það* in structurally ambiguous constructions with a clausal argument in IcePaHC, 1150-2008.

This finding is in line with the proposal made in this section. Regardless of whether the clausal argument is analysed as a subject or an object in such contexts, *það* is expected to be frequently present in the prefinite position on the basis of the results for unambiguous constructions with a clausal subject and those with a clausal object (Sections 3.1 and 3.2 respectively).

4 THE EMERGENCE OF THE EXPLETIVE

4.1 Say-type predicates as a bridging context

So far, I have shown that cataphoric *það* is already established in Old Icelandic as a topic position placeholder in topicless impersonal constructions with a clausal object, as in the examples repeated in (25).

- (25) a. En **það**_i er vitanda, [að þá göfgum vér réttlega
but CATPH be.PRS know.PRS.PTCP COMP then honour.PRS we.NOM rightly
postula guðs alla]_i...
apostles god.GEN all...
‘But one should know that we then rightly honour all of God’s apostles...’
(1150, Homiliubok.304)

I leave a full comparison of the Danish data with the Icelandic data for future work.

- b. **Það_i** er af Helga að segja [að hann fékk fæð mikla]_i.
 CATPH be.PRS of Helgi to say.INF COMP he.NOM get.PST melancholy great
 ‘One can say of Helgi that he got a great melancholy.’ (1400, Gunnar.876)

As mentioned in Section 1, in Old Icelandic topicless impersonal constructions which lack a clausal object, *það* is by contrast overwhelmingly absent, rendering V1 sentences, e.g. (26); see also (3) above.

- (26) a. ...og má eigi einum munni allt senn segja.
 and may NEG one.DAT mouth.DAT all at once say.INF...
 ‘...and one cannot say everything at once with one mouth.’
 (1260, Jomsvikingar.875)
- b. Var síðan leitað vandlega.
 be.PST then search.PASS.PTCP carefully
 ‘It was then searched carefully.’ (1210, Jarstein.537)

The IcePaHC results confirm that impersonal constructions which lack a clausal object in the earlier periods are predominantly V1 constructions which lack *það*, see Table 13. There is a dramatic increase in prefinite *það* which – at least in the corpus data – is visible in the period 1901-2008.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1350	2	80	82	2.4%
1351-1550	1	47	48	2.1%
1551-1750	0	24	24	0.0%
1751-1900	3	32	35	8.6%
1901-2008	24	7	31	77.4%
All periods	30	190	220	

Table 13: Frequency of prefinite *það* in impersonal constructions without a clausal object in IcePaHC, 1150-2008.

Nevertheless, there are three early examples in the data prior to 1550 with a prefinite *það* which can be analysed as an expletive. I show these in (27).

- (27) a. **Það** er mælt um sakir þær allar sem hér eru
 EXPL be.PRS speak.PASS.PTCP about things DEM all REL here be.PRS
 taldar, um frumhlaup og um sár og um víg
 tell.PASS.PTCP about personal-assault and about wound and about manslaughter
 og lagalöstu alla...
 and law-evasions all
 ‘There is talk about all those things which are told here, about personal assault and about injury and about manslaughter and all evasions of the law...’
 (1270, Gragas.334)¹²
- b. **Það** er nú að segja frá Hrafn. Hann kom á Eyri og...
 EXPL be.PRS now to say.INF from Hrafn he.NOM come.PST to Eyri and
 ‘There is now talk of Hrafn. He came to Eyri and...’
 (1350, Finnbogi.1394)¹³

¹²This sentence introduces a new chapter, and so one can rule out the possibility that the clause-initial *það* has anaphoric reference to something in the preceding context.

¹³This example initiates a new discourse and so one can again rule out the possibility that *það* is anaphorically referential.

- c. **Það** er nú sagt af einum ríkum manni og mikilhæfum...
 EXPL be.PRS now say.PASS.PTCP of one rich man and talented
 Hann tók sótt hættliga.
 he.NOM take.PST sickness dangerous
 ‘There is now talk of one rich and talented man... He took dangerously ill.’
 (1475, *Ævintyri.477*)¹⁴

All three of these exceptional examples with *það* have a say-type predicate (*mæla* ‘speak’, *segja* ‘say’). Say-type predicates are precisely those transitive predicates which participate in constructions with a clausal argument, i.e. the contexts where cataphoric *það* is already robustly attested in Old Icelandic, see (25) above. The examples in (27) – while scarce – are evidence that *það* can already appear in early Icelandic in constructions with say-type predicates which lack a clausal argument. Furthermore, there is a possibility that *það* in these examples is cataphoric – with reference to the proposition in the later discourse – though it is hard to be sure whether *það* is expletive or cataphoric here.

It is only in the last two periods – and specifically in texts as of c.1850 – that an unambiguously expletive *það* is attested in topicless impersonal constructions with other types of predicate, beyond the say-type. I show examples from 1850-2008 in (28), which exhibit expletive *það* occurring in impersonal constructions with a range of predicate types.

- (28) a. **Það** þurfti ekki að vitja um Hans...
 EXPL need NEG to check about Hans
 ‘It was not necessary to check up on Hans...’ (1883, *Voggur.81*)
- b. ...**það** var barið; Geimundur og Snjólaug komu jafnsnemma
 EXPL be.PST knock.PASS.PTCP Geimundur and Snjólaug come.PST same-time
 til dyranna...
 to doors.DEF
 ‘...someone knocked; Geimundur and Snjólaug came to the door at the same time...’
 (1902, *Fossar.1623*)
- c. **Það** má reyna að telja einhverjum öðrum en mér trú um það.
 EXPL may try.INF to convince.INF someone other than I.DAT truth about DEM
 ‘One may try to convince someone other than me of that.’
 (1908, *Ofurefli.1638*)
- d. **Það** á að spara með þessu.
 EXPL ought to save.INF with DEM
 ‘One ought to be sparing with this.’ (1985, *Margsaga.232*)
- e. Og **það** þarf að draga skipin úr sjó á haustum...
 and EXPL need to pull.INF ships.DEF out sea in autumns
 ‘And one needs to pull the ships in out of the sea in autumn...’ (2008, *Ofsi.390*)
- f. ...og **það** á aldrei að hafa hrædda menn í forystu fyrir neinu.
 and EXPL ought never to have.INF afraid men in leadership for nothing
 ‘And one ought never to have men who are afraid in the lead for anything.’
 (2008, *Ofsi.732*)

I suggest that the early examples of *það* in (27) are indication that constructions with say-type predicates and no clausal argument serve as a bridging context (e.g. Heine, 2002) in facilitating the spread of *það* to impersonal constructions with all types of predicate.¹⁵ Moreover, since the

¹⁴This sentence initiates a new chapter and so cannot be anaphoric. The example is also one of those cited by Rögnvaldsson (2002, 22).

¹⁵An anonymous reviewer suggests that the early examples with *það* in (27) are too few to support the claim that say-type predicates have special status in the diachronic development. I acknowledge this point, but maintain that the fact that *það* is attested several centuries earlier in this context than with other predicate types is still

status of *það* in (27) as either cataphoric or expletive is ambiguous, such examples appear to provide a bridging context from the unambiguous cataphoric to the unambiguous expletive type.

The examples in (28) thus represent the third and final stage in the diachronic development by which *það* becomes established in the clause-initial prefinite position in the full range of topicless impersonal constructions, see Table 14.

	say-type pred with clausal argument	say-type pred without clausal argument	any pred without clausal argument
Stage 1	+	–	–
Stage 2	+	+	–
Stage 3	+	+	+

Table 14: The rise of *það* in Icelandic topicless impersonal constructions.

Between stages 2 and 3, the spread of *það* is likely to have proceeded via lexical diffusion, but the IcePaHC data lacks the necessary detail to show the actual stepwise progression. More research of texts beyond IcePaHC would be necessary to show this development in greater detail.

4.2 A parallel development in presentational constructions

I now discuss how the emergence of the expletive in topicless impersonal constructions as shown in Section 4.1 has a parallel diachronic development in presentational constructions. Presentational constructions – which are ‘all new’ and hence topicless – are another context where *það* functions as a topic position placeholder in present-day Icelandic, e.g. (29).

- (29) a. **Það** voru mýs í baðkerinu í gær.
EXPL be.PST mice.NOM in bathtub.DEF yesterday
‘There were mice in the bathtub yesterday.’
b. Í gær voru (***það**) mýs í baðkerinu.
yesterday be.PST (*EXPL) mice.NOM in bathtub.DEF
‘Yesterday there were mice in the bathtub.’

In Old Icelandic, the expletive is absent in such constructions, rendering V1 structures, e.g. (30).

- (30) Voru þar tvö skip í búnaði.
be.PST there two.NOM ships.NOM in preparations
‘There were two ships in the preparations.’ (1250, Sturlunga.408.710)

Booth (2018) conducted a corpus-based study of IcePaHC examining the status of the expletive in presentational constructions. I display the findings for prefinite *það* in such contexts in Table 15.

a striking finding which fits with the account. It would be enlightening to look at data beyond IcePaHC on this point in future.

Time period	Prefinite <i>það</i>	No <i>það</i> (V1)	Total	% Prefinite <i>það</i>
1150-1350	0	39	39	0.0%
1351-1550	5	33	38	13.2%
1551-1750	4	21	25	16.0%
1751-1900	35	19	54	64.8%
1901-2008	86	5	91	94.5%
All periods	130	117	247	

Table 15: Frequency of prefinite *það* in presentational constructions in IcePaHC, 1150-2008.

The first examples with a prefinite *það* occur in the period 1351-1550. However, the frequency of *það* remains low from this point onwards, until the last two periods where a stark increase in *það* is visible in the corpus data for 1751-2008. I show examples from each period in (31).

- (31) a. **Það** var einn kvinna er fastaði við brauð og vatn
EXPL be.PST.SG one.NOM.SG woman.NOM.SG REL fast.PST with bread and water
fyrir Marju messu Magdalena.
for María.GEN mass Magdalena
‘There was one woman who fasted with bread and water for Mary Magdalene’s
mass.’ (1475, Ævintýri.17)
- b. **Það** eru margir kimar niður við sjómálið...
EXPL be.PRS.PL many.NOM.PL bilges.NOM.PL down by high-waterline.DEF
‘There are many bilges down by the high waterline.’ (1661, Indiafari.73.1287)
- c. **Það** rísu upp tveir nýir kaupmenn.
EXPL stand.PST.PL up two.NOM new.NOM.PL merchant.NOM.PL
‘There stood up two new merchants.’ (1888, Grimur.126)
- d. **Það** komu nokkrir vopnaðir menn af næstu
EXPL come.PST.PL some.NOM.PL armed.NOM.PL man.NOM.PL from next
bæjum...
farms
‘There came some armed men from the nearby farms...’ (2008, Ofsi.634)

It is striking that the increase in prefinite *það* in presentational constructions approximately coincides with the development shown for impersonal constructions in Section 4.1, whereby *það* generalises in its roll as a topic position placeholder to all types of topicless impersonal construction. I interpret these two developments as representing one overall change – the establishment of *það* as a topic position placeholder across presentational and impersonal constructions. As shown above, in earlier stages of Icelandic the topic position was typically unoccupied in such contexts, rendering V1 structures. Strikingly, the growing establishment of this topic position placeholder coincides with the change involving cataphoric ‘subject’ *það* shown in Section 3.1, which also develops towards becoming a topic position placeholder at this point in the diachrony.

5 CONCLUSION

In this paper, I have presented a diachronic account for the development whereby the Icelandic expletive *það* emerged as a topic position placeholder in topicless impersonal constructions, on the model of an earlier restricted set of cataphoric contexts. I showed that there are two types of cataphoric *það* which are robustly established in earlier Icelandic (pre-1750):

1. A cataphoric *það* which co-occurs with a clausal subject and behaves positionally like a subject (Section 3.1).

2. A cataphoric *það* which co-occurs with a clausal object and behaves positionally like a structural placeholder for the topic position in topicless sentences (Section 3.2).

I also discussed constructions with passive transitive say-type predicates where cataphoric *það* co-occurs with a clausal argument. I claimed that such constructions allow for two analyses (promotional or non-promotional passive) and showed how this difference can be modelled with a version of LFG’s Mapping Theory developed by Kibort (2007, 2008, 2014). Furthermore, I argued that the mixed positional distribution of *það* in such contexts supports the availability of the two analyses.

I showed that the overwhelming pattern in Icelandic prior to 1850 for impersonal constructions which lack a clausal argument is for *það* to be absent. However, sparse early examples with *það* were observed and found to share a particular property: having a say-type predicate (Section 4.1). Only in the latter half of the 19th century do the first examples of *það* with other predicate types appear, with *það* attested with a broad range of predicates in the data for 1850-2008. This led me to propose a pathway of change via generalisation, whereby *það* spread from a more restricted set of topicless impersonal constructions with a clausal object (cataphoric *það*) to all types of topicless impersonal constructions – including those which lack a clausal object and in which *það* cannot be cataphorically referential (expletive *það*). I argued that this change was facilitated by impersonal constructions with a say-type predicate, which act as a bridging context between the older context with a clausal object, and the innovative context with no clausal object.

Strikingly, three changes shown in this paper occur at approximately the same point in the Icelandic diachrony: the generalisation of *það* in its function as a topic position placeholder to all types of impersonal construction roughly coincides with the change whereby ‘subject’ cataphoric *það* begins to lose its subject status and transitions towards becoming a topic position placeholder, i.e. the same function as *það* in topicless impersonal constructions. Furthermore, the change whereby expletive *það* emerges in Icelandic presentational constructions – which are always topicless – also approximately coincides with the two changes involving cataphoric ‘subject’ *það* and topicless impersonal constructions. I interpret these three developments as representing one overall change – the establishment of *það* as a topic position placeholder in topicless sentences – occurring across three broad construction types (constructions with a clausal argument, impersonal constructions; presentational constructions). Furthermore, an IcePaHC-based study of other word order phenomena (V1 declaratives, subject position) by Booth et al. (2017) observed significant changes towards reduced word order freedom occurring in Icelandic in c.1900, which was interpreted as a shift towards increased syntactic configurationality and the growth of left peripheral structure. The changes involving cataphoric and expletive *það* presented in this paper similarly point towards this broader development in Icelandic clause structure at this relatively late stage in the diachrony. The exact nature of the interaction between these various changes, I leave for future research.

ACKNOWLEDGMENTS

The research for this paper was funded by an Arts & Humanities Research Council UK (AHRC) doctoral scholarship. Particular thanks go to the editors of this issue and two anonymous reviewers for their feedback on this work. I thank audiences at the *Manchester Forum in Linguistics* (April 2018), *Grammatik i Norden* (May 2018), the *25th South of England LFG Meeting* (May 2018) and the *Workshop on Word Order in Scandinavian Languages* at the University of Konstanz (December 2018) for comments on this work. For their valuable insights I am especially grateful to Kersti Börjars, Tine Breban, John Payne, Helge Lødrup, Joan Maling, Halldór Sigurðsson and Christin Schätzle.

References

- Andrews, Avery. 1990. Unification and morphological blocking. *Natural Language & Linguistic Theory* 8(4), 507–557.
- Asudeh, Ash & Ida Toivonen. 2009. Lexical-functional grammar. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 425–458. Oxford: Oxford University Press.
- Axel, Katrin. 2007. *Studies on Old High German Syntax: Left Sentence Periphery, Verb Placement and Verb-Second*. Amsterdam: John Benjamins.
- Bennis, Hans. 1986. *Gaps and Dummies*. Dordrecht: Foris publications.
- Berman, Judith, Stefanie Dipper, Christian Fortmann & Jonas Kuhn. 1998. Argument clauses and correlative *es* in German – deriving discourse properties in a unification analysis. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'98 Conference, University of Queensland, Brisbane*, Stanford, CA: CSLI Publications.
- Berman, Judith. 2003. *Clausal Syntax of German*. Stanford, CA: CSLI Publications.
- Booth, Hannah. 2018. *Expletives and Clause Structure: Syntactic Change in Icelandic*. Doctoral dissertation, University of Manchester.
- Booth, Hannah, Schätzle, Christin, Börjars, Kersti, & Butt, Miriam. (2017). Dative subjects and the rise of positional licensing in Icelandic. In Butt, Miriam, & King, Tracy Holloway (Eds.): *Proceedings of the LFG'17 Conference, University of Konstanz*, 104–124. Stanford, CA: CSLI Publications.
- Breivik, Leiv Egil. 1990. *Existential There: A synchronic and diachronic study*. Oslo: Novus Press.
- Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1), 1–50.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-Functional Syntax*, 2nd Edition. Oxford: Blackwell.
- Butt, Miriam, Tina Bögel, Kristina Kotcheva, Christin Schätzle, Christian Rohrdantz, Dominik Sacha, Nicole Dehé & Daniel Keim. 2014. V1 in Icelandic: A multifactorial visualization of historical data. In *Proceedings of VisLR: Visualization as added value in the development, use and evaluation of Language Resources*, 33–40. Workshop at the 9th edition of the Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.
- Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'00 Conference, University of California, Berkeley*, 104–121. Stanford, CA: CSLI Publications.
- Eythórsson, Thórhallur. 2008. The New Passive in Icelandic really is a passive. In Thórhallur Eythórsson (ed.), *Grammatical Change and Linguistic Theory: The Rosendal papers*, 173–219. Amsterdam: John Benjamins.
- Eythórsson, Thórhallur & Sigríður Sæunn Sigurðardóttir. 2016. A brief history of Icelandic weather verbs: syntax, semantics and argument structure. *Working Papers in Scandinavian Syntax* 96, 91–125.

- Falk, Cecilia. 1993. *Non-referential Subjects in the History of Swedish*. Department of Scandinavian Languages, Lund University.
- Faarlund, Jan Terje. 1990. *Syntactic Change: Toward a Theory of Historical Syntax*. Berlin: de Gruyter.
- Haiman, John. 1974. *Targets and Syntactic Change*. The Hague: Mouton.
- Håkansson, David. 2017. Transitive expletive constructions in Swedish. *Nordic Journal of Linguistics* 40(3), 255–285.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 83–101. Amsterdam: John Benjamins.
- Hróarsdóttir, Þorbjörg. 1998. *Setningafræðilegar breytingar á 19. öld: þróun þriggja málbreytinga*. [Syntactic changes in the 19th century: The development of three linguistic changes]. Reykjavík: Institute of Linguistics, University of Iceland.
- Ingham, Richard. 2001. The structure and function of expletive *there* in pre-modern English. *Reading Working Papers in Linguistics* 5, 231–249.
- Jónsson, Jóhannes Gísli. 2009. The new impersonal as a true passive. In Artemis Alexiadou, Jorge Hankamer, Thomas McFadden, Justin Nuger & Florian Schäfer (eds.), *Advances in Comparative Germanic Syntax*, 281–306. Amsterdam: John Benjamins.
- Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'07 Conference, Stanford University*, 250–270. Stanford, CA: CSLI Publications.
- Kibort, Anna. 2008. On the syntax of ditransitive constructions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'08 Conference, University of Sydney*, 312–332. Stanford, CA: CSLI Publications.
- Kibort, Anna. 2014. Mapping out a construction inventory with LFG's (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'14 Conference, Ann Arbor, Michigan*, 262–282. Stanford, CA: CSLI Publications.
- Kibort, Anna & Joan Maling. 2015. Modelling the syntactic ambiguity of the active vs. passive impersonal in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'15 Conference, Tokyo, Japan*, 145–165. Stanford, CA: CSLI Publications.
- Kinn, Kari. 2016. Referential vs. non-referential null subjects in Middle Norwegian. *Nordic Journal of Linguistics* 39(3), 277–310.
- Kroch, Anthony & Ann Taylor. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. Second edition. <http://www.ling.upenn.edu/hist-corpora/>.
- Kuhn, Jonas. 1999. The syntax and semantics of split NPs in LFG. In Francis Corblin, Carmen Dobrovie-Sorin & Jean-Marie Marandin (eds.), *Empirical Issues in Formal Syntax and Semantics 2, Selected Papers from the Colloque de Syntaxe et Sémantique à Paris (CSSP 1997)*, vol. 2, 145–166. The Hague: Thesus.
- Kuhn, Jonas. 2001. Resource sensitivity in the syntax-semantics interface and the German split NP construction. In Detmar Meurers, W. & Tibor Kiss (eds.), *Constraint-Based Approaches to Germanic Syntax*, 177–216. Stanford, CA: CSLI Publications.

- Lenerz, Jürgen. 1985. Zur Theorie syntaktischen Wandels: das expletive *es* in der Geschichte des Deutschen. In Werner Abraham (ed.), *Erklärende Syntax des Deutschen*, 99–136. Tübingen: Gunter Narr.
- Lødrup, Helge. 2011. Lexical-Functional Grammar: functional structure. In Robert D. Borsley & Kersti Börjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, 141–180. Oxford: Blackwell.
- Maling, Joan & Sigríður Sigurjónsdóttir. 2002. The new impersonal construction in Icelandic. *The Journal of Comparative Germanic Linguistics* 5(1), 97–142.
- Platzack, Christer. 1983. Existential sentences in English, German, Icelandic and Swedish. In Fred Karlsson (ed.), *Papers from the 7th Scandinavian Conference of Linguistics*, 80–100. Department of General Linguistics, University of Helsinki.
- Platzack, Christer. 1985. Narrative inversion in Old Icelandic. *Íslenskt mál* 7, 127–144.
- Randall, Beth. 2005. CorpusSearch2 User's Guide. Philadelphia: Dept. of Linguistics, University of Pennsylvania. Philadelphia: Dept. of Linguistics, University of Pennsylvania. <http://corpus-search.sourceforge.net/CS-manual/Contents.html>.
- Richards, Marc & Theresa Biberauer. 2005. Explaining Expl. In Marcel den Dikken & Christina Tortora (eds.), *The Function of Function Words and Functional Categories*, 115–153. Amsterdam: John Benjamins.
- Rögnvaldsson, Eiríkur & Höskuldur Thráinsson. 1990. On Icelandic word order once more. In Joan Maling & Annie Zaenen (eds.), *Syntax and Semantics: Modern Icelandic Syntax*, 3–40. San Diego, CA: Academic Press.
- Rögnvaldsson, Eiríkur. 2002. ÞAÐ í fornu máli – og síðar. [ÞAÐ (it, there) in Old Icelandic – and later]. *Íslenskt mál* 24, 7–30.
- Rögnvaldsson, Eiríkur, Einar Freyr Sigurðsson, Anton Karl Ingason & Joel Wallenberg. 2012. The Icelandic Parsed Historical Corpus (IcePaHC). In *Proceedings of LREC 2012*, 1978–1984.
- Santorini, Beatrice. 2010. Annotation manual for the Penn Historical Corpora and the PCEEC. Department of Linguistics, University of Pennsylvania. <http://www.ling.upenn.edu/hist-corpora/annotation/index.html>.
- Sells, Peter. 2005. The peripherality of the Icelandic expletive. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'05 Conference*, 408–428. Stanford, CA: CSLI Publications.
- Sigurðsson, Halldór Ármann. 1990. V1 declaratives and verb raising in Icelandic. In Joan Maling & Annie Zaenen (eds.), *Syntax and Semantics: Modern Icelandic Syntax*, 41–69. San Diego, CA: Academic Press.
- Sigurðsson, Halldór Ármann. 2007. Argument features, clausal structure and the computation. In Tammoy Bhattachayra, Eric Reuland & Giorgos Spathas (eds.), *Argument Structure*, 121–158. Amsterdam: John Benjamins.
- Sigurðsson, Halldór Ármann & Verner Egerland. 2009. Impersonal null subjects in Icelandic and elsewhere. *Studia Linguistica* 63(1), 158–185.
- Silva-Villar, Luis. 1996. The diachronic syntax of expletive creation. *International Journal of Basque Linguistics and Philology* 30(1), 173–193.

- Thráinsson, Höskuldur. 1979. *On Complementation in Icelandic*. New York: Garland.
- Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
- Vikner, Sten. 1995. *Verb Movement and Expletives in the Germanic Languages*. Oxford: Oxford University Press.
- Wallenberg, Joel C., Anton Karl Ingason, Einar Freyr Sigurðsson & Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC), version 0.9. http://linguist.is/icelandic_treebank.
- Zaenen, Annie. 1983. On syntactic binding. *Linguistic Inquiry* 14(3), 469–504.