



**UNIVERSITY  
OF ICELAND**

## **Clausal nominalization in Icelandic**

Mirko Garofalo

Dissertation towards the degree of Doctor of Philosophy  
2025

**SCHOOL OF HUMANITIES  
FACULTY OF ICELANDIC AND COMPARATIVE CULTURAL STUDIES**

Íslensku- og menningardeild Háskóla Íslands  
hefur metið ritgerð þessa hæfa til varnar  
við doktorspróf í íslenskri málfræði

Reykjavík, 5. júní 2025

Gauti Kristmannsson  
Deildarforseti

The Faculty of Icelandic and Comparative Cultural Studies  
at the University of Iceland  
has declared this dissertation eligible for a defence  
leading to a Ph.D. degree in Icelandic linguistics

Doctoral committee:  
Jóhannes Gísli Jónsson, supervisor  
Anton Karl Ingason  
Matthew Whelpton

*Clausal nominalization in Icelandic*

© Mirko Garofalo  
Reykjavík 2025

Dissertation for a doctoral degree at the University of Iceland. All rights reserved.  
No part of this publication may be reproduced in any form  
without written permission of the author

ISBN: 978-9935-9811-9-6  
ORCID: 0000-0001-8304-3003

# Abstract

This doctoral dissertation is the first extensive analysis of nominalized clauses in contemporary Icelandic. Its main objective is to explain: a) the role of the demonstrative pronoun *það* ‘that’ (see Garofalo (2020)) when it introduces a clausal complement; b) its syntactic distribution; and c) the structural difference between nominalized and non-nominalized clauses. The results of this analysis are mainly built upon naturalistic examples from the Icelandic Gigaword Corpus (Steingrímsson (2019)), and judgment tasks, which consist of 20 questionnaires with 25 sentences each, as well as interviews.

This dissertation claims that the main role of *það* is to check the case, gender and number features that target the relevant clausal complement and which can only be checked by DPs. In the absence of such features, *það* is unnecessary and is therefore dropped. However, *það* is not dropped if a head that subcategorizes for a clausal complement bears a feature that triggers Merge of a DP (see Heck and Müller (2007)), or if a clause is moved from its merge position to check a D-feature, e.g. to Spec,T. The hypothesis presented here differs from previous accounts on clausal nominalization in Icelandic, in particular Thráinsson (1979), who claimed, following Kiparsky and Kiparsky (1971), that factivity is the main trigger of clausal nominalization in Icelandic: factive predicates would tend to select more complex structures for their complements (i.e., nominalized clauses) than non-factives. The results presented in this thesis show that factivity is inadequate to clearly explain the distribution of clausal nominalization across syntactic positions.

Various facts support the main hypothesis of this dissertation. Non-nominalized clauses are preferable in positions where nominative and accusative case are assigned, except in e.g. Spec,T and Spec,Appl, where nominalization is mandatory in absence of extraposition. On the other hand, dative and genitive are highly likely to trigger clausal nominalization. A similar contrast can also be observed with post-copular clauses, where the determiner, which agrees with the gender and number of the subject, is dispreferred if it has to surface in neuter singular, but it is more likely to emerge if the gender and/or the number features of the pronoun are non-default. These results indicate that default features related to case, gender and number, such as nominative, singular or neuter, are not formal features to check in narrow syntax, which cause *það* to be dropped. Such a conclusion aligns with theories which, for instance, consider nominative and accusative configurationally derived, like *Dependent Case Theory* (see e.g. Yip et al. (1987); Marantz (2000); Preminger (2011)).

From a structural perspective, this dissertation also proposes that nominalized clauses are DPs and non-nominalized clauses emerge as CPs at the surface. However, all clauses that are merged in DP positions are originally merged as DPs and only undergo a process of structural removal if they land in a position in which no feature needs to be checked by *það*, yielding a

non-nominalized clause. These claims are supported by the fact that: a) an item can only be extracted from a clausal argument if it is not nominalized; b) nominalized clauses as well as DPs are ungrammatical when they replace the clausal complement of verbs like *þvinga* 'force' or *hjálpa* 'help'. As for structural removal, it is supported by the fact that obligatory nominalization caused by lexical case assignment can be altered by extraction, which causes *það* to become ungrammatical while the extracted DP item has its case overwritten with lexical case. This suggests that the DP shell is not longer present to check case.

By comparing clausal nominalization in Icelandic and other languages (Swedish, German, Persian and Russian), this dissertation also shows that, independently of whether a language displays a morphological case system, structural case does not trigger clausal nominalization. Lexical case, on the other hand, triggers clausal nominalization specifically in those languages that have a morphological case system. Moreover, languages without a morphological case system tend to generalize how nominalized clauses are distributed: nominalization can become optional across all syntactic positions (as in Swedish) or it can become mandatory *in situ* as a general rule (this is the case of Persian). However, all the languages studied in this comparison show that clausal subjects must be nominalized in Spec,T unless they can escape that position. This suggests that D-feature checking is a common trigger of clausal nominalization cross-linguistically.

# Ágrip

Þessi doktorsritgerð er fyrsta ítarlega greiningin á nafnyrtum fallsetningum í nútímaíslensku. Aðalmarkmið hennar er að útskýra a) hlutverk ábendingarfornafnsins *það* þegar það tekur með sér fallsetningu (sjá Garofalo (2020)), b) setningarlega dreifingu setningafornafnsins og c) formgerðarmuninn á nafnyrtum og ónafnyrtum fallsetningum. Niðurstöður þessarar greiningar eru aðallega byggðar á náttúrulegum gögnum úr Risamálheildinni (sjá Steinþór Steingrímsson (2019)) og dómaprófum, þ.e. 20 könnunum með 25 setningum, en einnig á viðtölum við málhafa.

Í þessari ritgerð eru færð rök fyrir því að aðalhlutverk setningafornafnsins *það* sé að gáta fall-, tölu- og kynþætti sem fallsetning þarf en einungis ákveðniliðir geta gátað. Þegar þessir þættir eru ekki til staðar er *það* ónauðsynlegt og er því sleppt. Hins vegar birtist það ef haus sem tekur fallsetningu sem fyllilið ber þátt sem kallar á grunnmyndun ákveðniliðar (sjá Heck og Müller (2007)) eða ef fallsetningin er færð úr grunnstöðu til þess að gáta D-þátt, t.d. í frumlagssæti setningar (Spec,T). Tilgátan sem hér er kynnt er ólík fyrri greiningum á nafnyrðingu fallsetninga í íslensku, sérstaklega greiningu Höskuldar Þráinssonar (1979), sem taldi með hliðsjón af Kiparsky og Kiparsky (1971) að staðreyndarmerking í sögnum væri aðalhvati nafnyrðingar í íslensku. Staðreyndarsagnir hefðu því tilhneigingu til að velja flóknari formgerðir fyrir fylliliði sína (þ.e. nafnyrtar fallsetningar) en álitssagnir. Niðurstöðurnar sem kynntar eru í þessari ritgerð sýna hins vegar að staðreyndarmerking getur ekki útskýrt dreifingu nafnyrtra fallsetninga þvert á setningastöðu.

Það sem styður megintilgátu ritgerðarinnar er að ónafnyrtar fallsetningar eru algengastar í mismunandi sætum þegar nefnifalli og þolfalli er úthlutað, nema t.d. í ákvæðisliðarsæti tíðarliðar (Spec,T) og ákvæðisliðarsæti þáguliðar (Spec,Appl) þar sem nafnyrðing er skyldubundin þegar fráfærsla fallsetningar verkar ekki. Þágufall og eignarfall kalla hins vegar yfirleitt á nafnyrðingu fallsetninga. Svipaðan mun má sjá í fallsetningum með sögninni *vera* þar sem setningafornafnið, sem samræmist frumlaginu eftir kyni og tölu, er sjaldan notað í hvorugkyni eintölu en mun oftast ef kyn- og/eða töluþættir fornafnsins eru ekki sjálfgefnir. Niðurstöður þessarar ritgerðar benda því til þess að sjálfgefnir fall-, kyn- og töluþættir eins og nefnifall, eintala eða hvorugkyn eru ekki formlegir þættir sem gáta þarf í setningafræði, sem veldur því að *það* sé fellt brott. Slík ályktun rímar við kenningar um að nefnifall og þolfall séu háð venslum á milli ákveðniliða eins og í kenningunni um tengifall (e. *dependent case*; sjá t.d. Yip et al. (1987), Marantz (2000) og Preminger (2011)).

Varðandi formgerð fallsetninga eru færð rök fyrir því að nafnyrtar fallsetningar séu ákveðniliðir (DPs) en ónafnyrtar fallsetningar séu tengiliðir (CPs) í yfirborðsgerð. Þó verki sneyðing (e. *structural removal*) á allar fallsetningar sem eru grunnmyndaðar í ákveðniliðarsæti ef þær lenda í stöðu þar sem ekki þarf að gáta neinn þátt og þær verði því ónafnyrtar. Þetta fær stuðning af því að a) færsla úr fallsetningu getur aðeins átt sér stað ef fallsetningin er ónafnyrt og b) nafnyrtar fallsetningar og ákveðniliðir eru almennt ótæk sem fylliliðir sagna eins og *þvinga* eða *hjálpa*.

Sneyðing styðst hins vegar við þau rök að færsla úr aukasetningu getur haft áhrif á skyldubundna nafnyrðingu þegar orðasafnsfalli er úthlutað. Í þessu tilviki verður það ótækt en fall ákveðniðar sem færður er út úr aukasetningunni er yfirskrifað með orðasafnsfallinu. Þetta bendir til þess að ákveðniðar fallsetningar sé ekki lengur til staðar til að gáta fallþætti.

Samanburður á íslensku og öðrum málum (sænsku, þýsku, rússnesku og persnesku) sýnir að formgerðarfall leiðir ekki til nafnyrðingar fallsetninga hvort sem tungumál er með ríkulegt fallakerfi eða ekki. Hins vegar kallar orðasafnsfall á nafnyrðingu í þeim tungumálum sem eru með ríkulegt fallakerfi. Tungumál með fátæklegt fallakerfi hafa hins vegar tilhneigingu til að afhæfa dreifingu nafnyrtra fallsetninga í tiltekinni setningastöðu en nafnyrðing getur verið valfrjáls í ákveðniðarstöðum (eins og í sænsku) eða skyldubundin í grunnstöðu (eins og í persnesku). En öll tungumál sem könnuð voru í þessum samanburði sýna að fallsetningar verða að vera nafnyrtar í frumlagssæti (Spec,T) nema þær færast út úr þeirri stöðu. Þetta bendir til þess að gátun á D-þætti sé algengur hvati fyrir nafnyrðingu þvert á tungumál.

# Contents

|                                                                                     |           |
|-------------------------------------------------------------------------------------|-----------|
| <b>List of Tables</b>                                                               | <b>13</b> |
| <b>List of Figures</b>                                                              | <b>15</b> |
| <b>Acknowledgements</b>                                                             | <b>17</b> |
| <b>1 Introduction</b>                                                               | <b>19</b> |
| 1.1 Topic and research questions . . . . .                                          | 19        |
| 1.2 Hypothesis . . . . .                                                            | 21        |
| 1.3 Methodology . . . . .                                                           | 23        |
| 1.4 Structure . . . . .                                                             | 24        |
| <b>2 Clausal nominalization</b>                                                     | <b>25</b> |
| 2.1 Introduction . . . . .                                                          | 25        |
| 2.2 A brief overview . . . . .                                                      | 26        |
| 2.2.1 Distribution of <i>pað</i> . . . . .                                          | 26        |
| 2.2.2 Demonstrative <i>pað</i> as CN . . . . .                                      | 29        |
| 2.2.3 A distal demonstrative as best candidate for clausal nominalization . . . . . | 31        |
| 2.2.4 Clausal extraposition . . . . .                                               | 34        |
| 2.3 The main puzzle of clausal nominalization . . . . .                             | 35        |
| 2.3.1 Subjects . . . . .                                                            | 35        |
| 2.3.2 Clausal objects . . . . .                                                     | 37        |
| 2.3.3 Complements of prepositions . . . . .                                         | 39        |
| 2.3.4 Complements of adjectives . . . . .                                           | 41        |
| 2.3.5 Clausal indirect objects . . . . .                                            | 42        |
| 2.4 Structure of clausal arguments with and without <i>pað</i> . . . . .            | 43        |
| 2.4.1 Introduction . . . . .                                                        | 43        |
| 2.4.2 Generalized DP/NP hypothesis . . . . .                                        | 44        |
| 2.4.3 Generalized CP hypothesis . . . . .                                           | 51        |
| 2.4.4 DP-shell vs. bare CP hypothesis . . . . .                                     | 53        |
| 2.5 Trigger and distribution of clausal nominalization . . . . .                    | 54        |
| 2.5.1 Introduction . . . . .                                                        | 54        |
| 2.5.2 Factivity hypothesis . . . . .                                                | 55        |
| 2.5.3 Other semantic influences of matrix verbs on clausal nominalization . . . . . | 58        |
| 2.5.4 DP selection hypothesis . . . . .                                             | 60        |
| 2.5.5 Case-Marking hypothesis . . . . .                                             | 64        |

|          |                                                                           |            |
|----------|---------------------------------------------------------------------------|------------|
| 2.5.6    | Less common hypotheses . . . . .                                          | 67         |
| 2.5.7    | The contribution of this dissertation: a preliminary hypothesis . . . . . | 69         |
| 2.6      | Summary . . . . .                                                         | 72         |
| <b>3</b> | <b>Empirical analysis</b>                                                 | <b>75</b>  |
| 3.1      | Introduction . . . . .                                                    | 75         |
| 3.2      | Methodology . . . . .                                                     | 76         |
| 3.2.1    | Clausal nominalization as a syntactic variable . . . . .                  | 76         |
| 3.2.2    | Two types of data . . . . .                                               | 78         |
| 3.2.3    | Corpora . . . . .                                                         | 79         |
| 3.2.4    | Judgment tasks . . . . .                                                  | 82         |
| 3.2.5    | Preliminary guide to the analysis of results . . . . .                    | 83         |
| 3.3      | Data . . . . .                                                            | 86         |
| 3.3.1    | Subjects . . . . .                                                        | 86         |
| 3.3.2    | Clausal direct objects . . . . .                                          | 92         |
| 3.3.3    | Post-copular clauses . . . . .                                            | 96         |
| 3.3.4    | Prepositional phrases . . . . .                                           | 98         |
| 3.3.5    | Complements of nouns . . . . .                                            | 113        |
| 3.3.6    | Complements of adjectives . . . . .                                       | 114        |
| 3.3.7    | Clausal indirect objects . . . . .                                        | 118        |
| 3.3.8    | Nominalization as a feature checking operation . . . . .                  | 122        |
| 3.3.9    | Clausal nominalization based on clause type . . . . .                     | 125        |
| 3.3.10   | Data on age and gender groups . . . . .                                   | 126        |
| 3.3.11   | Data on sources . . . . .                                                 | 126        |
| 3.4      | Summary . . . . .                                                         | 127        |
| <b>4</b> | <b>Theoretical implications</b>                                           | <b>129</b> |
| 4.1      | Introduction . . . . .                                                    | 129        |
| 4.2      | Core assumptions . . . . .                                                | 131        |
| 4.2.1    | Basic concepts of Minimalism . . . . .                                    | 131        |
| 4.2.2    | Main clausal structure . . . . .                                          | 134        |
| 4.2.3    | Case . . . . .                                                            | 137        |
| 4.3      | Deriving NomCIs and non-NomCIs . . . . .                                  | 140        |
| 4.3.1    | Introduction . . . . .                                                    | 140        |
| 4.3.2    | The problem of nominative subjects . . . . .                              | 141        |
| 4.3.3    | Option 1: DPs and CPs merged as such . . . . .                            | 142        |
| 4.3.4    | Option 2: All-CP merge hypothesis . . . . .                               | 148        |
| 4.3.5    | Option 3: All-DP merge hypothesis . . . . .                               | 150        |
| 4.3.6    | Structural removal . . . . .                                              | 153        |
| 4.3.7    | Four possible configurations . . . . .                                    | 157        |

|          |                                                                     |            |
|----------|---------------------------------------------------------------------|------------|
| 4.3.8    | Stylistic fronting . . . . .                                        | 161        |
| 4.3.9    | Postcopular clauses: mandatory extraposition? . . . . .             | 166        |
| 4.4      | Case, gender and number . . . . .                                   | 170        |
| 4.4.1    | Introduction . . . . .                                              | 170        |
| 4.4.2    | Structural vs. lexical case . . . . .                               | 170        |
| 4.4.3    | The trigger of case . . . . .                                       | 171        |
| 4.4.4    | Nominative and accusative are not syntactic case features . . . . . | 173        |
| 4.4.5    | The problem of prepositional complements . . . . .                  | 177        |
| 4.4.6    | Beyond case: gender and number . . . . .                            | 182        |
| 4.5      | Cross-linguistic analysis . . . . .                                 | 185        |
| 4.5.1    | Introduction . . . . .                                              | 185        |
| 4.5.2    | Swedish . . . . .                                                   | 186        |
| 4.5.3    | German . . . . .                                                    | 188        |
| 4.5.4    | Persian . . . . .                                                   | 194        |
| 4.5.5    | Russian . . . . .                                                   | 198        |
| 4.6      | Summary . . . . .                                                   | 201        |
| <b>5</b> | <b>Conclusions</b>                                                  | <b>205</b> |
| <b>A</b> | <b>Data from corpora</b>                                            | <b>209</b> |
| A.1      | Introduction . . . . .                                              | 209        |
| A.2      | Further information on methodology . . . . .                        | 209        |
| A.2.1    | Special queries for clausal subjects . . . . .                      | 209        |
| A.2.2    | Exclusions from the analysis . . . . .                              | 211        |
| A.2.3    | Values attributed to each token . . . . .                           | 212        |
| A.2.4    | Statistical summary . . . . .                                       | 214        |
| A.3      | Data . . . . .                                                      | 216        |
| A.3.1    | Subjects in first position . . . . .                                | 217        |
| A.3.2    | Subjects in Spec,T . . . . .                                        | 218        |
| A.3.3    | Clausal objects . . . . .                                           | 220        |
| A.3.4    | Post-copular clauses . . . . .                                      | 221        |
| A.3.5    | Prepositional phrases . . . . .                                     | 221        |
| A.3.6    | Complements of nouns . . . . .                                      | 224        |
| A.3.7    | Complements of adjectives . . . . .                                 | 225        |
| A.3.8    | Active, passive and middle voice . . . . .                          | 225        |
| A.3.9    | Data on sources based on the frequency of <i>pað</i> . . . . .      | 226        |
| <b>B</b> | <b>Data from questionnaires</b>                                     | <b>227</b> |
| B.1      | Introduction . . . . .                                              | 227        |
| B.2      | Structure . . . . .                                                 | 227        |

|                                                    |            |
|----------------------------------------------------|------------|
| B.3 Questionnaire template . . . . .               | 228        |
| B.4 Participants . . . . .                         | 230        |
| B.5 Spreadsheets . . . . .                         | 230        |
| B.6 Statistical tests . . . . .                    | 231        |
| B.7 Relevant tokens in questionnaires . . . . .    | 231        |
| B.8 Clausal indirect objects: Interviews . . . . . | 239        |
| B.9 Data on age and gender groups . . . . .        | 241        |
| <b>Bibliography</b>                                | <b>243</b> |

# List of Tables

|      |                                                                                        |     |
|------|----------------------------------------------------------------------------------------|-----|
| 2.1  | Inflection of <i>pað</i> . . . . .                                                     | 26  |
| 2.2  | Complete inflection of CN <i>pað</i> . . . . .                                         | 29  |
| 2.3  | Summary - Clausal nominalization puzzle . . . . .                                      | 43  |
| 3.1  | Example of results from the questionnaires . . . . .                                   | 83  |
| 3.2  | Frequency bands for competition between variants in Bermel and Knittl (2012) . . . . . | 85  |
| 3.3  | Revisitation of the frequency bands proposed in Bermel and Knittl (2012) . . . . .     | 85  |
| 3.4  | Example of results from corpora . . . . .                                              | 85  |
| 3.5  | Questionnaires - Nominative subjects (not extraposed) after the finite verb . . . . .  | 87  |
| 3.6  | RMH - Nominative subjects (not extraposed) after the finite verb . . . . .             | 87  |
| 3.7  | Questionnaires - Nominative subjects (not extraposed) in first position . . . . .      | 87  |
| 3.8  | RMH - Nominative subjects (not extraposed) in first position . . . . .                 | 88  |
| 3.9  | Passivized clausal objects in Spec,T - Tokens with extraposition . . . . .             | 89  |
| 3.10 | Questionnaires - Clausal subjects with lexical case . . . . .                          | 90  |
| 3.11 | RMH - Clausal subjects with lexical case . . . . .                                     | 91  |
| 3.12 | Questionnaires - Nominative and accusative clausal objects . . . . .                   | 93  |
| 3.13 | RMH - Nominative and accusative clausal objects . . . . .                              | 94  |
| 3.14 | Questionnaires - Dative and genitive clausal objects . . . . .                         | 95  |
| 3.15 | RMH - Dative and genitive clausal objects . . . . .                                    | 95  |
| 3.16 | Classes of subjects - Postcopular clauses . . . . .                                    | 96  |
| 3.17 | Questionnaires - Postcopular clauses . . . . .                                         | 97  |
| 3.18 | RMH - Postcopular clauses . . . . .                                                    | 97  |
| 3.19 | Questionnaires - Accusative prepositional complements . . . . .                        | 102 |
| 3.20 | RMH - Accusative prepositional complements . . . . .                                   | 102 |
| 3.21 | Questionnaires - Dative prepositional complements . . . . .                            | 102 |
| 3.22 | RMH - Dative prepositional complements . . . . .                                       | 103 |
| 3.23 | Questionnaires - Genitive prepositional complements . . . . .                          | 103 |
| 3.24 | RMH - Genitive prepositional complements . . . . .                                     | 104 |
| 3.25 | Questionnaires - Double case prepositional complements . . . . .                       | 106 |
| 3.26 | RMH - Preposition <i>yfir</i> and <i>í</i> with accusative and dative case . . . . .   | 107 |
| 3.27 | Questionnaires - Functional prepositions preceded by adverbs . . . . .                 | 108 |
| 3.28 | RMH - Functional prepositions preceded by adverbs . . . . .                            | 109 |
| 3.29 | Questionnaires - Lexical prepositions . . . . .                                        | 110 |
| 3.30 | Judgments on 8A-08 and 8B-18 . . . . .                                                 | 110 |
| 3.31 | RMH - Lexical prepositions . . . . .                                                   | 110 |

|      |                                                                                            |     |
|------|--------------------------------------------------------------------------------------------|-----|
| 3.32 | Questionnaires - Complements of nouns . . . . .                                            | 113 |
| 3.33 | RMH - Complements of nouns . . . . .                                                       | 113 |
| 3.34 | Questionnaires - Accusative adjectival complements . . . . .                               | 115 |
| 3.35 | Judgments on 8A-20 and 8B-06 . . . . .                                                     | 115 |
| 3.36 | Questionnaires - Genitive adjectival complements . . . . .                                 | 116 |
| 3.37 | RMH - Genitive adjectival complements . . . . .                                            | 116 |
| 3.38 | Questionnaires - Dative adjectival complements . . . . .                                   | 116 |
| 3.39 | RMH - Dative adjectival complements . . . . .                                              | 116 |
| 3.40 | Questionnaires - Clausal indirect objects . . . . .                                        | 119 |
| 3.41 | Scale applied to judgments from interviews . . . . .                                       | 120 |
| 3.42 | Oral interviews - Clausal indirect objects . . . . .                                       | 120 |
| 3.43 | Summary of the findings . . . . .                                                          | 123 |
| 3.44 | RMH - Exceptional instances based on clause type . . . . .                                 | 125 |
|      |                                                                                            |     |
| 4.1  | Competing configurations across syntactic positions . . . . .                              | 160 |
| 4.2  | Competing configurations across syntactic positions - revisited . . . . .                  | 165 |
| 4.3  | Data from RMH on <i>lýsing</i> , <i>skilgreining</i> and <i>gagnrýni</i> . . . . .         | 182 |
|      |                                                                                            |     |
| A.1  | Nominative subjects in first position . . . . .                                            | 217 |
| A.2  | Genitive subjects in first position . . . . .                                              | 217 |
| A.3  | Most common matrix verbs with a nominative clausal subject in first position . . . . .     | 217 |
| A.4  | Constituents following <i>vera</i> and distribution of <i>það</i> . . . . .                | 218 |
| A.5  | Nominative subjects (not extraposed) in Spec,T . . . . .                                   | 218 |
| A.6  | Most common matrix verbs with a nominative clausal subject after the finite verb . . . . . | 218 |
| A.7  | Passivized accusative clausal objects in Spec,T . . . . .                                  | 219 |
| A.8  | Genitive and dative subjects in Spec,T . . . . .                                           | 219 |
| A.9  | Nominative and accusative clausal objects . . . . .                                        | 220 |
| A.10 | Dative and genitive clausal objects . . . . .                                              | 220 |
| A.11 | Post-copular clauses . . . . .                                                             | 221 |
| A.12 | Prepositions assigning accusative case only . . . . .                                      | 221 |
| A.13 | Prepositions assigning dative case only . . . . .                                          | 222 |
| A.14 | Prepositions assigning genitive case only . . . . .                                        | 222 |
| A.15 | Preposition <i>í</i> with accusative and dative case . . . . .                             | 223 |
| A.16 | Preposition <i>yfir</i> with accusative and dative case . . . . .                          | 223 |
| A.17 | Adverb+preposition constructions . . . . .                                                 | 224 |
| A.18 | Complements of nouns . . . . .                                                             | 224 |
| A.19 | Complements of adjectives . . . . .                                                        | 225 |
| A.20 | Active vs. passive . . . . .                                                               | 225 |
| A.21 | Nominative clausal subjects - Middle voice . . . . .                                       | 226 |
| A.22 | Exceptions based on sources . . . . .                                                      | 226 |

|      |                                                                                         |     |
|------|-----------------------------------------------------------------------------------------|-----|
| B.1  | Questionnaires - Comparison between version A and B . . . . .                           | 227 |
| B.2  | Participants in questionnaires - overview . . . . .                                     | 230 |
| B.3  | General results from questionnaires - Part 1 . . . . .                                  | 232 |
| B.4  | General results from questionnaires - Part 2 . . . . .                                  | 233 |
| B.5  | General results from questionnaires - Part 3 . . . . .                                  | 234 |
| B.6  | General results from questionnaires - Part 4 . . . . .                                  | 235 |
| B.7  | General results from questionnaires - Part 5 . . . . .                                  | 236 |
| B.8  | General results from questionnaires - Part 6 . . . . .                                  | 237 |
| B.9  | General results from questionnaires - Part 7 . . . . .                                  | 238 |
| B.10 | Judgments in interviews . . . . .                                                       | 239 |
| B.11 | Judgments from speakers on a scale from 1 to 3 on indirect objects . . . . .            | 240 |
| B.12 | Judgments from speakers on a scale from 1 to 3 on indirect objects - with extraposition | 240 |
| B.13 | Summary of oral interviews - Clausal indirect objects . . . . .                         | 241 |
| B.14 | Summary of oral interviews - Clausal indirect objects - detailed . . . . .              | 241 |
| B.15 | Comparison between age groups . . . . .                                                 | 242 |
| B.16 | Comparison between gender groups . . . . .                                              | 242 |



# List of Figures

- 3.1 Extended tab in RMH (2019) . . . . . 80
- 3.2 Example of query (*fagna* + dative *pað* + *that*-clause) . . . . . 80
  
- A.1 Example of token list (from the *spá* dataset) . . . . . 213
- A.2 Statistical summary (example from *opinbera*) . . . . . 215
- A.3 Green box - Statistical summary (example from *opinbera*) . . . . . 215
- A.4 Example of a *Trends* tab . . . . . 216
  
- B.1 Questionnaires: general overview (example of questionnaire 1B) . . . . . 230



# Acknowledgements

This doctoral project could not have been completed without the precious help of various people, in Iceland and abroad. First of all, I would like to thank the chair of my doctoral committee, Jóhannes Gísli Jónsson, for supporting me during this long journey; for guiding me, especially during those moments in which I really found difficult to discuss the theoretical implications of my analysis; also for challenging my views on the role and distribution of clausal nominalization so that I could look at things from a different perspective.

I am also extremely grateful to the other members of the doctoral committee, Anton Karl Ingason and Matthew Whelpton, for their precious feedback and suggestions. Their comments have been particularly useful to define in a clearer way the methodology I applied to my empirical analysis as well as the results from both corpora and questionnaires.

I am also very thankful to my colleagues in Icelandic as a second language at the University of Iceland, from whom I received feedback and help on various examples I needed to test. My thanks in this regard go in particular to Kolfinna Jónatansdóttir, María Anna Garðarsdóttir, Gísli Hvanndal Ólafsson, Sigríður Dagný Þorvaldsdóttir, Kári Páll Óskarsson, Sigríður Kristinsdóttir, Helga Jónsdóttir, Laufey Guðnadóttir, Salome Lilja Sigurðardóttir, Védis Ragnheiðardóttir and Gísli Rúnar Harðarson. My special thanks also go to Stefanie Bade, for her support during these years, for sharing with me her experience as a doctoral student, for her precious pieces of advice as well as for her useful feedback on certain examples from German I needed to test. Among the staff members of the Árni Magnússon Institute, I would also like to thank Einar Freyr Sigurðsson for his suggestions on case marking and also his feedback when I was a student of his in the course *Föll og fallafræði* 'Cases and case theory' a few years ago during the first years of my doctoral studies at the University.

Very special thanks should also go to my wife Zara. Although she has always considered grammar as her arch-enemy in her teenage years, she has always been highly supportive of my work, especially in those moments when I found difficult to go ahead with my research. Similarly, I would like to thank my family as well as my wife's for their support and love from Sicily and Germany respectively, and also my brothers in faith Mansoor Ahmad Malik, Andrey Klavins and Gunnlaugur Briem here in Iceland. I would also like to thank my Arabic teachers Mazen Maarouf here in Iceland and Tareq Sayed in Egypt, with whom I tried to explore clausal nominalization in Semitic languages as well, although in the end those observations could not be discussed in my dissertation for reasons of space.

Lastly, I also want to extend my thanks to all my former and current students (especially my Russian, Iranian and Afghan students, who helped me to explore the characteristics of clausal nominalization in Persian and Russian) and friends who I discussed my project with and who contributed, even if just a little, to its completion.



# 1 Introduction

## 1.1 Topic and research questions

With the term *nominalized clauses* (which I adopt from Roussou (1991); henceforth called Nom-Cls), in this dissertation I refer to the clausal complements preceded<sup>1</sup> by a determiner, like a demonstrative pronoun (see (1)) or a definite article (as in (2)), or by a personal pronoun (generally one that also functions as an expletive, see (3)), as illustrated in the following examples from Icelandic, Modern Greek and English:<sup>2</sup>

- (1) **Það** að læra á hljóðfæri er eins og að læra að hjóla  
that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
'Learning to play an instrument is like learning to ride a bike.'  
(Garofalo 2023:106) ICELANDIC
- (2) **To** oti lei psemata apodhiknii enohi tis  
the that tells lies proves guilt her  
'The fact that she tells lies proves her guilt.' (Roussou 1991:87) M. GREEK
- (3) I hate **it** that I have to go with my parents out of town this weekend

NomCls are a frequent phenomenon from a cross-linguistic perspective and coexist in the same language with clausal complements that are not preceded by a determiner or pronoun. The distribution of these two types of clausal variants often differs, as the *clausal nominalizer* (henceforth CN) can be optional, mandatory, or ungrammatical under certain conditions (see e.g. Thráinsson (1979); Roussou (1991); Wood (2012); Ingason (2018); Knyazev (2016); Hartman (2012); also Garofalo (2020; 2023) among many others), which also indicates that their structure must be different.

The existence of NomCls alongside bare clauses (henceforth called non-NomCls) raises two important questions. The first one is how NomCls and non-NomCls differ from a structural perspective. In the linguistic literature, the answer to this question is still an object of discussion, as it has been proposed for both variants that they are complementizer phrases (CPs), noun phrases

---

<sup>1</sup>Within this definition, I focus exclusively on those instances where the clausal nominalizer precedes the clause. However, it is also true that there are other languages like Korean and Japanese (see e.g. Hara et al. (2013); Han (2008); also cf. Moulton et al. (2020)) where the nominalizing element occurs at the end of the clause. In this dissertation, I will mainly focus from a cross-linguistic perspective on those languages where the nominalizing element is cataphoric, since Icelandic, which is the target language of this analysis, belongs to this category.

<sup>2</sup>The phenomenon of NomCls has also been defined as *nouny propositional expressions* or *nouny CPs* (see Moulton et al. (2020)), or *clausal prolepsis* (see e.g. Angelopoulos (2022)), while the determiner involved has also been defined as an anticipatory pronoun (see Kaltenböck (2003)) and expletive (e.g. Jahromi (2011)). I prefer to adopt the term *clausal nominalization* for this phenomenon because, as will be shown in the following pages, the clausal argument is embedded in a DP shell of which the determiner itself is part.

(NPs) or determiner phrases (DPs) (see e.g. Yoon (2001); Knyazev (2016); Borsley and Kornfilt (2000); Pietraszko (2019); Hartman (2012); Garofalo (2020; 2023); Han (2005); Roussou (1991); Ingason (2018); Stroik (1996); Kastner (2015) among others; see in particular §2.4 for a detailed discussion on the structure of both variants).<sup>3</sup> The second question is why there is a need for CNs before clauses or, alternatively, why they are allowed to surface in the first place if non-NomCl complements already exist. Many analyses that have been proposed to explain the role and the distribution of CNs are based on various languages, for the most part English (see e.g. Rosenbaum (1967); Kaltenböck (2003); Kallulli (2006); Kim and Sag (2005); Shahar (2008); Gentens (2016) among many others), but also German (e.g. Sudhoff (2016); Axel-Tober et al. (2016); Zimmermann (2016)), Modern Greek (e.g. Roussou (1991)), Persian (Farudi (2007); Jahromi (2011)), Russian (Hartman (2012); Knyazev (2016)), Ndebele (Pietraszko (2019)), Spanish (Delicado Cantero (2013); Picallo (2002)), Hungarian (De Cuba and Ürögdi (2010)) and Hebrew (Zuckermann (2006); Kastner (2015)). Some of these analyses link clausal nominalization to various triggering factors like factivity (e.g. Thráinsson (1979); Kallulli (2006)) or other semantic characteristics of matrix verbs selecting clausal arguments (e.g. Kastner (2015)), case marking (Roussou (1991); Knyazev (2016)), movement (e.g. Shahar (2008)) and DP selection (e.g. Borsley and Kornfilt (2000); Kornfilt and Whitman (2011a); Panagiotidis and Grohmann (2009)), but more data are needed in order to propose a more precise syntactic model of clausal complementation that can also explain the existence and distribution of NomCls.

There are also studies of clausal nominalization in Icelandic, which is the main topic of this dissertation. The seminal work in this regard is Chapter 4 in Thráinsson's (1979) doctoral dissertation, which had a major focus on extraposed subject and object clauses and analyzed to some extent the presence of determiners before clausal arguments. His analysis concluded that factivity plays a major role in the distribution of NomCls (see §2.5.2), in the sense that emotive factive predicates tend to select NomCls more often. After his work, however, clausal nominalization has been put aside in the linguistic research on Icelandic. More recent accounts discuss the phenomenon to some extent (see Wood (2012); Ott (2014), Garofalo (2015; 2020; 2023) and Ingason (2018)), but it has never been investigated thoroughly. In fact, research on Icelandic clausal nominalization, but also on clausal nominalization cross-linguistically, has focused too often on subjects and direct objects only (in the case of Icelandic, Garofalo's (2015; 2020; 2023) accounts are an exception, where prepositional objects, indirect objects and postcopular clauses have been analyzed to some extent) and, therefore, the phenomenon has not been sufficiently investigated. What is more, all of the cross-linguistic analyses listed above are either partially or entirely problematic when applied to Icelandic clausal nominalization (see §2.5), especially because their predictions on the distribution of NomCls and non-NomCls are incompatible with some of the empirical facts of Icelandic. Hence, this dissertation will present the very first ex-

<sup>3</sup>Certain analyses consider the existence of clausal arguments embedded into a shell where no element is phonetically realized to be plausible, e.g. Knyazev (2016), Kastner (2015) or Takahashi (2010). In a way, these clausal arguments can also be considered nominalized. However, it will be demonstrated in this dissertation that NomCls surface as DPs in Icelandic and non-NomCls as CPs, so they do not constitute an issue.

tensive analysis of Icelandic clausal nominalization, in which I will revise the previous proposals and fill in the gaps left in the linguistic literature on Icelandic, examining all the syntactic positions where NomCIs can surface.

From a cross-linguistic perspective, one of the ultimate objectives of this work is to argue for an interpretation that can account for clausal nominalization in Icelandic and other languages as well. For this purpose, it is pivotal to compare how clausal nominalization functions in Icelandic and other languages (both closely and distantly related), and discover what Icelandic clausal nominalization can tell us about other languages as well as what NomCIs in other languages can tell us about Icelandic.

The research questions that will be answered in this dissertation are therefore as follows:

1. What is the role of clausal nominalization in Icelandic?
2. What is the structure of NomCIs in Icelandic? How do they differ from non-NomCIs?
3. What factors determine the syntactic distribution of NomCIs and non-NomCIs?
4. What can the facts about clausal nominalization in Icelandic tell us about other languages and vice versa?

The first question I will discuss in this dissertation is on the structure of clausal arguments, after presenting an overview of clausal nominalization in Icelandic and a discussion on the relevant literature (see Chapter 2). Then, after determining the structural difference between NomCIs and non-NomCIs and observing a wide sample of empirical data, I will provide information on what factors determine the distribution of NomCIs and, consequently, the role of CNs in Icelandic (see Chapter 3). Then I will discuss the derivation of NomCIs and non-NomCIs as well as the relevant theoretical implications, and will also compare clausal nominalization in Icelandic with four other languages (see Chapter 4).

## 1.2 Hypothesis

The main hypothesis I propose to interpret the role and the distribution of clausal nominalization in Icelandic is illustrated below:<sup>4</sup>

(4) *Clausal Nominalization Mechanism (CNM)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and can only surface if (it is enough to satisfy one condition):

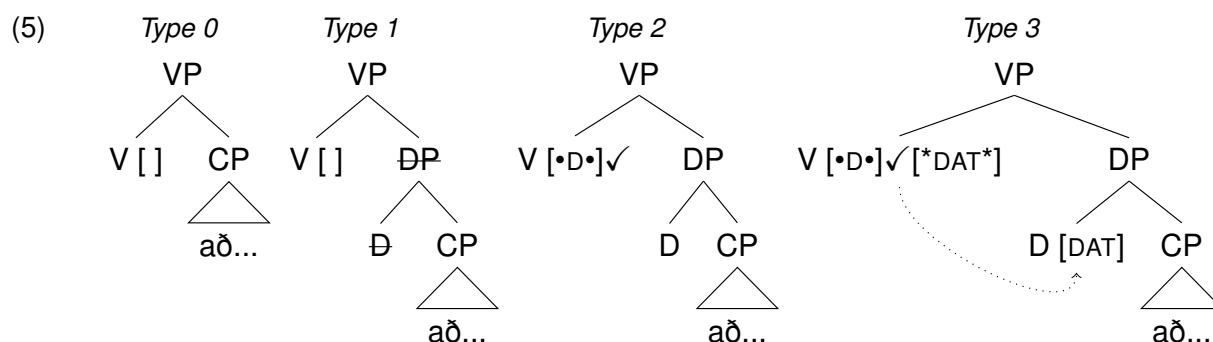
- a. CN checks at least one D-,  $\varphi$ - or case feature, or
- b. the clause merges with an item bearing a [ $\bullet$ D $\bullet$ ] feature

---

<sup>4</sup>I assume the feature notation presented in Heck and Müller (2007), where bullet features [ $\bullet$ x $\bullet$ ] trigger Merge of a specific category. Features that trigger Agree are noted as [ $\ast$ x $\ast$ ]. Further discussion is offered in Chapter 4.

CNM addresses to a great extent the first three research questions presented in the previous paragraph. The role of a CN is to check syntactic features that can only be checked by DPs (e.g. case, gender and number features) and that a CP by itself is consequently unable to check (see §3.3.8). From a structural perspective, NomCIs are subsumed within the distribution of DPs, which indicates that they are DPs as well; non-NomCIs, by contrast, emerge as CPs (see §2.4.4 for a detailed discussion on the structural difference between these two clausal variants). In particular, CNM suggests that, if no syntactic feature among those specified in condition (a) or (b) needs to be checked by the relevant clause, then the clause can emerge as a non-NomCI (more discussion is offered in §4.3). As for the factors determining the distribution of both variants, CNM indicates that feature checking is the criterion based on which *það* either emerges or does not. For a clause to be nominalized, it is enough that one feature needs to be checked by the relevant clause, otherwise it emerges as a non-NomCI.

From a derivational perspective, more specifically, I propose that the distribution of NomCIs and non-NomCIs is mainly accounted for by four different configurations, which are sketched below (the tree diagrams refer to direct object position, but they can be extended to other positions as well).<sup>5</sup>



*Type 0* accounts for non-DP positions, where NomCIs are ungrammatical by virtue of the fact that they are DPs. The remaining configurations compete in DP positions (a detailed mapping of the distribution of these configurations is offered in §4.3.7). *Type 3* is triggered if conditions (a) and (b) of CNM are satisfied, while *Type 2* follows from condition (b), as a  $[•D•]$  feature requires Merge of a DP complement. As for *Type 1*, I propose that those clauses that are merged in DP positions and emerge as non-NomCIs are originally DPs that undergo removal of the DP shell if neither condition in CNM is satisfied (a detailed discussion is offered in §4.3.5).<sup>6</sup>

An important implication of my hypothesis from a theoretical perspective is that nominative and accusative case in Icelandic do not correspond to formal features to be checked in narrow syntax. By contrast, dative and genitive correspond to features to check and trigger clausal nominalization

<sup>5</sup>I propose, however, a slightly different variant of *Type 3* for post-copular clauses introduced by the pronouns *sá* ‘that (masculine)’ and *sú* ‘that (feminine)’. This proposal is discussed in §4.3.9.

<sup>6</sup>There are already some accounts on the existence of a structural removal rule in the linguistic literature. Some of them claim that movement is the trigger of such an operation (see e.g. Pesetsky (2021)), while others claim that certain syntactic features trigger structural removal (see e.g. Müller (2017); Korsah and Murphy (2019)). These accounts will also be discussed in §4.3.5.

(see §4.4 for a more detailed discussion). A similar reasoning can be extended to singular and neuter in comparison with non-default gender and number (cf. Garofalo (2023); see §4.4.6).

As for the answer to the fourth research question I presented in the previous paragraph, a comparison between Icelandic and four other languages (Swedish, German, Persian and Russian) shows that lexical case only triggers clausal nominalization when a language displays a morphological case system. As for structural case, it does not trigger clausal nominalization independently of whether the language has a morphological case system or not. Languages without a case system can allow clausal nominalization to be optional as a general rule (as in Swedish, see §4.5.2) or tend to require DPs *in situ* (as in Persian, see §4.5.4). However, all the languages studied show that clausal subjects can only be NomCIs in Spec,T unless they escape this position. This suggests that D-feature checking is a common trigger of clausal nominalization cross-linguistically.

### 1.3 Methodology

Since the main purpose of this dissertation is to propose an analysis of clausal nominalization in Icelandic valid for all syntactic positions, and due to the fact that previous analyses of Icelandic do not fully capture the whole picture of clausal nominalization (see §2.5), I collected empirical data on both NomCIs and non-NomCIs from two main sources. On the one hand, I retrieved examples from the so-called *Icelandic Gigaword Corpus* (*Risamálheildin* in Icelandic, henceforth RMH; see Steingrímsson (2019)). All the valid examples have been classified based on the presence or absence of the CN, the clause type, the case assigned, the item selecting the clausal complement as well as the source the token has been taken from. On the other hand, I administered a series of questionnaires (20, with 25 sentences each) to different samples of native speakers of Icelandic. The choice of resorting to two sets of data is due to various reasons. Firstly, observing both naturalistic data and judgments from native speakers can ensure more solid results (see Hoffmann (2006)), but, at the same time, it can reveal more details on the distribution of NomCIs and non-NomCIs that only one set of data would be unable to display.<sup>7</sup> Secondly, naturalistic data are needed in order to prepare the examples that will have to be tested in questionnaires. Thirdly, since speakers differ in their judgments and might be biased towards giving the ‘correct’ answers, or those answers they think the researcher is hoping for, it is necessary to compare judgment tasks with naturalistic data (cf. Þráinsson et al. (2013); Hoffmann (2006); Myers (2009); also §3.2.2 for a more detailed discussion on this methodology).

Moreover, regarding questionnaires specifically, I also briefly observed whether clausal nominalization exhibits any variation across age and gender groups (§3.3.10). In the case of clausal indirect objects, I also interviewed some speakers orally, as the results from questionnaires turned out to be unfruitful to a certain degree (§3.3.7) and there were no data from corpora available.

---

<sup>7</sup>The results in Chapter 3 confirm the importance of observing both data from corpora and questionnaires, as one can notice a systematic difference between the two sets of data.

## 1.4 Structure

This dissertation is organized as follows. In the second chapter, I will discuss the main properties of clausal nominalization and the relevant hypotheses that have been presented in the literature on Icelandic and other languages, as well as a brief preliminary overview of my hypothesis. In the third chapter, I will discuss the results of both naturalistic data and questionnaires (and also interviews) and will refine my hypothesis to explain the distribution of NomCIs and non-NomCIs, in light of the data retrieved in the empirical analysis. In the fourth chapter, I will discuss the distribution of these two clausal variants from a derivational perspective and will argue in detail for the four configurations I have presented in this introduction, as well as specific theoretical implications in relation to case, gender and number categories. Then I will also discuss clausal nominalization from a cross-linguistic perspective and compare Icelandic with Swedish, German, Persian and Russian. In the fifth and final chapter, the main findings of this dissertation will be summarized.

## 2 Clausal nominalization

### 2.1 Introduction

In this chapter, I will take a first close look at the properties of clausal nominalization in Icelandic and discuss its distributional puzzle in light of the previous relevant hypotheses that have been presented for Icelandic and in the cross-linguistic literature. It is important to carefully observe the phenomenon analyzed here, as the distribution of *það* as CN in Icelandic is quite complex compared to other languages (e.g. English, where *it* can only precede clausal subjects and objects), due to the number of positions where it can occur, as well as other properties related to clausal nominalization like extraposition.

Once the most relevant properties and issues of this phenomenon in Icelandic have been observed, I will focus on the theoretical part of this chapter, in which the most relevant hypotheses that have been presented in the literature on clausal nominalization (both in Icelandic and other languages) will be discussed, in particular in relation to the structure of NomCIs and non-NomCIs and their distribution across syntactic positions. Subsequently, I will present a preliminary overview of my hypothesis on the distribution and the purpose of clausal nominalization on the basis of the information discussed in this chapter.

There are two main findings of this chapter. Firstly, I will demonstrate that NomCIs and non-NomCIs surface as DPs and CPs respectively,<sup>1</sup> as: a) data from extraction indicate that an item can only be moved from a clausal argument if *það* does not surface; and b) data from the clausal complements of verbs like *þvinga* ‘force’ or *hjálpa* ‘help’ show that Icelandic NomCIs can only appear in positions where DPs are allowed, otherwise only a CP is possible. Secondly, I propose that the various hypotheses that have been presented in the linguistic literature for Icelandic clausal nominalization as well as clausal nominalization cross-linguistically are either completely or partially problematic with the Icelandic data. In particular, I argue that the factivity hypothesis presented by Thráinsson (1979), based on which Icelandic NomCIs tend to surface when the matrix verb is factive (especially when the clause is in direct object position), fails to capture the distribution of clausal nominalization across syntactic positions. There are, in fact, examples of NomCIs that surface when the verb is non-factive as well as instances of non-NomCIs that optionally surface after factive verbs.

An important reason why the various hypotheses presented in the linguistic literature are unable to properly map the distribution of Icelandic clausal nominalization is pronominal optionality. Various accounts that will be discussed in the following sections focus on explaining the reason

---

<sup>1</sup>Note that, at this point of the analysis, I will only focus on how NomCIs and non-NomCIs look like after the derivation. After observing more closely all the empirical data in Chapter 3, I will propose in Chapter 4 that both NomCIs and non-NomCIs are originally merged as DPs in DP positions, while CPs are merged as such in non-DP positions.

why a NomCI is needed or why a NomCI is disallowed. However, the matter becomes less clear when both variants are acceptable in the same position. If a CP can already be a suitable complement of a certain item, there is no necessity for a NomCI to appear. Yet, a NomCI is still allowed. In this regard, I propose at the end of this chapter a hypothesis that can explain both pronominal optionality and pronominal obligatoriness as the product of a similar mechanism for NomCIs. More specifically, I argue that NomCIs in Icelandic are systematically licensed in positions where a DP can appear. However, pronominal obligatoriness only differs from pronominal optionality in that non-NomCIs are ruled out by feature checking operations like case assignment, in line with Stowell's (1981) *Case Resistance Principle*.

This chapter will be divided into five main sections. In the first one, I will present the main properties of clausal nominalization in Icelandic. In the second, I will present the main puzzle of this dissertation, i.e. the distribution of NomCIs and non-NomCIs. In the third, I will discuss the structural models proposed in the literature for both clausal variants. In the fourth, I will discuss the previous hypotheses that have been presented to explain clausal nominalization, both in Icelandic and other languages, and I will also broadly delineate a preliminary version of my hypothesis, in light of the information offered in the current chapter, to explain the purpose of CN *það* and its distribution. In the fifth subsection, the most relevant findings of this chapter will be summarized.

## 2.2 A brief overview

### 2.2.1 Distribution of *það*

The determiner that is allowed to precede clauses in Icelandic is the demonstrative pronoun *það* 'that'.<sup>2</sup> This pronoun can be inflected for case, as shown in Table 2.1, and most often does not require any previous context in order to surface and does not contribute to the meaning of the sentence, as will be shown in many of the examples presented in this dissertation (see in particular the data from judgment tasks in Appendix B, which are tokens extrapolated from any context).<sup>3</sup>

| Nom | Acc | Dat | Gen  |
|-----|-----|-----|------|
| það | það | því | þess |

Table 2.1: Inflection of *það*

Clausal nominalization in Icelandic has a very extended distribution. Due to the fact that Icelandic displays a V2-structure (see Thráinsson (2007) among many others), NomCI subjects can be found before as well as after the finite verb ((1a–b) respectively; see §2.3.1 for further discussion

<sup>2</sup>In §2.2.2, I will discuss how *það* can only be a demonstrative and not a personal pronoun (see also Garofalo (2020)). In footnote 5 in the current chapter, I explain why I list the CN as neuter *það* and not as masculine *sá*.

<sup>3</sup>Thráinsson (1979), however, considers certain instances of *það* before clausal subjects and objects as discourse-anaphoric. This will be discussed in §2.4.2 and §2.5.2.

on these two positions). Clausal nominalization is also displayed in direct objects (see (1c)), postcopular position (see (1d)), prepositional complements (see (1e)), indirect objects (see (1f)) and complements of nouns (see (1g)) and adjectives (see (1h)) (some of these examples can also be found in Garofalo (2020:6-7)):<sup>4</sup>

- (1) a. [**Það** að læra á hljóðfæri] er eins og að læra að hjóla  
that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
'Learning to play an instrument is like learning to ride a bike.' (Garofalo 2023:106)
- b. Er [**það** að læra á hljóðfæri] eins og að læra að hjóla?  
is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
'Is learning to play an instrument like learning to ride a bike?' (Garofalo 2023:106)
- c. María krafðist [**þess** að skýrslunni yrði skilað strax]  
María demanded that<sub>GEN</sub> that report.the became submitted immediately  
'María demanded that the report be submitted immediately.'
- d. Vandamálið er [**það** að við skuldum meira núna]  
problem.the is that<sub>NOM</sub> that we owe more now  
'The problem is that we owe more money now.'
- e. Allardyce er svekktur yfir [**því** að hafa fengið sparkið]  
Allardyce is annoyed over that<sub>DAT</sub> to have got kick.the  
'Allardyce was annoyed about the fact that he was fired.'
- f. Vopnahléið svipti [**það** að hermenn fórnúðu lífi sínu] öllum hetjuljóma  
armistice.the deprived that<sub>ACC</sub> that soldiers sacrificed life their all<sub>DAT</sub> heroism<sub>DAT</sub>  
'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'  
(Garofalo 2023:113)
- g. Þessi samningur er gerður til verndar [**því** að starfsmennirnir séu ekki þvingaðir til að senda formlegt kvörtunarbréf]  
this contract is made to prevention that<sub>DAT</sub> that workers.the are not compelled to to send formal letter of complaint  
'This contract is made to prevent the workers from feeling compelled to send a formal letter of complaint.'
- h. Ég er feginn [**því** að þú skulir vera kominn]  
I am satisfied that<sub>DAT</sub> that you shall be arrived  
'I'm happy that you have come.'

The typical clausal complements that can be preceded by a determiner in Icelandic are *that*-clauses, infinitives and indirect questions (cf. Kress (1982); see various examples of clausal complements in Thráinsson (1979); also Garofalo (2015; 2020; 2023)). One example for each category is illustrated in (2), in the same order:

- (2) a. Ég ákvað [**það** að María þarf ekki að sjá um þetta lengur]  
I decided that<sub>ACC</sub> that María needs not to see about this longer  
'I decided that María does not have to take care of this any longer.'

<sup>4</sup>Since the distributional value of *það* (e.g. mandatory or optional) in one specific complement position varies depending on the lexical item selecting the clausal argument (see Chapter 3), I am not going to specify in the examples here whether *það* is optional or mandatory.

- b. Það snýst um [það að vernda viðskiptavini okkar]  
it is about that<sub>ACC</sub> to protect customers our  
'It is about protecting our customers.'
- c. Við spurðum hann um [það hversu mikið er hægt að fjárfesta]  
we asked him about that<sub>ACC</sub> how much is possible to invest  
'We asked him how much it is possible to invest.'

Let us now take a closer look at postcopular clauses (as in (1d)). These are the only ones where, whenever nominalization occurs, *það* inflects for gender, number and case in agreement with the subject. Evidence of this agreement can be found in particular in instances of *Exceptional Case Marking* (henceforth ECM), where the determiner surfaces in accusative case instead of nominative, and gender and number features are still assigned to the pronoun. Compare the nominative feminine determiner *sú* 'that' with the corresponding accusative feminine determiner *þá* in the examples below:

- (3) a. Afleiðingin er [sú að við skuldum meira núna]  
consequence.the<sub>NOM-F-SG</sub> is that<sub>NOM-F-SG</sub> that we owe more now  
'The consequence is that we owe more money now.' (adapted from Garofalo (2020:9))
- b. Ég tel afleiðinguna vera [þá að við skuldum meira núna]  
I consider consequence.the<sub>ACC-F-SG</sub> be that<sub>ACC-F-SG</sub> that we owe more now  
'I consider the consequence to be that we owe more money now.' (adapted from Garofalo (2020:9))

For the sake of completeness, in Table 2.2 I present the inflectional forms we can find for the determiner<sup>5</sup> if we also count all the possible outcomes with postcopular clauses based on gender and number.<sup>6</sup>

<sup>5</sup>The inflectional forms missing in the table do exist in Icelandic, but they never occur as inflectional forms of CN *það*. Postcopular clauses can only surface in nominative or accusative case (via ECM), but not in dative or genitive case. An important question the reader might ask about the word forms presented in the table is whether at this point it is not better to call the Icelandic CN *sá* instead of *það* since the lemma of this pronoun is the masculine singular nominative form. Although this is true, since the neuter form is much more extended across syntactic positions as CN, I believe it is more convenient and easier to call the determiner *það*.

<sup>6</sup>The distribution of *það* in Icelandic is also more extended due to the fact that it can also precede temporal PPs (see Garofalo (2020)):

- (i) a. Verðin eru enn þá að lækka frá því [pp í júlí]  
prices.the are still to decrease from that<sub>DAT</sub> in July  
'The prices have still been decreasing since July.' (adapted from Garofalo (2020:22))
- b. Þetta kemur fram í reglugerð heilbrigðisráðherra frá því [pp um helgina]  
this comes forth in regulation health.minister from that<sub>DAT</sub> in weekend.the  
'This is stated in the health minister's regulation from last weekend.'  
(adapted from Garofalo (2020:22))

This phenomenon, however, is outside the scope of this dissertation as PPs are not clausal complements and will not be analyzed here.

| <b>Sing.</b> | <b>Masculine</b> | <b>Feminine</b> | <b>Neuter</b> |
|--------------|------------------|-----------------|---------------|
| Nom.         | sá               | sú              | það           |
| Acc.         | þann             | þá              | það           |
| Dat.         |                  |                 | því           |
| Gen.         |                  |                 | þess          |
| <b>Plur.</b> | <b>Masculine</b> | <b>Feminine</b> | <b>Neuter</b> |
| Nom.         | þeir             | þær             | þau           |
| Acc.         | þá               | þær             | þau           |

Table 2.2: Complete inflection of CN *það*

## 2.2.2 Demonstrative *það* as CN

At the beginning of this section, I defined CN *það* as a demonstrative pronoun straight away. However, one might raise an objection to such a claim by saying that it may well be a personal pronoun (since *það* can also mean ‘it’ in Icelandic), especially if one considers the fact that other Germanic languages like English and German provide cataphoric pronouns as well, which are personal pronouns (*it* and *es* respectively, see e.g. Kaltenböck (2003); Kim and Sag (2005); Sudhoff (2016) among many others). However, we have already observed that, in postcopular clauses, when a CN surfaces in masculine and feminine singular, the pronouns that surface are *sá* or *sú* ‘that’, which are demonstrative pronouns. But the personal pronouns *hann* ‘he’ or *hún* ‘she’ are disallowed:

- (4) **Afleiðingin** er [sú/\*hún að við skuldum meira núna]  
 consequence.the<sub>NOM-F-SG</sub> is that<sub>NOM-F-SG</sub>/she that we owe more now  
 ‘The consequence is that we owe more money now.’ (Garofalo 2020:9)

There are further arguments<sup>7</sup> in favor of a demonstrative *það*. Firstly, it was suggested in Práins-son et al. (2005:339, footnote 10) that *það* is a demonstrative pronoun due to the fact that complex DPs are also generally introduced by a demonstrative pronoun:

- (5) **Sú** staðreynd að ...  
 that<sub>NOM-F-SG</sub> fact<sub>NOM-F-SG</sub> that ...  
 ‘The fact that...’

Secondly, Icelandic also provides two other determiners which can precede clauses, but only when they refer to a previous discourse. They are *þetta* ‘this’ and *hitt* ‘the other’, which are both demonstratives and not personal pronouns (the examples are from RMH; see also some discussion in Thráinsson (1979:255-256) on this topic):

<sup>7</sup>Some of these arguments have also been presented in Garofalo (2020), where it is claimed that *það* is a demonstrative pronoun.

- (6) a. **Petta** að menn hjá okkur hugsi um afkvæmi grunar mig að sé  
 this<sub>NOM</sub> that men among us think about offspring suspects me that is  
 vegna menningar  
 because of culture  
 'I suspect that the fact men here think about their offspring is because of culture.'
- b. **Hitt** að hvítt kjöt í Bandaríkjunum sé ódýrt vissum við fyrir löngu  
 the other<sub>ACC</sub> that white meat in United States.the is cheap knew we for long  
 'The other fact – that white meat in the US is cheap – is something that we knew a  
 long time ago.'

Thirdly, from a cross-linguistic perspective, languages that display a demonstrative as CN tend to be characterized by optional clausal extraposition. This is the case of languages like Persian (see (7); cf. Farudi (2007), Hartman (2012) and Jahromi (2011); also §4.5.4) and German (with the CN *das* 'that', as shown in (8); see Frey (2016); more discussion on German clausal nominalization is offered in §4.5.3). Icelandic follows the same trend, see (9):<sup>8</sup>

- (7) a. [**In** ke Maryam raft] maalum e  
 this that Maryam left clear is  
 'It is clear that Maryam left.' (Hartman 2012:36)
- b. **In** maalum e [ke Maryam raft]  
 this clear is that Maryam left  
 'It is clear that Maryam left.' PERSIAN
- (8) a. Ich kann mich [**dessen**, dass er seine Mutter angelogen hat], nicht entsinnen  
 I can me<sub>ACC</sub> that<sub>GEN</sub> that he his mother lied has not remember  
 'I cannot recall him ever lying to his mother.'
- b. Ich kann mich **dessen** nicht entsinnen, [dass er seine Mutter angelogen hat]  
 I can me<sub>ACC</sub> that<sub>GEN</sub> not remember that he his mother lied has  
 'I cannot recall him ever lying to his mother.' GERMAN
- (9) a. [**Það** að hann skuli vera farinn] er skrítið  
 that<sub>NOM</sub> that he shall be gone is strange
- b. **Það** er skrítið [að hann skuli vera farinn]  
 that<sub>NOM</sub>/it? is strange that he shall be gone  
 'It is strange that he left.'

By contrast, in languages where clauses are introduced by personal pronouns, clausal extraposition is mandatory. This is the case of English (see e.g. Kaltenböck (2003) among many others) or Dutch (Angelopoulos (2021)), but also German when it is expletive *es* 'it' that introduces clausal complements (Axel-Tober et al. (2016); also §4.5.3):

- (10) a. **It** is surprising [that John went to London] (Kaltenböck 2003:235)  
 b. \***[It** that John went to London] is surprising

<sup>8</sup>In example (9), *það* is ambiguous after clausal extraposition due to the fact that it can either be a CN or an expletive. Therefore, I wrote both readings in the relevant glosses.

- (11) a. Jan zei dat hij **het** toch hoopte [dat deze beslissing genomen was]  
 Jan said that he it yet hoped that this decision taken was  
 'Jan said that he hoped this decision would be made.'
- b. \*Jan zei dat hij **het** [dat deze beslissing genomen was] toch hoopte  
 Jan said that he it that this decision taken was yet hoped  
 (Angelopoulos 2021:14) DUTCH
- (12) a. Er hat **es** bedauert, dass er schwarz gewählt hat  
 he has it regretted that he black voted has  
 'He regretted having voted black.'
- b. \*Er hat **es**, dass er schwarz gewählt hat, bedauert  
 he has it that he black voted has regretted  
 (Axel-Tober et al. 2016:52) GERMAN

The Icelandic pattern of extraposition is similar to the languages that take a demonstrative as CN, which suggests that *það* is a demonstrative pronoun and not a personal pronoun.

### 2.2.3 A distal demonstrative as best candidate for clausal nominalization

One important issue about *það* as a demonstrative pronoun, however, needs to be addressed, namely why it is that this specific pronoun is selected as CN in Icelandic by default and not another item. Icelandic features different categories of pronouns and determiners: the proximal demonstrative *þessi* 'this', the personal pronoun *það* 'it' homophonous with CN *það*, the suffixed definite article *-(i)nn* 'the' as well as a free article, *hinn* 'the'. Let us briefly consider the various options enumerated here.

A suffixed definite article can hardly be able to introduce a clausal complement in Icelandic. The definite article is only suffixed to nouns and not to whole constituents:

- (13) \*[Að læra íslensku] **-(i)ð** er skemmtilegt  
 to learn Icelandic the<sub>N-SG</sub> is fun  
 'Learning Icelandic is fun.'

However, as just mentioned, Icelandic also has a free article, *hinn*, which precedes nouns. But *hinn* is problematic as well, as it cannot simply precede nouns. It must always be modified by an adjective (see Ingason (2016)), even if no noun is phonetically realized (see (14c)):

- (14) a. **hinn** bjarti dagur  
 the<sub>M-SG</sub> bright<sub>M-SG</sub> day<sub>M-SG</sub>  
 'The bright day'
- b. \***hinn** dagur  
 the<sub>M-SG</sub> day<sub>M-SG</sub>  
 Intended: 'The day'
- c. **hið** opinbera  
 the<sub>N-SG</sub> official<sub>N-SG</sub>  
 'The authorities'

Personal pronouns can also not be valid alternatives to introduce clauses. They are generally not able to select a complement, so they are very often isolated (see Práinsson et al. 2005:66–67). For instance, let us consider the following examples:

- (15) a. *Bíllinn hennar*  
 car.the<sub>M-SG</sub> her  
 ‘Her car’  
 b. \***Hann** hennar  
 he her  
 Literally: ‘Her he’ (where *he* refers to the car)

The noun *bíll* ‘car’ is a masculine noun and can be replaced with the appropriate masculine personal pronoun. However, the masculine pronoun *hann* ‘he’ cannot select the genitive pronoun *hennar* ‘her’ as its complement. By contrast, other pronouns, like *sá* ‘that’, as well as quantifiers are capable of doing so:<sup>9</sup>

- (16) a. **Nokkrar** ykkar ...  
 some<sub>F-PL</sub> of you  
 ‘Some of you’  
 b. **Þær** ykkar ...  
 those<sub>F-PL</sub> of you  
 ‘Those of you’

The same restriction on personal pronouns can be applied while juxtaposing a CP as a complement of a personal pronoun. Consider, for instance, relative clauses, as shown in the following examples:

- (17) a. *Ég synjaði umsókninni* [sem þú vildir ekki lesa yfir]  
 I rejected application.the<sub>F-SG</sub> that you wanted not read over  
 ‘I rejected the application that you didn’t want to check.’  
 b. \**Ég synjaði henni* [sem þú vildir ekki lesa yfir]  
 I rejected her that you wanted not read over  
 Literally: ‘I rejected her (=the application) that you didn’t want to check.’  
 c. ?*Ég synjaði þeirri* [sem þú vildir ekki lesa yfir]  
 I rejected that<sub>F-SG</sub> that you wanted not read over  
 Literally: ‘I rejected that (=the application) that you didn’t want to check.’

A personal pronoun is not suitable to replace the noun in dative case *umsókninni* ‘the application’, as it is generally not able to take a complement. A demonstrative like *sá* is better, but it has an anaphorical function in this case (see e.g. Práinsson et al. (2005:82)), therefore more context is needed in the example above for it to be fully grammatical.

At this point, we are left with two demonstrative pronouns, *þessi* and *sá*, which means that we

<sup>9</sup>Note that *þær* is also homophonous with the feminine plural personal pronoun *þær* ‘they’. But the example in (16b) cannot be an example of a personal pronoun, otherwise (15b) should also be grammatical. Since that is not the case, *þær* can only be interpreted as a demonstrative in the example (see Práinsson et al. (2005:67)).

need to find out why it is a distal demonstrative that has been selected as a determiner. Recall that we know for a fact that all demonstrative pronouns, even *hitt* ‘the other’, are able to introduce clausal complements (the examples in (6) are re-illustrated below for the ease of the reader):

- (18) a. **Petta** að menn hjá okkur hugsi um afkvæmi grunar mig að sé  
 this<sub>NOM</sub> that men among us think about offspring suspects me that is  
 vegna menningar  
 because of culture  
 ‘I suspect that the fact men here think about their offspring is because of culture.’
- b. **Hitt** að hvítt kjöt í Bandaríkjunum sé ódýrt vissum við fyrir löngu  
 the other<sub>ACC</sub> that white meat in United States.the is cheap knew we for long  
 ‘The other fact – that white meat in the US is cheap – is something that we knew a long time ago.’

The question is whether *sá* displays some characteristics that *þessi* does not have and that causes the former to be the default CN. One possible clue to solve this puzzle can be found in Sigurðsson (2006:199) who suggested that *sá*, in absence of a linguistic context, requires the presence of a relative clausal complement. An example is shown below:

- (19) a. **Sá** sem aldrei hefur komið á Hornstrandir á mikið eftir  
 that that never has come to Hornstrandir has much after  
 ‘Whoever has never visited Hornstrandir still has a lot to see.’  
 (Práinsson et al. 2005:82)
- b. ?\***Sá** á mikið eftir  
 that has much after  
 ‘That has a lot to see.’

If the clause does not follow the distal demonstrative, the sentence is not felicitous. By contrast, pronouns like *þessi* are not problematic in this regard (note that *þessi* can refer to an individual here):

- (20) **Þessi** á mikið eftir  
 this has much after  
 ‘This has a lot to see.’

In other words, it appears that *sá* has an intrinsic predisposition for taking CP complements, which could have made it more likely than *þessi* to introduce NomCIs as it does not necessarily require a CP in a default context.

Let us also not forget that complex DPs are commonly introduced by *sá* and not by *þessi*, which also corroborates the choice of a distal demonstrative as CN (example (5) is re-illustrated below):

- (21) **Sú** staðreynd að ...  
 that<sub>F-SG</sub> fact<sub>F-SG</sub> that ...  
 ‘The fact that...’ (Práinsson et al. 2005:339, footnote 10)

These facts indicate that *sá* is the best candidate among the pronouns and determiners to introduce NomCIs in Icelandic.

## 2.2.4 Clausal extraposition

Let us now take a closer look at clausal extraposition with non-NomCIs and NomCIs. Generally speaking, clausal extraposition is usually optional in Icelandic, as we have already seen in the previous paragraphs. It is true, however, that instances of mandatory extraposition also exist, in line with the so-called *Principle of End-Weight* (cf. e.g. Quirk (1985)), based on which complex structures tend to be postponed compared to lighter constituents. The following is one of those examples where the non-extraposed version is degraded (see also Thráinsson (1979:27)):

- (22) a. Hún hefur oft sagt (**það**) við mig [að hún sé orðin leið á starfinu]  
 she has often told that<sub>ACC</sub> to me that she is become tired of job.the  
 'She has often told me that she has got tired of her job.'
- b. ??Hún hefur oft sagt [(**það**) að hún sé orðin leið á starfinu] við mig  
 she has often told that<sub>ACC</sub> that she is become tired of job.the to me

It is not among the objectives of my analysis to determine the cause of the contrast between mandatory and optional clausal extraposition in Icelandic. But my point here is that this contrast is to a great extent trivial in our discussion on clausal nominalization, primarily due to the fact that the distribution of *það* usually remains unchanged independently of whether clausal extraposition occurs or not. This can be observed in the examples of mandatory extraposition in (22), but also in the following, where extraposition is optional:

- (23) a. Ég veitti [\*(**því**) að Jón var að gráta] enga athygli  
 I gave that<sub>DAT</sub> that Jón was to cry no attention  
 'I paid no attention to the fact that Jón was crying.' (Garofalo 2020:18)
- b. Ég veitti \*(**því**) enga athygli [að Jón var að gráta]  
 I gave that<sub>DAT</sub> no attention that Jón was to cry  
 'I paid no attention to the fact that Jón was crying.'<sup>10</sup>
- (24) a. Ég gerði ráð fyrir [(**því**) að hann væri búinn að skila skýrslunni] eftir  
 I assumed that<sub>DAT</sub> that he was finished to submit report.the after  
 fundinn okkar í gærkvöldi  
 meeting.the our last night  
 'I assumed that he had submitted the report after our meeting last night.'
- b. Ég gerði ráð fyrir (**því**) eftir fundinn okkar í gærkvöldi [að hann væri búinn  
 I assumed that<sub>DAT</sub> after meeting.the our last night that he was finished  
 að skila skýrslunni]  
 to submit report.the  
 'I assumed after our meeting last night that he had submitted the report.'

<sup>10</sup>The data from oral interviews indicate that clausal extraposition of dative indirect objects is allowed, see Appendix B.

Considering this data, there is little need to specifically test examples of extraposed and non-extraposed clauses in our empirical analysis of NomCIs and non-NomCIs (see Chapter 3), as the distribution of *það* is normally insensitive to clausal extraposition. This simplifies and reduces the number of examples we need to test.

However, it is also true that there are exceptions to this trend. An important case is one of postcopular clauses, which are characterized by an optional CN in the default word order. As we can see in the following examples, it appears that a NomCI introduced by a pronoun like *sú* ‘that (feminine)’ cannot be topicalized unless the CP is extraposed. Note that the pronoun in the first position cannot be dropped, contrary to (25a), due to the fact that the first position must be occupied by an item:

- (25) a. **Afleiðingin** er [(**sú**) að við skuldum meira núna]  
 consequence.the<sub>NOM-F-SG</sub> is that<sub>NOM-F-SG</sub> that we owe more now  
 ‘The consequence is that we owe more money now.’ (Garofalo 2020:9)
- b. \*[**Sú** að við skuldum meira núna] er afleiðingin  
 that<sub>NOM-F-SG</sub> that we owe more now is consequence.the<sub>NOM-F-SG</sub>
- c. \*(**Sú**) er afleiðingin [að við skuldum meira núna]  
 that<sub>NOM-F-SG</sub> is consequence.the<sub>NOM-F-SG</sub> that we owe more now

At first glance, these facts suggest that clausal extraposition is mandatory with post-copular clauses, as the CN and CP are unable to emerge together in the first position. I will discuss this exceptional instance in more detail in §4.3.9, proposing that it is an effect of a structural issue related specifically to NomCIs introduced by *sá* ‘that (masculine)’ and *sú*.

## 2.3 The main puzzle of clausal nominalization

### 2.3.1 Subjects

Now that we have taken a look at the main properties of clausal nominalization in Icelandic and discussed the nature of *það*, let us focus on the main issues in its distribution.<sup>11</sup> Figuring out why the distribution of clausal nominalization is the way it is is a prerequisite for formulating a valid hypothesis about the role of *það* in clausal nominalization.

The first issue we need to tackle about the distribution of *það* is in sentential subjects. When *það* and its associate clausal subject are in a clause-initial position, clausal nominalization is optional, but when they surface immediately after the finite verb, *það* becomes mandatory (this also occurs in Mainland Scandinavian languages, see e.g. Josefsson (2006:footnote 12) for Swedish; Lohndal (2017) for Norwegian; for Icelandic cf. Garofalo (2020; 2023); see §4.5.2 as well):

<sup>11</sup>In this section, we are not going to discuss clausal postcopular clauses nor clausal complements of nouns as the distribution of *það* in these positions is quite regular. Postcopular clauses are optionally preceded by *það* (although, as we will see in §3.3.3, it is better to drop the pronoun when CN is in its neuter singular form whereas it is better to nominalize the clause if CN is assigned feminine, masculine or plural in accordance with the subject), while complements of nouns are obligatorily preceded by *það* (see §3.3.5).

- (26) a. (**Það**) að læra á hljóðfæri er eins og að læra að hjóla  
 that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
 ‘Learning to play an instrument is like learning to ride a bike.’ (Garofalo 2023:106)
- b. (**Það**) að hann sé farinn sannar að við vitum ekki enn þá allan sannleikann  
 that<sub>NOM</sub> that he is gone proves that we know not yet all truth.the  
 ‘The fact that he left proves that we do not know the whole truth yet.’
- (27) a. Er **\*(það)** að læra á hljóðfæri eins og að læra að hjóla?  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 ‘Is learning to play an instrument like learning to ride a bike?’ (Garofalo 2023:106)
- b. Sannar **\*(það)** að hann sé farinn að við vitum ekki enn þá allan sannleikann?  
 proves that<sub>NOM</sub> that he is gone that we know not yet all truth.the  
 ‘Does the fact that he left prove that we do not know the whole truth yet?’

From a cross-linguistic perspective, the data from Icelandic are puzzling because non-NomCl subjects are able to emerge, which does not happen in other languages such as Modern Greek, Persian, and Russian (see Roussou (1991); Hartman (2012); Knyazev (2016); Jahromi (2011); also Takahashi (2010) and Farudi (2007)), where nominative clausal subjects must be preceded by a CN (see §4.5.4 and §4.5.5 for further discussion on Persian and Russian):

- (28) a. **\*(To)** oti lei psemata apodhiknii tin enohi tis  
 the<sub>NOM</sub> that tells lies proves the<sub>ACC</sub> guilt her  
 ‘The fact that she tells lies proves her guilt.’ (Roussou 1991:87) M. GREEK
- b. **\*(In)** ke Maryam raft ma’alum e  
 this<sub>NOM</sub> that Maryam left clear is  
 ‘It is clear that Maryam left.’ (Hartman 2012:36) PERSIAN
- c. **\*(To)**, čto on èto skazal, dokazyvaet ego nevinovnost  
 that<sub>NOM</sub> that he<sub>NOM</sub> this<sub>ACC</sub> said proves his innocence<sub>ACC</sub>  
 ‘The fact that he said this proves his innocence.’ RUSSIAN  
 (adapted from Knyazev (2016:23))

What we need to figure out about this puzzle is why non-NomCl subjects are allowed in a clause-initial position in sentences like (26), as nominative clausal subjects after the finite verb are in line with the examples from other languages as *það* is obligatory.<sup>12</sup>

However, there is another important problem related to subjects, in particular the ones surfacing after the finite verb, as in (27). We said earlier that extraposition does not normally affect the distribution of NomCl subjects. Considering the fact that nominative clausal subjects in (27) are preceded by mandatory pronouns, we would expect *það* to be systematically mandatory when extraposition occurs. But there are certain exceptions to this trend, as shown below (this issue will be discussed further in §2.5.2 and §4.3.3):

<sup>12</sup>In §3.3.1.2, we will also see that clausal subjects in lexical case (I observed genitive case in particular) are mandatory even if they surface before the finite verb. Therefore, this puzzling contrast only exists with nominative subjects.

- (29) a. Er [**(það)** að læra á hljóðfæri] eins og að læra að hjóla?  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 'Is learning to play an instrument like learning to go ride a bike?' (Garofalo 2023:106)
- b. Er (**það**) hörmulegt að Jón skuli hafa barið Maríu?  
 is it deplorable that Jón shall have beaten María  
 'Is it deplorable that Jón has beaten María?' (Thráinsson 1979:192, 194)

(29b) is problematic due to the fact that *það* is optional after the finite verb. At least at the surface, the NomCI variant appears to be an instance of clausal extraposition.<sup>13</sup> So, the question is why this construction is characterized by pronominal optionality. This constitutes an exception that we need to take into account in order to explain how clausal nominalization functions in subjects.

### 2.3.2 Clausal objects

Another interesting, albeit puzzling, characteristic of Icelandic is that accusative clausal direct objects are systematically allowed to be nominalized, independently of the matrix verb involved that selects the clausal argument. I am not aware of any exception to this rule. Here are some examples:

- (30) a. Ég harma (**það**) að ég sé ekki búinn að senda umsóknina  
 I regret that<sub>ACC</sub> that I am not finished to send application.the  
 'I regret the fact that I did not send my application.'
- b. Karl samþykkir (**það**) að einhver annar stjórni fundinum  
 Karl accepts that<sub>ACC</sub> that someone else leads meeting.the  
 'Karl accepts the fact that someone else will lead the meeting.'
- c. Ég held (**það**) að góður liðsandi sé mikilvægur fyrir fótboltið  
 I think that<sub>ACC</sub> that good team.spirit is important for football.team  
 'I think a good team spirit is important for a football team.'
- d. Þeir vona (**það**) að liðið sigri  
 they hope that<sub>ACC</sub> that team.the wins  
 'They hope that the team wins.'
- e. Umræðan endurspeglar (**það**) að löggin eru ekki almenn lög  
 discussion.the reflects that<sub>ACC</sub> that laws.the are not general laws  
 'The discussion reflects the fact that the laws are not general laws.'

The same trend can be observed in nominative objects as well, as shown in the following examples (these have also been tested in questionnaires, see Appendix B):

<sup>13</sup>I will discuss these structures in §2.5.2 in the current chapter and will argue that optionality in (29b) is determined by the fact that the clausal complement can also be merged as a CP complement of the adjective (see e.g. Roussou (1991)). Within this configuration, if we look at the non-NomCI variant, the CP is not moved from its position, which means that extraposition has not taken place. However, consequently, it also means that the subject position needs to be filled by a covert expletive.

- (31) a. Þeim leiddist (**það**) að ég væri alltaf að stoppa og taka myndir  
 them<sub>DAT</sub> were upset that<sub>NOM</sub> that I was always to stop and take pictures  
 ‘They were upset about the fact that I was always stopping and taking pictures.’  
 (Garofalo 2023:111)
- b. Mér líkar (**það**) að bókin sé myndskreytt  
 me<sub>DAT</sub> likes that<sub>NOM</sub> that book.the is illustrated  
 ‘I like the fact that the book is illustrated.’ (Garofalo 2023:111)

At the surface, this regularity in nominative and accusative clausal complements should not constitute an issue, but it does in relation to previous analyses in the linguistic literature on languages like English (see e.g. Kiparsky and Kiparsky (1971), Alrenga (2005), Takahashi (2010), Kastner (2015)), Hungarian (see De Cuba and Ürögdi (2010)) and even Icelandic itself (Thráinsson (1979)). These analyses propose that the meaning conveyed by a matrix verb (e.g. factive vs. non-factive, referential vs. non-referential) has an impact on the structure of the clausal complement selected (these proposals will be discussed in much more detail in §2.5), so that matrix verbs that yield a certain meaning (true factives, for instance) would require a NomCl complement because they need an argument with a more complex structure, while other verbs of a different category (like non-factives) would opt for a non-NomCl, which has a simpler structure. The question here is why Icelandic neutralizes all these distinctions, unlike other languages. It is true, however, that some differences among verbs selecting clausal complements have been observed in Icelandic, but only when dative case is assigned. Thráinsson (1979) showed, in line with the hypothesis that the meaning of the matrix verb influences the structure of the clausal complement, that non-factives like *spá* ‘predict’ can select non-NomCls, while true factives like *fagna* ‘rejoice’ must be followed by a NomCl (see §2.5.2). As for genitive clausal objects, the pronoun is mandatory, see (32c) (the judgments in the examples below are from Thráinsson (1979)).<sup>14</sup>

- (32) a. Jón spáði (**því**) að María væri farin  
 Jón predicted that<sub>DAT</sub> that María was gone  
 ‘Jón predicted that María would have left.’ (Thráinsson 1979:216)
- b. Ég fagna \*(**því**) að þú skulir vera kominn  
 I rejoice that<sub>DAT</sub> that you shall be come  
 ‘I am happy that you came.’ (Thráinsson 1979:230)
- c. Ég sakna \*(**þess**) að María skuli ekki vera hér  
 I miss that<sub>GEN</sub> that María shall not be here  
 ‘I miss that María is not here.’ (Thráinsson 1979:230)

If we consider Thráinsson’s judgment, the question is why there is a difference between nominative/accusative, dative and genitive clausal objects. Thráinsson (1979:236) suggested that dative as well as genitive case need to be expressed morphologically through *það*, which would entail pronominal obligatoriness. But even in the hypothesis that lexical case assignment triggers the

<sup>14</sup>Unexpectedly, however, naturalistic data and questionnaires suggest in relation to *spá* that the pronoun is actually preferable, differently from Thráinsson’s judgment (see §3.3.2.3). For the moment, I will just present his point of view, but in Chapter 3 I will discuss the relevant empirical data that contrast with his analysis.

contrast between (30) and (32), then it is not clear why *spá* selects, based on Thráinsson’s judgment, NomCIs optionally if it assigns dative case. Factivity by itself cannot explain this property of *spá*, especially if we consider the fact that there are instances of true factives where *það* is optional (see e.g. *harma* ‘regret’ in the examples in (30); see further discussion in §2.5.2 with other examples that contradict the claim that factivity triggers clausal nominalization).<sup>15</sup>

### 2.3.3 Complements of prepositions

Among clausal complements of prepositions, the distribution of NomCIs and non-NomCIs is more complex than in clausal direct objects. Prepositions in Icelandic are able to assign accusative, dative and genitive case. Certain prepositions that only assign accusative case like *um* ‘about’ tend to be optionally followed by a NomCI, but there are other prepositions that tend to require the presence of *það* like the complex preposition *í kringum* ‘around’.<sup>16</sup> As for those prepositions that can assign accusative case but also dative, like *í* ‘in’, they are optionally followed by a NomCI, regardless of case (I will discuss this in more detail in §3.3.4).<sup>17</sup> If we take into account the fact that accusative clausal direct objects as in (30) are systematically followed by an optional *það*, accusative complements of prepositions also follow this trend, except with prepositions like *í kringum*. It is possible that such complex prepositions like *í kringum* are generally followed by a NomCI, but the question is whether this can also be confirmed by other similar prepositions that assign dative or genitive case.<sup>18</sup>

- (33) a. Hann er mjög meðvitaður um (**það**) hvaða áhrif notkun snjallsíma hefur á  
 he is very aware of that<sub>ACC</sub> which impact use smartphones has on  
 börn  
 children  
 ‘He is very aware of the influence the use of smartphones has on children.’
- b. Það var mikið ferli í kringum \*(**það**) að byggja þetta hótél  
 there was great process around that<sub>ACC</sub> to build this hotel  
 ‘There was a great process around the construction of this hotel.’

<sup>15</sup>In §3.3.2.3, we will also see that *það* is not exactly as optional as Thráinsson claims it is.

<sup>16</sup>The example on *í kringum* has been collected from RMH. My judgment here is based on the high frequency of *það* that resulted from the data from corpora (see Appendix A) as well as the data from questionnaires, which suggest that the NomCI variant is significantly more acceptable than the non-NomCI variant (see Appendix B).

<sup>17</sup>The empirical data in §3.3.4 will show in particular that, although double case prepositions allow optional nominalization, a CN is preferable in instances of dative case assignment, while it is much less common when accusative case is assigned.

<sup>18</sup>Note that *í kringum* is not the only exception to the rule, as other only-accusative prepositions like *umfram* ‘above, beyond’ also seem to need *það*:

- (i) Það munar miklu að taka við slíku fjósi umfram \*(**það**) að byggja nýtt frá grunni  
 it differs much to receive such cowshed beyond that<sub>ACC</sub> to build new from scratch  
 ‘It makes a big difference to receive such a cowshed in addition to the construction of a new one from scratch.’ (from *Bændablaðið*)

- c. Hún er ekki tilbúin í (**það**) að eignast barn svona ung  
 she is not ready in that<sub>ACC</sub> to have child so young  
 'She is not ready to have a child at such a young age.'

As for dative case, we can find a similar contrast between prepositions. Those that assign dative case only, like *hjá* 'by', need to be followed by a NomCl, while those that also assign accusative case, like *yfir* 'over', accept *það* optionally. This difference is similar to the one between dative clausal direct objects selected by *fagna* and *spá* based on Thráinsson's judgment:

- (34) a. Enginn kemst hjá \*(**því**) að kaupa sér mat  
 nobody avoids that<sub>DAT</sub> to buy oneself food  
 'No one can avoid buying food.'
- b. Allardyce er svekktur yfir (**því**) að hafa fengið sparkið  
 Allardyce is annoyed over that<sub>DAT</sub> to have got kick.the  
 'Allardyce was annoyed about the fact that he was fired.' (adapted from *ftbolti.net*)

Regarding genitive case, the most common preposition introducing clausal complements is *til* 'to', which is optionally followed by *það*. Other prepositions like *utan* 'apart from' tend to need clausal nominalization:<sup>19</sup>

- (35) a. Ég ætlast ekki til (**þess**) að Bjarna sé sagt upp  
 I expect not to that<sub>GEN</sub> that Bjarni is fired  
 'I don't expect Bjarni to be fired.'
- b. Mig langar til (**þess**) að þakka þeim sem sendu umsagnir  
 me wants to that<sub>GEN</sub> to thank those that sent comments  
 'I want to thank those who sent comments.'
- c. Gögn þátttakenda eru ekki aðgengileg utan \*(**þess**) að aldur þeirra er  
 data participants are not accessible apart from that<sub>GEN</sub> that age their is  
 birtur á vefsíðunni  
 published on website.the  
 'Data on participants are not accessible except for their age being published on the website.'

In other words, independently of the case assigned, certain prepositions allow nominalization while some other prepositions require it. At first glance, one might propose that this contrast is caused by the fact that those prepositions that are selected by verbs tend to allow nominalization, while prepositions that emerge in an adjunct position, like *í kringum* or *utan*, require NomCl (see further discussion on this distinction in §3.3.4). However, the matter is not that simple, as there are prepositions like *hjá* which are selected by verbs but still require NomCl. Thus, there could be more than one factor at play in the distribution of NomCl and non-NomCl in PPs which need to be defined (these factors will be discussed in §3.3.4 and also §4.4.5).

<sup>19</sup>The judgment on *utan* is based on the results from questionnaires as well as the ones from RMH on *that*-clauses and infinitives that are introduced by this preposition (see §3.3.4.6).

### 2.3.4 Complements of adjectives

In addition, among adjectives, one can find a contrast between those that require clausal nominalization and those that allow it. Here is an example with the adjectives *feginn* ‘happy, satisfied’, *háður* ‘dependent’, and also *sammála* ‘agreeing’, which assign dative case to their clausal complements. As we can see, *það* can be optional or mandatory depending on the relevant adjective.<sup>20</sup>

- (36) a. *Ég er feginn (því) að þú skulir vera kominn*  
I am satisfied that<sub>DAT</sub> that you shall be arrived  
‘I’m happy that you have come.’
- b. *Sum hagkerfi heimsins eru háð \*(því) að fá orku úr jarðgasi*  
some economic systems world.the<sub>GEN</sub> are dependent that<sub>DAT</sub> to get energy from natural gas  
‘Some economic systems depend on getting energy from natural gas.’
- c. *Flestir eru sammála (því) að framleiðslan er of mikil*  
most are agreeing that<sub>DAT</sub> that production.the is too much  
‘Most people agree that the production is too much.’ (adapted from *Bleikt.is*)

Adjectives can also assign genitive case and, rarely, accusative case. Here are some examples with the adjective *minnugur* ‘mindful’, which assigns genitive case, and *viðstaddur* ‘present’, which assigns accusative case. In both instances, clausal nominalization is mandatory:<sup>21</sup>

- (37) a. *KR-ingar voru minnugir \*(þess) að tapa 2-0 fyrir Völsurum*  
KR-players were mindful that<sub>GEN</sub> to lose 2-0 against Valur-players  
‘KR players were mindful of having lost 2-0 against Valur.’
- b. *Ég sagðist ekki vilja vera viðstödd \*(það) að hann væri sprautaður*  
I said not want be present that<sub>ACC</sub> that he was injected  
‘I said I didn’t want to be present at his injection.’

Just by looking at all these examples, it seems that optional nominalization with dative-assigning adjectives is the exception to the rule, since the remaining adjectives observed here trigger clausal nominalization obligatorily independently of the case assigned. Hence, the question is what causes this different distribution of *það* among dative-assigning adjectives, i.e. whether the main factor is, for example, the subcategorization pattern of the adjective involved or whether the meaning of the adjective itself plays a role in this contrast. This issue will be discussed in particular in §3.3.6.

<sup>20</sup>Unlike English, where an adjective is generally followed by a preposition when it takes a complement, in Icelandic, adjectives can assign case directly without the help of a preposition.

<sup>21</sup>The data on *viðstaddur*, as we will see in Chapter 3, were not so satisfying as both variants were degraded (the criteria on general acceptability of a certain example are described in §3.2.5), but the NomCl variant was still better than the non-NomCl variant. RMH, unfortunately, only provided one example, which contains a NomCl.

### 2.3.5 Clausal indirect objects

Clausal indirect objects are more rare than clausal complements in other syntactic positions (cf. Garofalo (2020; 2023)). However, they exhibit interesting characteristics. Indirect objects in Icelandic can either surface in accusative or dative case, as shown in the following list of possible patterns (see also Jónsson (2005b:404–405)):

- (38) dative+accusative, e.g. *gefa* ‘give’  
 dative+dative, e.g. *lofa* ‘promise’  
 dative+genitive, e.g. *óska* ‘wish’  
 accusative+dative, e.g. *svipta* ‘confiscate, deprive’  
 accusative+genitive, e.g. *spyrja* ‘ask’  
 accusative+accusative (very rare), e.g. *kosta* ‘cost’

Clausal indirect objects need to surface as NomCIs independently of case, as shown in the examples below with the verb *veita* ‘provide’ for indirect objects in dative case and *svipta* ‘deprive’ for indirect objects in accusative case (also cf. Garofalo (2020; 2023)).<sup>22</sup> The most important and puzzling aspect of clausal indirect objects is that, in active voice, both dative and accusative CNs are mandatory. This is unexpected if we consider, for instance, accusative clausal direct objects, which are generally characterized by optional nominalization. In passive voice, however, the accusative clausal indirect object promoted to subject can be a non-NomCI in a clause-initial position.<sup>23</sup>

- (39) a. *Ég veitti [\*(því) að Jón var að gráta] enga athygli*  
 I gave that<sub>DAT</sub> that Jón was to cry no<sub>ACC</sub> attention<sub>ACC</sub>  
 ‘I paid no attention to the fact that Jón was crying.’ (Garofalo 2020:18)
- b. *[(því) að Jón var að gráta] var engin athygli veitt*  
 that<sub>DAT</sub> that Jón was to cry was no<sub>NOM</sub> attention<sub>NOM</sub> paid  
 ‘No attention was paid to the fact that Jón was crying.’ (Garofalo 2020:25)
- (40) a. *Vopnahléið svipti [\*(það) að hermenn fórnúðu lífi sínu] öllum hetjuljóma*  
 armistice.the deprived that<sub>ACC</sub> that soldiers sacrificed life their all<sub>DAT</sub> heroism<sub>DAT</sub>  
 ‘The armistice deprived the soldiers’ sacrifice of their lives of all heroism.’
- b. *[(Það) að hermenn fórnúðu lífi sínu] var svipt öllum hetjuljóma*  
 that<sub>NOM</sub> that soldiers sacrificed life their was deprived all<sub>DAT</sub> heroism<sub>DAT</sub>  
 ‘The fact that soldiers sacrificed their lives was deprived of all heroism.’  
 (Garofalo 2023:113)

<sup>22</sup>From a cross-linguistic perspective, pronominal obligatoriness is attested, for instance, in the *bokmål* variant of Norwegian, where the CN *det* ‘that’ needs to surface before a clausal indirect object (see Faarlund (2019:143)):

- (i) *De gav \*(det) at boligprisene steg, skylden for bankkrisen*  
 the gave that that housing.prices.the rose blame.the for bank.crisis.the  
 ‘They blamed the bank crisis on the rising housing prices.’

NORWEGIAN BOKMÅL

<sup>23</sup>These judgments have been confirmed in oral interviews with a group of native speakers of Icelandic and, to some extent, also in questionnaires (see §3.3.7; also Garofalo (2023)).

| Position                   | Status of <i>það</i> | Remarks                                 |
|----------------------------|----------------------|-----------------------------------------|
| Subjects 1st position      | OPT                  | -                                       |
| Subjects after finite verb | OBL                  | OPT with extraposition                  |
| Direct objects NOM/ACC     | OPT                  | Independently of the verb involved      |
| Direct objects DAT/GEN     | OBL/OPT              | Apparently dependent on the verb        |
| Prepositional complements  | OBL/OPT              | Apparently dependent on the preposition |
| Complements of adjectives  | OBL/OPT              | Apparently dependent on the adjective   |
| Clausal indirect objects   | OBL                  | OPT with passivized ACC objects         |

Table 2.3: Summary - Clausal nominalization puzzle

The contrast in (40) is very similar to the one observed with nominative clausal subjects, where *það* is optional if the clause surfaces before the finite matrix verb. The question we need to answer is why this is also allowed with passivized accusative indirect objects while the ones in active voice require nominalization (unlike accusative direct objects, where non-NomCIs are allowed in active voice, see (30)).

Before we move to the next section, I summarized in Table 2.3, for convenience, the puzzle of the distribution of *það* we have discussed in this section (OPT stands for optional, OBL for obligatory).

## 2.4 Structure of clausal arguments with and without *það*

### 2.4.1 Introduction

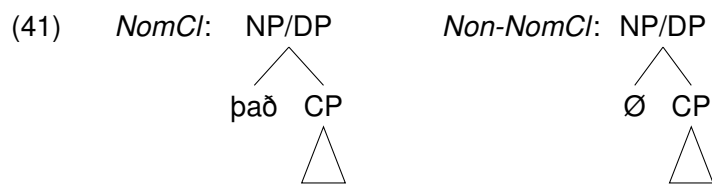
Now that we have taken a look at the main properties of clausal nominalization as well as its main distributional puzzle, let us discuss NomCIs and non-NomCIs from a structural perspective, in light of the linguistic literature on the topic. Various structural models have been proposed to analyze both clausal types. The main question we need to answer in this regard is whether their structure is the same or not, and most importantly, what the structure is in both cases. I propose in this subsection that, in Icelandic, NomCIs surface as DPs while non-NomCIs surface as CPs.<sup>24</sup> I assume in this structural analysis that *það* and its associate clause form together one constituent (cf. Garofalo (2020)). I am aware that certain accounts like Ott and de Vries (2014) and Ott (2014) challenge this view, proposing that the clausal complement must be extra-sentential, as it would otherwise compete with *það* for the same  $\theta$ -role. However, the fact that both items can occupy the same position together in Icelandic (e.g. in clause-initial position or indirect object position) does not support this alternative analysis.<sup>25</sup>

<sup>24</sup>We will see in Chapter 4, however, that non-NomCIs that we find in DP positions are not merged as CPs in Icelandic, but rather as DPs that undergo a process of structural removal. For the moment, in this section we will focus on the distribution of NomCIs and non-NomCIs based on how they appear after all movements have taken place. But, in Chapter 4, we will also discuss the derivation of both clausal variants.

<sup>25</sup>The only exception in this regard could be post-copular clauses, as they are characterized by mandatory extraposition (see §2.2.4).

## 2.4.2 Generalized DP/NP hypothesis

Let us look at the first structural model of our discussion on clausal nominalization, which I call the *Generalized DP hypothesis* (henceforth GDPH; alternatively, Generalized NP hypothesis based on previous approaches). It can be found in accounts like Thráinsson (1979) for Icelandic and also in cross-linguistic analyses like e.g. Knyazev (2016) for Russian:



Let us focus on Thráinsson's proposal since it analyzes Icelandic. His model, unlike many other accounts mentioned in this chapter which are generally framed in the *Minimalist Program* (see Chomsky (1993; 1995 and subsequent work); also §4.2.1), is built upon *Transformational Generative Grammar* (TGG; see Chomsky (1957; 1965)), which posits a much simpler syntactic structure compared to more recent theoretical frameworks. Thráinsson's proposal is based on the idea that an argument position consists of an NP slot (I use here his terminology) occupied by an item. When the clause is nominalized, both *það* and the CP surface inside the NP slot, while only the CP is embedded when the clause is not nominalized. Note that, within TGG, an NP is allowed to immediately dominate a CP by unary branching, which is excluded in recent frameworks. Nevertheless, let us at least consider the arguments that support Thráinsson's proposal. Thráinsson observed that clausal arguments behave like NPs in Icelandic as they undergo the same syntactic transformations. These similarities can be tested, for instance, via different diagnostic tests. One of them is left dislocation (see in particular Thráinsson (1979:69)):

- (42) a. **[Friðrik]** er langbestur  
 Friðrik is best  
 'Friðrik is the best.' (Thráinsson 1979:61)
- b. **[Friðrik]<sub>i</sub>, hann<sub>i</sub>** er langbestur  
 Friðrik he is best  
 'Friðrik, he is the best.' (Thráinsson 1979:61)
- (43) a. **[Að Jón tali sjálfur við kennarann]** er langbestur  
 that Jón talks himself to teacher.the is best  
 'It is best that Jón talk to the teacher himself.' (Thráinsson 1979:31)
- b. **[Að Jón tali sjálfur við kennarann]<sub>i</sub>, það<sub>i</sub>** er langbestur  
 that Jón talks himself to teacher.the it is best  
 'That Jón talks to the teacher himself, that is best.'

The argument proposed by Thráinsson is that the left-dislocated position is actually an NP position. In fact, the constituent [Friðrik] in (42), which is clearly a nominal category, is co-referential with the pronoun *hann* 'he'. Now, the fact that CPs can also be moved to that position and that the

pronoun *það* is placed as an argument of the predicate in the matrix clause (see (43)) suggests that they are also co-referential and that the embedded clause is an argument originally placed in an NP slot.<sup>26</sup>

Another piece of evidence Thráinsson presented for the claim that clausal complements have the same distribution as NPs and should be placed in an NP slot is that they can be coordinated with fully-fledged NPs (Thráinsson 1979:95):<sup>27</sup>

- (44) a. Jón spurði [**þessarar spurningar**] og [**hvort María væri farin**]  
 Jón asked this question and whether María were gone  
 ‘Jón asked this question and whether María was gone.’  
 b. Þeir töluðu um [**veðrið**] og [**að þessi vetur væri óvenjulega harður**]  
 they talked about weather.the and that this winter were unusually hard  
 ‘They talked about the weather and that this winter was unusually hard.’

Thráinsson also used passivization as a diagnostic test in order to claim that clausal complements are NPs. Here is an example where a clausal object is passivized and turns into a subject. In Thráinsson’s view, passivization of the clausal object here indicates that it occupies an NP slot:<sup>28</sup>

- (45) a. Fjölskyldan ákvað [**að María skyldi heimsækja Ólaf**]  
 family.the decided that María should visit Ólafur  
 ‘The family decided that María will visit Ólafur.’  
 b. [**Að María skyldi heimsækja Ólaf**] var ákveðið  
 that María should visit Ólafur was decided  
 ‘It was decided that María will visit Ólafur.’ (Thráinsson 1979:90)

Provided that clausal complements are positioned in an NP slot, let us now focus more on *það*

<sup>26</sup>It is important to clarify here that left-dislocation is different from topicalization in Icelandic syntax. As commonly known, topicalization in a V2 language entails that if a constituent other than the subject is moved to the position before the finite verb, the word order is such that the subject surfaces immediately after the finite verb, as shown in the following example:

- (i) [**Á morgun**]; ætla ég að fara í bíó  $t_i$   
 tomorrow intend I to go to cinema  
 ‘Tomorrow I will go to the cinema.’

In (43), by contrast, the word order is the default word order, which means that left-dislocation, which also might require a prosodic interruption immediately after the dislocated element, is different from topicalization, at least in Icelandic (cf. e.g. Þráinsson et al. (2005:590); Jónsson (2005a:443ff.)).

<sup>27</sup>It is also true, however, that there are instances where two different types of syntactic phrases can be coordinated if they are semantically comparable, as in the following example:

- (i) Áhrif tóbaks fara [**AdvP hægt**] og [**PP án þess að við veitum þeim athygli**]  
 impact tobacco go slowly and without that<sub>GEN</sub> that we provide them attention  
 ‘The impact of tobacco comes slowly and without us paying attention to it.’

The two coordinate phrases in this example differ from a structural perspective, but both express how tobacco affects somebody, so that they are comparable from a semantic perspective and can be coordinated.

<sup>28</sup>This resonates with Takahashi (2010), who claimed that a clausal complement can only be moved if it surfaces in a position in which a DP is allowed to appear.

before clauses. Thráinsson showed a contrast between predicates like *vera líklegt* ‘be likely’, which is a non-factive, and *vera hörmulegt* ‘be deplorable’, which is an emotive factive. When *það* is clause-initial and the clausal subject is extraposed, there is no apparent difference between predicates (see (46)). But, in Thráinsson’s view, *það* is problematic when it surfaces after the finite verb with non-factives if the sentence is uttered out of the blue, although it can become grammatical if it refers to a previous context, e.g. the content of the extraposed clausal subject has already been mentioned (the following judgments are from Thráinsson):<sup>29</sup>

- (46) a. **Það** er líklegt að Jón hafi barið Maríu  
it is likely that Jón has beaten María  
‘It is likely that Jón has beaten María.’ (Thráinsson 1979:184)
- b. **Það** er hörmulegt að Jón skuli hafa barið Maríu  
it is deplorable that Jón shall have beaten María  
‘It is deplorable that Jón has beaten María.’ (Thráinsson 1979:191)
- (47) a. Er (?\***það**) líklegt að Jón hafi barið Maríu?  
is it likely that Jón has beaten María  
‘Is it likely that Jón has beaten María?’ (Thráinsson 1979:183–184)
- b. Er (**það**) hörmulegt að Jón skuli hafa barið Maríu?  
is it deplorable that Jón shall have beaten María  
‘Is it deplorable that Jón has beaten María?’ (Thráinsson 1979:192, 194)

Based on the fact that *það* is degraded in (47a) unless it surfaces as a discourse anaphor (see Thráinsson (1979:183)) while it is grammatical in (47b), Thráinsson suggested that *það* is base-generated inside the NP slot when the matrix verb is an emotive factive before clausal extraposition takes place. On the other hand, *það* can be inserted into the NP slot after extraposition of the clausal argument when the matrix verb is non-factive.<sup>30</sup> But, in that case, *það* can only be a discourse anaphor, and not an expletive, as expletives only occupy clause-initial position in Icelandic (also cf. Práinsson et al. (2005:338)).

Thráinsson’s analysis, in the context of clausal nominalization, tells us that all clausal complements are NPs independently of whether *það* is absent, base-generated or inserted, due to the fact that they all surface embedded in argument slots, which are NPs. If we assume for a moment that Thráinsson’s analysis is valid, then the only difference between NomCIs and non-NomCIs consists in the presence or absence of *það* inside the NP slot, which, based on our observations so far across syntactic positions, would be more likely to be determined by base-generation rather than by insertion, considering the fact that many instances of clausal nominalization take place without *það* being discourse-anaphoric.

Thráinsson’s insights have also been supported by other cross-linguistic analyses like Alrenga (2005) and, in particular, Takahashi (2010), who proposed that a clausal complement can only be moved if it is merged in a position in which a DP is allowed to appear:

<sup>29</sup>The pronoun *það* is defined as *it* in the examples below in accordance with Thráinsson’s analysis. But, as explained earlier, *það* as CN is a demonstrative pronoun rather than a personal pronoun.

<sup>30</sup>Thráinsson followed here Rosenbaum (1967) with the idea that extraposition is a rightward movement.

(48) *Moved Clausal Complement Generalization* (Takahashi (2010); henceforth MCGG)

A clausal complement is allowed to move only if its base-generated position is one in which a DP is allowed to appear

Verbs like *capture* or *attribute* in English support this hypothesis, as they take a DP complement that can be passivized or topicalized (the examples are adapted from Alrenga (2005:184, 192) and Takahashi (2010:347); see further discussion on this topic in §2.5.4). Note that, if the complement remains *in situ*, it must emerge as a complex DP introduced by *the fact*, but this is not necessary after movement:

- (49) a. This formulation of the rule **expresses/captures/reflects/brings out** \*(the fact) that these nouns behave differently  
b. [That these nouns behave differently] **is captured** by this formulation of the rule  
c. [That these consonants behave exceptionally], we can **attribute** to the fact that they are coronals

Takahashi proposed that a moved clausal complement in English is a DP even if there is no overt determiner surfacing, as it is originally merged in a DP position, in line with GDPH. This hypothesis is also confirmed by a contrast between verbs like *capture* and verbs like *hope* or *reason*, as the latter can only subcategorize for CPs and not DPs and, consequently, their clausal complement is unable to move via passivization or topicalization (the examples are adapted from Alrenga (2005:183) and Takahashi (2010:347)):

- (50) a. Most baseball fans **hoped/felt/wished/insisted/reasoned** that the Giants would win the World Series  
b. \*[That the Giants would win the World Series] **was reasoned** by most baseball fans  
c. \*[That the Giants would probably win the World series], most baseball fans **reasoned**

Among other cross-linguistic accounts, Knyazev (2016) also came to a similar conclusion for DP positions in Russian.<sup>31</sup> On the basis of the account in Pesetsky and Torrego (2004; 2011), complement clauses are unable to have case, as they can appear in positions in which no case is assigned. Adjectives in English, for instance, do not assign case directly to their complement, but require the presence of a preposition. However, when a CP surfaces, there is no need for a preposition at all, which suggests that CPs do not require case licensing (the examples are from Pesetsky and Torrego (2004:502)):

- (51) a. Bill was afraid \*(**of**) the storm  
b. Bill was afraid that the storm will be destructive

Knyazev, however, observed that Russian clausal direct objects (note that his analysis focuses

<sup>31</sup>Although Russian is a distant relative of Icelandic, I believe it is important to discuss some of its data here, as there is an important contrast between Russian and Icelandic related to extraction from a clausal complement that shows that Icelandic non-NomCIs and NomCIs cannot surface both as DPs, as we will see in the following pages.

specifically on *that*-clauses) actually require case licensing when the subject is an agent,<sup>32</sup> as shown in the following examples with the verb *govorit* ‘say, speak’ (Knyazev 2016:1–2). In (52), the verb can select either a direct object or a PP where the CN must surface. These two selectional options are also confirmed in (53) where the clausal complements are replaced by interrogative pronouns (see in particular the pronoun *čto* ‘what’ in accusative case for direct objects):

- (52) a. Učenyje govorjat, čto na ètoj territorii ran’she žili ljudi  
 scientists<sub>NOM</sub> say that on this territory<sub>LOC</sub> earlier lived people<sub>NOM</sub>  
 ‘Scientists say that earlier people used to live on this territory.’
- b. Učenyje govorjat o \*(**tom**), čto na ètoj territorii ran’she žili ljudi  
 scientists<sub>NOM</sub> say about that<sub>LOC</sub> that on this territory<sub>LOC</sub> earlier lived people<sub>NOM</sub>  
 ‘Scientists are talking about the fact that earlier people used to live on this territory.’
- (53) Čto / O čem govorjat učenyje?  
 what<sub>ACC</sub> / about what<sub>LOC</sub> say scientists  
 ‘What are scientists saying?’
- RUSSIAN

Considering the data in (51) as well as the insights in Takahashi (2010), Knyazev proposed that case licensing is possible in (52) due to the fact that the clausal complement is embedded into an invisible DP projection. Consequently, Knyazev showed that there are DP clauses introduced by a CN, as in (52b), and DP clauses that are not introduced by any CN, as in (52a).<sup>33</sup> As a result, clausal complements in Russian are embedded into a DP projection independently of whether a pronoun introduces them or not.<sup>34</sup>

Let us now discuss whether GDPH is applicable to Icelandic as well. Assuming that Knyazev’s analysis is correct, applying GDPH to Russian entails some interesting implications that can help us to identify the structure of NomCIs and non-NomCIs in Icelandic. Since DPs are known in the linguistic literature for blocking extraction (see e.g. Ross (1967); Davies and Dubinsky (2003); Sedrins (2015) among others), if, as Knyazev claimed, accusative clausal complements are DPs

<sup>32</sup>Knyazev (2016:2) showed that, when the subject is not an agent, the accusative complement clause is disallowed, while the PP is still possible:

- (i) a. \*Èti naxodki govorjat, čto na ètoj territorii ran’she žili ljudi  
 these findings<sub>NOM</sub> say that on this territory<sub>LOC</sub> earlier lived people<sub>NOM</sub>  
 Intended: ‘These findings indicate that earlier people used to live on this territory.’
- b. Èti naxodki govorjat o **tom**, čto na ètoj territorii ran’she žili ljudi  
 these findings<sub>NOM</sub> say about that<sub>LOC</sub> that on this territory<sub>LOC</sub> earlier lived people<sub>NOM</sub>  
 ‘These findings indicate that earlier people used to live on this territory.’
- RUSSIAN

<sup>33</sup>Knyazev extended the idea that clausal complements are DPs to subjects as well, since clausal subjects must be introduced by a CN. He explained pronominal obligatoriness with clausal subjects by following Landau (2007), who proposed that the D-head in subjects must be overt:

- (i) \*(**To**), čto on èto skazal, dokazyvaet ego nevinovnost  
 that<sub>NOM</sub> that he<sub>NOM</sub> this<sub>ACC</sub> said proves his innocence<sub>ACC</sub>  
 ‘The fact that he said this proves his innocence.’ (adapted from Knyazev (2016:23))
- RUSSIAN

<sup>34</sup>Further discussion on Russian NomCIs and non-NomCIs is offered in §2.5.5 and §4.5.5.

whether the CN surfaces or not, then we expect them to be extraction islands.<sup>35</sup> This is confirmed, for instance, in (54), where *čto*-clauses do not allow *wh*-extraction (see Dyakonova (2009:215)):<sup>36</sup>

- (54) a. ??/\***Kogo**<sub>i</sub> Olga skazala [čto oni videli t<sub>i</sub>]?  
 who<sub>ACC</sub> Olga said that they saw?  
 'Who did Olga say that they saw?'  
 b. ??/\***Kto**<sub>i</sub> Olga skazala [čto t<sub>i</sub> videl Raju]?  
 who<sub>NOM</sub> Olga said that saw Raja<sub>ACC</sub>  
 'Who did Olga say that saw Raja?' RUSSIAN

Icelandic, interestingly, differs from Russian in this regard. Thráinsson (1979:110-111, 219, 235; the examples are presented below) himself showed instances where extraction<sup>37</sup> from a clausal argument is possible in Icelandic, e.g. with infinitives and *that*-clauses, but only if *það* does not surface (similar facts based on Thráinsson's account are also presented in Wood (2012); Ott (2014); Ingason (2018); Garofalo (2020; 2023)). *Það* is also ungrammatical if the whole *that*-clause or the infinitive clause is moved leftwards, leaving the CN behind. This is shown in particular in the c-examples below:

- (55) a. Þeir ákváðu [(**það**) að heimsækja Ólaf]  
 they decided that<sub>ACC</sub> to visit Ólafur<sub>ACC</sub>  
 'They decided to visit Ólafur.'  
 b. Ólaf<sub>i</sub> ákváðu þeir [(**\*það**) að heimsækja \_<sub>i</sub>]  
 Ólafur<sub>ACC</sub> decided they that<sub>ACC</sub> to visit  
 c. [Að heimsækja Ólaf]<sub>i</sub> ákváðu þeir [(**\*það**) \_<sub>i</sub>]  
 to visit Ólafur<sub>ACC</sub> decided they that<sub>ACC</sub>

<sup>35</sup>Although Knyazev claimed that clausal complements are DPs independently of the presence of a CN and the data in Dyakonova (2009) confirm the impossibility of extraction, it is also true that various accounts in the linguistic literature on Slavic languages like Russian claim that languages that do not provide a definite article do not display DP projections either, and argument nominals are simply NPs (see in particular Bošković (2012)). This view is supported, for instance, by the fact that *Left Branch Extraction* is only allowed in languages without a definite article, like Russian (see e.g. Bošković (2005)). However, the status of argument nominals in Slavic languages is still an object of debate, as other linguists have also shown that Slavic languages present DP characteristics (consider e.g. Pereltsvaig (2007); Progovac (1998)). A detailed discussion on this debate is offered in Pereltsvaig (2013). Note, however, that the purpose here is not to determine whether DPs exist or do not exist in Russian. At the end of the day, Icelandic provides a definite article, which indicates that Icelandic can definitely license DP projections.

<sup>36</sup>Infinitives, on the other hand, allow extraction in Russian (Dyakonova 2009:216). Note that Knyazev's DP analysis focused specifically on *čto*-clauses rather than infinitives:

- (i) **Komu**<sub>i</sub> Ira nadeetsja otdat' t<sub>i</sub> kotjat?  
 who<sub>DAT</sub> Ira<sub>NOM</sub> hopes give away<sub>INF</sub> kittens<sub>ACC</sub>  
 'Who does Ira hope to give the kittens to?' RUSSIAN

<sup>37</sup>Thráinsson, assuming that NomCIs are systematically extraposed clauses, showed these examples to support the idea that extraposition causes an NP to be an extraction island (following Ross (1967)). As already discussed so far, I do not accept the assumption that nominalization is systematically correlated to clausal extraposition, as there are instances where *það* and its associate clausal complement are unambiguously non-extraposed, e.g. clausal subjects after the finite verb or clausal indirect objects.

- (56) a. Ég harma [(**það**) að Jón skuli hafa barið Maríu]  
 I regret that<sub>ACC</sub> that Jón<sub>NOM</sub> shall have beaten up María<sub>ACC</sub>  
 'I regret that Jón has beaten up María.'
- b. Maríu<sub>i</sub> harma ég [(**\*það**) að Jón skuli hafa barið \_\_\_<sub>i</sub>]  
 María<sub>ACC</sub> regret I that<sub>ACC</sub> that Jón<sub>NOM</sub> shall have beaten up
- c. [Að Jón skuli hafa barið Maríu<sub>i</sub>] harma ég (**\*það**) \_\_\_<sub>i</sub>  
 that Jón<sub>NOM</sub> shall have beaten up María<sub>ACC</sub> regret I that<sub>ACC</sub>

If we compare, in particular, *that*-clauses in Russian and Icelandic, the facts suggest that, if NomCIs can be interpreted as DPs in Icelandic with GDPH, non-NomCIs cannot, as they allow extraction. In Russian, by contrast, since *čto*-clauses do not allow extraction as a general rule, they are DPs even if the CN does not surface.

There are additional reasons for why it cannot be the case that NomCIs and non-NomCIs surface both as DPs in Icelandic. If we take into account the examples from Pesetsky and Torrego (2004) in (51) on English, Icelandic displays a similar behavior. If we consider an adjective like *ánægður* 'happy', it cannot select a DP as a complement and, therefore, it requires a preposition like *með* 'with'. Alternatively, it can select a clausal complement. This suggests that the position where the clause emerges cannot be a DP position, which is not in line with GDPH:

- (57) a. Ég er ánægður **\*(með)** nýja verkefnisstjórnann  
 I am happy with new project.manager.the<sub>ACC</sub>  
 'I'm happy with the new project manager.'
- b. Ég er ánægður að við skulum hafa fengið nýjan verkefnisstjóra  
 I am happy that we shall have got new project.manager  
 'I'm happy we got a new project manager.'

The same problem with GDPH also occurs in other syntactic positions. Consider the clausal complement of the verb *hjálpa* 'help', for instance. The examples below show that the infinitival complement, which cannot be a NomCI, cannot be replaced by a DP either (Garofalo 2023:107):

- (58) a. \*María hjálpaði honum **flutningana**  
 María helped him<sub>DAT</sub> moving out<sub>ACC</sub>  
 Intended: 'María helped him moving out.'
- b. María hjálpaði honum [(**\*það**) að flytja]  
 María helped him that<sub>ACC</sub> to move out  
 'María helped him moving out.'

Lastly, if we follow GDPH and assume that both clausal variants are DPs in Icelandic, pronominal obligatoriness<sup>38</sup> constitutes a problem for this hypothesis, as a non-NomCI, by virtue of being a DP, should be sufficient to occupy a DP position. But that is often not the case. Consider, for instance, the distribution of clausal indirect objects. GDPH can explain why *það* is optional in the first position after the clausal indirect object has been passivized (see (60b)), but it cannot offer

<sup>38</sup>Note that Thráinsson (1979) correlated pronominal obligatoriness in Icelandic to factivity. I do not agree with this hypothesis and will present some arguments against it in §2.5.2.

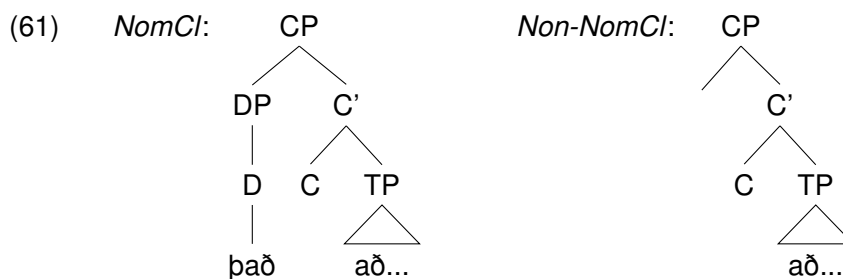
an answer as to why it is systematically mandatory *in situ* in active voice (see (59) and (60a)). We would have rather expected *það* to be optional in that position as well, but this does not happen.<sup>39</sup> Thus, we can infer that NomCIs and non-NomCIs cannot both be DPs:

- (59) Ég veitti [**(því)** að Jón var að gráta] enga athygli  
 I gave that<sub>DAT</sub> that Jón was to cry no<sub>ACC</sub> attention<sub>ACC</sub>  
 'I paid no attention to the fact that Jón was crying.' (Garofalo 2020:18)
- (60) a. Vopnahléið svipti [**(það)** að hermenn fórnðu lífi sínu] öllum hetjuljóma  
 armistice.the deprived that<sub>ACC</sub> that soldiers sacrificed life their all<sub>DAT</sub> heroism<sub>DAT</sub>  
 'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'  
 (Garofalo 2023:113)
- b. [**(Það)** að hermenn fórnðu lífi sínu] var svipt öllum hetjuljóma  
 that<sub>NOM</sub> that soldiers sacrificed life their was deprived all<sub>DAT</sub> heroism<sub>DAT</sub>  
 'The fact that soldiers sacrificed their lives was deprived of all heroism.' (Garofalo 2023:113)

In sum, if we interpret the facts from Icelandic presented here in light of GDPH, it is possible to view NomCIs as DPs, but not non-NomCIs, which appear to have a different structure.

### 2.4.3 Generalized CP hypothesis

If the previous model assumes the presence of an outer shell embedding clausal arguments independently of *það*, the hypothesis that is presented here, which I call *Generalized CP Hypothesis* (henceforth GCPH), does not assume any additional projection over a CP at all. NomCIs only differ from non-NomCIs due to the presence of the CN in the Spec,C position of the clausal complement (see e.g. Stroik (1996); Yoon (2001); Jahromi (2011) among others):



<sup>39</sup>Note that the accusative case assigned to the indirect object appears to be structural, as shown by the fact – commonly known about Icelandic – that a passivized indirect object takes up nominative case (also note the agreement in gender and number between the passivized indirect object and the past participle, which only happens when the subject is in structural case):

- (i) a. Lögreglan svipti **Mörð** ökuskírteininu  
 police.the confiscated Mörður<sub>ACC</sub> driving license.the<sub>DAT</sub>  
 'The police confiscated Mörður's driving license.'
- b. **Mörður** var sviptur ökuskírteininu  
 Mörður<sub>NOM-M-SG</sub> was confiscated<sub>NOM-M-SG</sub> driving license.the<sub>DAT</sub>  
 'Mörður's driving license was confiscated.'

Such a model has been proposed, for instance, for English (see e.g. Stroik (1996)), but also for other languages like Persian (see Jahromi (2011); see §4.5.4 for more discussion on clausal nominalization in Persian). Especially in the case of English, where *it* surfaces in extraposition constructions, the GCPH model is based on the assumption that rightward extraposition is excluded in syntax, consistently with Kayne (1994). In other words, it is the pronoun that is moved upwards in the syntactic structure (e.g. to Spec,AgrS or Spec,AgrO according to Stroik (1996)).

Icelandic, however, is incompatible with this proposal for a variety of reasons. First of all, if a CN occupies subordinate Spec,C, it is in complementary distribution with *wh*-movement to that same position. However, Icelandic embedded questions can be NomCIs, which indicates that *það* occupies a different position than Spec,C:

- (62) Við spurðum hann um **það** **hversu mikið** er hægt að fjárfesta  
 We asked him about that<sub>ACC</sub> how much is possible to invest  
 ‘We asked him how much it is possible to invest.’

The logical question in this regard is which position *það* occupies. By viewing a NomCI as a DP, we can easily posit that *það* is positioned, for example, in the D-head of the projection, while the interrogative pronoun occupies Spec,C in the embedded clause. But, within GCPH, one would wonder whether it is possible that the clause contains further projections in its left periphery (consider e.g. a Split-CP approach as in Rizzi (1997)). My view in this regard is that no position is available for *það* in the left periphery of the CP either. Consider once again the case of verbs like *hjálpa* ‘help’ (Garofalo 2023:107):

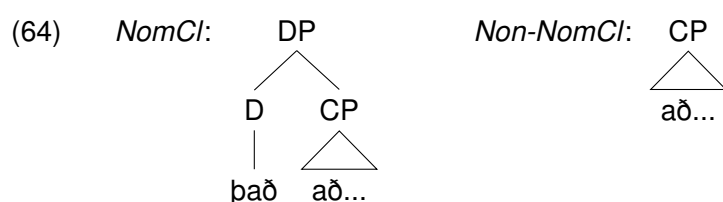
- (63) a. \*María hjálpaði honum **flutningana**  
 María helped him<sub>DAT</sub> moving out<sub>ACC</sub>  
 Intended: ‘María helped him moving out.’  
 b. María hjálpaði honum [(**\*það**) að flytja]  
 María helped him that<sub>ACC</sub> to move out  
 ‘María helped him moving out.’

These examples categorically exclude the possibility that DPs surface in the same position of the clausal complement. But, in light of GCPH, clausal complements, whether NomCIs or non-NomCIs, should be allowed to surface, as the relevant position is a non-DP position. In other words, a NomCI containing *það* in the left periphery should not be problematic as a complement of *hjálpa*. Despite this, *það* remains disallowed. More interestingly, it seems that every time a clause surfaces in a non-DP position, *það* is ungrammatical (while, by contrast, *það* in a direct object position is always possible). This fact can be explained if we interpret a NomCI as a DP, as the distribution of *það* would then be subsumed within the distribution of DPs, but it cannot be explained by GCPH.

In sum, we can see from these facts that GCPH cannot be a possible interpretation of the structure of NomCIs and non-NomCIs in Icelandic.

## 2.4.4 DP-shell vs. bare CP hypothesis

The next proposal to present is perhaps the most common across the literature over the last few decades (see e.g. Roussou (1991); Delicado Cantero (2013); Garofalo (2020; 2023); Rosenbaum (1967); Borsley and Kornfilt (2000); Kornfilt and Whitman (2011b); Kastner (2015) and many others). This time, we find a clear structural difference between NomCIs and non-NomCIs, which simply consists of embedding a CP into a DP projection in the case of the former type of clause. There is a slight difference across the various accounts based on whether the maximal projection of the NomCI is a DP or an NP or whether the CN (which represents the D-head in this proposal) is overt or not, but mainly they all present similar features. The structure of NomCIs presented in the literature is illustrated below. CN is the head of the DP shell, which embeds the CP:



If NomCIs and non-NomCIs surface as DPs and CPs respectively, it is possible to explain the various facts we have seen in our discussion on GDPH and GCPH. Let us consider, once again, the case of extraction (the examples are from Thráinsson (1979:110-111, 219, 235)):

- (65) a. Þeir ákváðu [(**pað**) að heimsækja Ólaf]  
 they decided that<sub>ACC</sub> to visit Ólafur<sub>ACC</sub>  
 ‘They decided to visit Ólafur.’  
 b. Ólaf<sub>i</sub> ákváðu þeir [(**\*pað**) að heimsækja    ]  
 Ólafur<sub>ACC</sub> decided they that<sub>ACC</sub> to visit  
 c. [Að heimsækja Ólaf]<sub>i</sub> ákváðu þeir [(**\*pað**)    ]  
 to visit Ólafur<sub>ACC</sub> decided they that<sub>ACC</sub>
- (66) a. Ég harma [(**pað**) að Jón skuli hafa barið Maríu]  
 I regret that<sub>ACC</sub> that Jón<sub>NOM</sub> shall have beaten up María<sub>ACC</sub>  
 ‘I regret that Jón has beaten up María.’  
 b. Maríu<sub>i</sub> harma ég [(**\*pað**) að Jón skuli hafa barið    ]  
 María<sub>ACC</sub> regret I that<sub>ACC</sub> that Jón<sub>NOM</sub> shall have beaten up  
 c. [Að Jón skuli hafa barið Maríu]<sub>i</sub> harma ég (**\*pað**)      
 that Jón<sub>NOM</sub> shall have beaten up María<sub>ACC</sub> regret I that<sub>ACC</sub>

The fact that extraction can only take place when *pað* does not surface suggests that NomCIs provide a DP barrier, while non-NomCIs do not and have a smaller structure.

Furthermore, let us consider the case of non-DP positions again. Verbs like *hjálpa* ‘help’, as we have already seen, are able to select a clausal complement, but, in that position, no DP can surface, nor a NomCI. Other verbs like *þvinga* ‘force’ display the same behavior:

- (67) a. \*María hjálpaði honum **flutningana**  
 María helped him<sub>DAT</sub> moving out<sub>ACC</sub>  
 Intended: 'María helped him moving out.' (Garofalo 2023:107)
- b. María hjálpaði honum [(**það**) að flytja]  
 María helped him that<sub>ACC/DAT/GEN</sub> to move out  
 'María helped him moving out.' (Garofalo 2023:107)
- (68) a. \*Það er enginn að þvinga fólk **birtingu þessara mynda** hérna  
 there is nobody to force people publication<sub>ACC</sub> these<sub>GEN</sub> photos<sub>GEN</sub> here  
 Literally: 'There is no one forcing people to the publication of these photos here.'
- b. Það er enginn að þvinga fólk [(**það**) að hafa heimasíður barnanna sinna]  
 there is nobody to force people that<sub>ACC</sub> to have homepages children.the their  
 hérna  
 here  
 'There is no one forcing people to have their children's homepages here.'  
 (from *bland.is*)

However, as soon as a DP is allowed in a specific position, a NomCl can also surface. Compare the examples of *þvinga* and *hjálpa* with the one below (which contains the matrix verb *segja* 'say, tell'):

- (69) a. Ég sagði honum **sannleikann**  
 I told him<sub>DAT</sub> truth.the<sub>ACC</sub>  
 'I told him the truth.'
- b. Ég sagði honum [(**það**) að hún væri farinn]  
 I told him<sub>DAT</sub> that<sub>ACC</sub> that she were gone  
 'I told him that she left.'

The fact that NomCls are allowed or disallowed, based on whether a DP is allowed or not in the same position, indicates that the distribution of NomCls is subsumed within that of DPs. To summarize, there is strong evidence in support of the idea that NomCls and non-NomCls display a different structure in Icelandic, namely DPs and CPs respectively.

## 2.5 Trigger and distribution of clausal nominalization

### 2.5.1 Introduction

Now that we have come to the conclusion that NomCls surface as DPs and non-NomCls as CPs, we can take a step forward in our analysis by discussing the most relevant hypotheses on the trigger and distribution of clausal nominalization in the linguistic literature, in light of the structural properties of both variants. I will focus on some semantic hypotheses first (which consider factivity or other semantic factors as a trigger of clausal nominalization) and, following this, some syntactic hypotheses (proposing DP selection and case as a solution to the nominalization puzzle). Lastly, after briefly discussing some less common accounts on clausal nominalization, I will

broadly delineate a preliminary version of my hypothesis in light of the information discussed in this chapter.

## 2.5.2 Factivity hypothesis

The first account to discuss is what I call the *Factivity hypothesis*, according to which the factive (or non-factive) meaning of a matrix verb affects the structure of its complement (see Kiparsky and Kiparsky (1971)). Thráinsson (1979), whose structural hypothesis has already been discussed in §2.4.2, proposed in Chapter 4 of his doctoral dissertation that factivity plays a role in the distribution of NomCIs in Icelandic. In particular, he argued that the adoption of a CN is a reflex of the tendency of factive predicates to select more complex arguments.<sup>40</sup> This view is supported by the fact that, in clausal objects, emotive factive predicates like *fagna* ‘rejoice’ or *sakna* ‘miss’ require a clausal argument and also that semi-factives like *vita* ‘know’ are able to select NomCIs. However, non-factives like *hugsa* ‘think’ are unable to do so, unless the pronoun refers to a previous context, e.g. the content of the clausal complement has already been mentioned previously (the judgments here below are Thráinsson’s):

- |      |                                                                                                                                                                                                 |               |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| (70) | a. Ég sakna *( <b>þess</b> ) að María skuli ekki vera hér<br>I miss that <sub>GEN</sub> that María shall not be here<br>‘I miss María not being here.’ (Thráinsson 1979:230)                    | TRUE FACTIVES |
|      | b. Ég fagna *( <b>því</b> ) að þú skulir vera kominn<br>I rejoice that <sub>DAT</sub> that you shall be come<br>‘I am happy that you came.’ (Thráinsson 1979:230)                               | TRUE FACTIVES |
|      | c. Ég veit ( <b>það</b> ) að María er farin alveg fyrir vísit<br>I know that <sub>ACC</sub> that María is gone entirely for sure<br>‘I know for sure that María is gone.’ (Thráinsson 1979:221) | SEMI-FACTIVES |
|      | d. Ég hugsa (*? <b>það</b> ) að Jón hafi barið Maríu<br>I think that <sub>ACC</sub> that Jón has beaten María<br>‘I think that Jón has beaten María.’ (Thráinsson 1979:216)                     | NON-FACTIVES  |

Thráinsson (1979:236) suggested that dative as well as genitive case needs to be expressed morphologically through *það*, which would entail pronominal obligatoriness in examples like (70a) and (70b). But he also observed that some non-factives like *spá* differ from this trend. They accept NomCIs as complements, but the pronoun is optional, even if dative case is assigned. He attributed this property, i.e. the possibility to select non-NomCIs (unlike *fagna*), to the fact that *spá* is a non-factive:

- |      |                                                                                                                                                                     |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (71) | Jón spáði ( <b>því</b> ) að María væri farin<br>Jón predicted that <sub>DAT</sub> that María was gone<br>‘Jón predicted that María was gone.’ (Thráinsson 1979:216) |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|

<sup>40</sup>Recall here that Thráinsson considered both non-NomCI and NomCI arguments as NPs because they surface in an NP slot. So, within GDPH, it is the presence of a CN specifically that makes the structure of the argument more complex.

Moreover, he claimed that factivity affects clausal subjects as well, as he also observed that subjects of factive predicates like *vera hörmulegt* ‘be deplorable’ are able to select NomCIs, while non-factives like *vera líklegt* ‘be likely’ cannot, unless the pronoun refers to a previous context (we have already discussed these examples in relation to GDPH):

- (72) a. Er (?\***það**) líklegt að Jón hafi barið Maríu?  
 is it likely that Jón has beaten María  
 ‘Is it likely that Jón has beaten María?’ (adapted from Thráinsson 1979:183-184)
- b. Er (**það**) hörmulegt að Jón skuli hafa barið Maríu?  
 is it deplorable that Jón shall have beaten María  
 ‘Is it deplorable that Jón has beaten María?’ (Thráinsson 1979:192, 194)

Nevertheless, there are various facts in Icelandic which do not support the factivity hypothesis presented by Thráinsson. First of all, regarding clausal subjects, stand-alone verbs like *teljast* ‘be considered’ or *valda* ‘cause’ (Thráinsson (1979:253) also mentioned this last verb in particular) require *það* after the finite verb. This occurs independently of clausal extraposition and factivity:

- (73) a. Telst \*(**það**) að Bjarki sé enn þá lifandi ótrúlegt kraftaverk?  
 is considered that<sub>NOM</sub> that Bjarki is still alive incredible miracle  
 ‘Is the fact that Bjarki is alive considered an incredible miracle?’
- b. Telst \*(**það**) ótrúlegt kraftaverk að Bjarki sé enn þá lifandi?  
 is considered that<sub>NOM</sub> incredible miracle that Bjarki is still alive  
 ‘Is the fact that Bjarki is alive considered an incredible miracle?’
- (74) a. Mér veldur \*(**það**) að María skuli elska Ólaf þungum áhyggjum  
 me causes that<sub>NOM</sub> that María shall love Ólafur heavy worries  
 ‘The fact that María loves Ólafur worries me a lot.’ (Thráinsson 1979:253)
- b. Mér veldur \*(**það**) þungum áhyggjum að María skuli elska Ólaf  
 me causes that<sub>NOM</sub> heavy worries that María shall love Ólafur  
 ‘The fact that María loves Ólafur worries me a lot.’

Pronominal obligatoriness also occurs with copula+adjective constructions independently of factivity if the whole clause surfaces after a subordinating conjunction, as shown in the examples below:<sup>41</sup>

- (75) a. ... [C' að [DP \*(**það**) að Jón sé hættur] sé líklegt]  
 that that<sub>NOM</sub> that Jón is quit is likely  
 ‘... that it is likely that Jón quit.’
- b. ... [C' að [DP \*(**það**) að hann skuli vera farinn] sé hörmulegt]  
 that that<sub>NOM</sub> that he shall be gone is deplorable  
 ‘... that it is deplorable that he left.’

An important question is, however, why the copula+adjective constructions in (72) contrast with the remaining counterexamples presented here. My proposal is that they differ due to the fact that

<sup>41</sup>Note that, in embedded clauses, since the conjunction is merged in C, the only position available for subjects is Spec,T. More discussion is offered in Chapter 4.

the clause can be merged as a complement of the adjective (see e.g. Roussou (1991) for Modern Greek). Adjectives like *líklegt* or *hörmulegt* are unable to subcategorize for a DP, so I assume they can only select non-NomCIs as complements, i.e. CPs:

- (76) a. \*Aðstæðurnar voru hörmulegar **þetta/þessu/þessa**  
 circumstances were deplorable this<sub>ACC/DAT/GEN</sub>  
 Literally: ‘The circumstances were deplorable this.’
- b. \*Fjarvist hans var líkleg **þetta/þessu/þessa**  
 absence his was likely this<sub>ACC/DAT/GEN</sub>  
 Literally: ‘His absence was likely this.’

CPs, however, are unable to be promoted to the subject position, as we would otherwise expect to see non-NomCIs after stand-alone verbs like *teljast* or *valda* as well, but that is not the case. In this configuration, the CP cannot be promoted to satisfy the *Extended Projection Principle* (Chomsky (1995) and considerable subsequent work within the Minimalist Program; I offer a more detailed discussion on the derivation of NomCIs and non-NomCIs as well as copula+adjective constructions in Chapter 4), therefore an invisible expletive emerges to satisfy EPP. This configuration is impossible with the stand-alone verbs we have just discussed.<sup>42</sup> In other words, *það* is optional with copula+adjective constructions because they possess a configuration where CPs can emerge, which is ungrammatical with stand-alone verbs.

Regarding non-factive predicates, *það* is not always discourse-anaphoric before a clausal subject in the first position. Consider the following example with the predicate *vera mögulegt* (‘be possible’):

- (77) **Það** að Guð sé til er mögulegt samkvæmt nýlegri grein í heimspeki  
 that<sub>NOM</sub> that God exists is possible according recent article in philosophy  
 ‘It is possible that God exists, according to a recent article in philosophy.’

This kind of example does not necessarily need a previous discourse for *það* to be allowed to surface (e.g. if we consider it as the very first sentence of a text), which indicates that *það* can play the role of a CN with non-factives as well.

As for clausal objects, I present two counterarguments against Thráinsson (1979). The first one is that there are examples of non-factives where *það* can surface before a clausal object without being discourse-anaphoric, as shown in the example below. This fact neutralizes the differences between accusative direct objects in Icelandic based on factivity (see §2.3.2):

- (78) Ég held (**það**) að góður liðsandi sé mikilvægur fyrir fótboltalið  
 I think that<sub>ACC</sub> that good team.spirit is important for football.team  
 ‘I think a good team spirit is important for a football team.’

The second counterargument against the factivity hypothesis is that emotive factive predicates

<sup>42</sup>We will see in the data presented in Chapter 3, however, that passive constructions function differently in this regard, as it is more common for an accusative clausal direct object to surface as a nominative non-NomCI after passivization.

subcategorizing for a PP like *vera svekkur yfir* ‘be annoyed about something’ and *vera pirraður yfir* ‘be annoyed about something’, as well as their verbal variants with the middle voice suffix *-st*, i.e. *svekkjast yfir* and *pirrast yfir* do not require a mandatory pronoun, contrary to what we would expect following the factivity hypothesis (all the examples below are actual examples from the Internet with some slight adaptations, but originally without *það*):

- (79) a. Allardyce er svekkur yfir (**því**) að hafa fengið sparkið  
Allardyce is annoyed over that<sub>DAT</sub> to have got kick.the  
‘Allardyce was annoyed about the fact that he was fired.’
- b. [Ég] [v]ar einmitt að koma inn úr labbi og svekkjast yfir (**því**) að sjá  
I was exactly to come inside from walk and be annoyed over that<sub>DAT</sub> to see  
skottuna ekki  
ghost.the not  
‘I’d just come in from a walk and was annoyed at not having seen the ghost.’
- (80) a. [Er] hann ekki bara pirraður yfir (**því**) að þú sért að vekja hann[?]  
Is he not just annoyed over that<sub>DAT</sub> that you are to wake up him  
‘Isn’t he just annoyed about the fact that you are waking him up?’
- b. Ég er að pirrast yfir (**því**) að núna vilja þeir ekki halda sætunum  
I am to be annoyed over that<sub>DAT</sub> that now want they not keep seats.the  
‘I am getting annoyed about the fact that they don’t want to keep the seats now.’

In light of these facts, we can safely assume that factivity has little influence over the distribution of NomCIs and non-NomCIs. I might concede the possibility that it can account for the difference between *spá* and *fagna*,<sup>43</sup> but it is not able to account for clausal subjects with stand-alone verbs (or non-factive copula+adjective constructions like (77) for that matter) as well as prepositional objects and accusative direct objects.

### 2.5.3 Other semantic influences of matrix verbs on clausal nominalization

Thráinsson’s approach is not the only one where verbal meaning is claimed to be able to affect the structure of clausal arguments. We find, once again in relation to English, another approach in Kastner (2015). He classified verbs into three different categories:

- (81) NON-STANCE: regret, know, remember, realize, notice...  
RESPONSE STANCE: deny, accept, agree, admit, confirm, verify...  
VOLUNTEERED STANCE: think, suppose, assume, claim, suspect...

Kastner claimed that non-stance and response stance predicates are the ones that subcategorize for a NomCI (although no CN in his approach is phonetically realized), while only volunteered stance verbs differ, as they select clausal complements without any need for a DP shell over CP. His argument built upon the following contrast (Kastner 2015:163):

<sup>43</sup>In Chapter 3, however, we will also see that *spá* is much more likely to subcategorize for a NomCI rather than a non-NomCI, as the frequency of *það* reaches 90%.

- (82) a. What do you **think** (that) John stole \_?  
 b. Where do you **think** John came from \_?  
 c. Who do you **think** \_ stole the cookies?  
 d. Why do you **think** that John stole the cookies \_?
- (83) a. What do you **remember/deny** that John stole \_?  
 b. Where do you **remember/deny** that John came from \_?  
 c. \*Who do you **remember/deny** \_ stole the cookies?  
 d. #Why do you **remember/deny** that John stole the cookies \_?

The former verb is a volunteered stance verb, while the latter is either a non-stance verb (*remember*) or a response stance verb (*deny*). The contrast shows that extraction of a subject or an adjunct from the clausal complement (see (c) and (d) examples) is disallowed with non-stance and response stance verbs, whereas it is possible with volunteered stance verbs like *think*, while extraction of a complement or from the complement of a preposition is allowed in all of the three classes (see (a) and (b) examples). This difference indicates that the structure of the clausal complement is simpler with volunteered stance verbs.

What is more, he observed a further contrast between these classes through fronting. With volunteered stance verbs, fronting is possible, contrary to the other two classes. This also suggests that the clausal complement of *think* has a simpler structure compared to the one of *regret* and *remember* (the judgments are Kastner's (2015:157)):

- (84) a. John thinks that [this book, Mary read]  
 b. I can assure you that [that film, I don't want to ever see again]
- (85) a. \*John regrets that [this book, Mary read]  
 b. \*John remembers that [this book, Mary read]

Considering the restrictions applied onto the complements of non-stance and response stance verbs, Kastner drew the conclusion that they subcategorize for a DP, which should be a NomCl, while this does not happen with volunteered stance verbs.

Icelandic, however, is incompatible with this pattern, as *það* neutralizes semantic differences between various predicates, interestingly all assigning accusative case (see §2.3.2):

- (86) a. Ég harma (**það**) að ég sé ekki búinn að senda umsóknina  
 I regret that<sub>ACC</sub> that I am not finished to send application.the  
 'I regret the fact that I did not send my application.'
- b. Karl samþykkir (**það**) að einhver annar stjórnir fundinum  
 Karl accepts that<sub>ACC</sub> that someone else leads meeting.the  
 'Karl accepts the fact that someone else leads the meeting.'
- c. Ég held (**það**) að góður liðsandi sé mikilvægur fyrir fótboltið  
 I think that<sub>ACC</sub> that good team.spirit is important for football.team  
 'I think a good team spirit is important for a football team.'

This contrasts with Kastner’s assumption, as verbs like *halda* ‘think’ can select a NomCI despite the fact that it is a volunteered stance verb. And this is not the only verb of the class that displays these properties. Let us take a look, for instance, at the verbs *gruna* ‘suspect’ and *vænta* ‘expect’:

- (87) a. Mig grunar (**það**) að einhver sé búinn að taka peninga úr  
 me<sub>ACC</sub> suspects that<sub>ACC</sub> that someone is finished to take money from  
 kassanum  
 cash-machine.the  
 ‘I suspect someone took money from the cash machine.’
- b. Ég vænti \*(**þess**) að þeir deili öllu með mér  
 I expect that<sub>GEN</sub> that they share everything with me  
 ‘I expect them to share everything with me.’

In both instances, the pronoun is definitely possible and, in particular, *það* is mandatory with the verb *vænta*. Therefore, as we can see, Icelandic does not confirm the distinction proposed in Kastner (2015). We can safely assume that any semantic distinction between matrix verbs that we can find in the linguistic literature is going to be problematic for Icelandic, due to the fact that accusative direct clausal objects can all be nominalized without exception. The only verbs that never allow nominalization to take place are verbs like *þvinga* ‘force’ or *hjálpa* ‘help’, which do not take DP complements in the same position where a clausal complement can surface (as we have already observed earlier).<sup>44</sup>

## 2.5.4 DP selection hypothesis

We now move on to what I call the DP selection hypothesis. Cross-linguistically, it has been observed in various accounts that there are some similarities between DPs and NomCIs, in the sense that both can be introduced or marked with a functional item expressing definiteness (see some cross-linguistic examples in Roussou (1991); Han (2008); Moulton et al. (2020); Hara et al. (2013)). Now, the presence of a determiner before clausal arguments, sometimes even obligatorily as is the case of Icelandic, suggests that a bare clausal argument by itself is at times not enough when selected by specific lexical items. Within the DP selection hypothesis, the use of a CN is explained by the fact that the item subcategorizing for a clausal complement requires a DP. A first set of studies, some of which are based on *Lexical-Functional Grammar* (Kaplan and Bresnan (1982)), viewed clausal complements as verbal categories, which are incompatible with items that can only subcategorize for nominal categories (see e.g. Borsley and Kornfilt (2000); Panagiotidis and Grohmann (2009); Kornfilt and Whitman (2011b; 2011a); also cf. Bresnan (1997) among others). For a clausal complement to become compatible with the relevant lexical item, it is embedded into a *nominal functional category* (see Borsley and Kornfilt (2000)), which contains a determiner.

<sup>44</sup>Note that the clausal complements that we find after verbs like *þvinga* or *hjálpa* are not direct objects, therefore these verbs do not constitute an exception for my assumption that accusative clausal direct objects can always be nominalized.

This view is supported by two facts across these studies: a) the extension of nominalization across syntactic positions; and b) the hierarchy of nominal and verbal categories in the syntactic structure. As for the former, we can find a good summary of the extension of nominalization across syntactic positions in Kornfilt and Whitman (2011a),<sup>45</sup> where it is shown that not only can nominalization take place at CP-level (which is the one relevant for our analysis), but also e.g. at vP level (as in Italian infinitives) and TP level (as in Turkish clausal complements; the reader can find various examples of these two types of nominalization in Kornfilt and Whitman (2011a) as well as other articles in the same volume of *Lingua*). As for the latter fact, it appears that, cross-linguistically speaking, nominal categories embed verbal categories or vice-versa but the distribution of both subtrees is uniform rather than interspersed (see Bresnan (1997)) and this can be observed at all levels where nominalization can take place.<sup>46</sup> As a result, we can see that in various SVO languages like English, Spanish or Italian, all the nominal components involved in nominalization are linearly on the left side of a nominalized item, while the verbal counterpart is on the right side. This can be exemplified with the following example from English gerunds:

(88) His **playing** the violin was quite remarkable

The gerund *playing*, which is nominalized, behaves like a verb as it selects a direct object on its right side. At the same time, it behaves like a noun by being preceded on its left side by modifiers like possessive pronouns. Icelandic clausal nominalization, in this regard, is not an exception since we have established that, in a NomCl, a DP shell embeds a CP. *Pað* represents the nominal component, and the CP the verbal one.

In addition, other analyses point to the possibility that DP selection is the trigger of clausal nominalization, in particular Alrenga (2005) and Takahashi (2010) which we have already discussed to some extent in relation to GDPH (see §2.4.2) and also Kim and Sag (2005). These studies identify two main classes of verbs in English, which subcategorize for a DP and a CP respectively. Verbs like *capture* or *blame* require the presence of a clausal complement embedded in a DP projection, whereas verbs like *hope* or *think* normally cannot, as they subcategorize for CPs only:

- (89) a. This formulation of the rule **expresses / captures / reflects / brings out** \*(the fact) that these nouns behave differently  
 b. That these nouns behave differently **is captured** by this formulation of the rule  
 (Takahashi 2010:354)

<sup>45</sup>The article in question summarizes the findings of the articles published in the same volume of *Lingua* on the topic of nominalization.

<sup>46</sup>Due to the double nature (nominal and verbal) of nominalized items, Bresnan (1997), Borsley and Kornfilt (2000), Kornfilt and Whitman (2011a; 2011b) and also Panagiotidis and Grohmann (2009) proposed the existence in nominalization constructions of a *mixed projection* (or *mixed extended projection*, see Kornfilt and Whitman (2011b)), which is „verbal up to a certain point in the structure, but nominal at and above the level where a nominal functional category is introduced“ (Kornfilt and Whitman 2011b:1160). See also other similar analyses like Marantz (1997).

- (90) a. Most baseball fans **hoped / felt / wished / insisted / reasoned** that the Giants would win the World Series  
 b. \*That the Giants would win the World Series **was reasoned** by most baseball fans (Takahashi 2010:354)
- (91) a. I blame \*(it) on you [that we can't go]  
 b. John thought (?it) to himself [that we had betrayed it] (Kim and Sag 2005:251)

As already mentioned in §2.4.2, Takahashi (2010) proposed that the contrast between *capture*-verbs and *hope*-verbs emerges due to the fact that clausal complements that are base-generated as CPs are unable to move (see (90b)), while those that are merged as DPs can do so (see (89b)). This is summarized in Takahashi's MCCG, which is re-illustrated below:

(92) *Moved Clausal Complement Generalization*

A clausal complement is allowed to move only if its base-generated position is one in which a DP is allowed to appear

MCCG is important in the analysis of NomCIs from a syntactic perspective due to the fact that it provides an answer for a dilemma related to clausal complements discussed in the previous linguistic literature, namely that when clausal complements move, their lower copies display DP properties (see e.g. Alrenga (2005); also Koster (1978)). These DP properties are explained by the fact that the moved clausal complement itself is base-generated as a DP and remains a DP after movement, although the DP projection remains silent.

As for the contrast presented by Kim and Sag (2005) in (91), the complement of *think* is generally a CP, so that the presence of a CN is problematic. However, since examples like the following exist, they suggest that *it* might emerge with CP-selecting verbs if the clause undergoes extraposition:

- (93) I thought **it** that it would be nearly impossible for the filmmakers to sustain such a level of excitement through the rest of the movie  
 (Kim and Sag 2005:262)

The *blame* class, on the other hand, makes a requirement for a nominal category, which explains why *it* is mandatory before the clausal complement. But Kim and Sag, unlike the analysis in Takahashi (2010), also show a third class of verbs, the *expect* class, where the CN is optional (Kim and Sag 2005:251):

- (94) Nobody expected (**it**) of you [that you could be so cruel]

They explained the contrast between verbs like *blame* and *expect* with the fact that the former involves a predicative PP, while the latter does not, as shown in the following examples:

- (95) a. The blame was **on me** (indicates that *PP* is predicative)  
 b. I blame **\*(it) on you** [that we can't go]
- (96) a. \*The expectation is **of me** (indicates that *PP* is non-predicative)  
 b. Nobody expected **(it) of you** [that you could be so cruel]

Considering the data discussed here, the main question is whether the DP selection hypothesis is also valid for Icelandic. First of all, it is true that this hypothesis is able to explain the presence of mandatory pronouns in clausal subjects in subject position, clausal indirect objects, dative and genitive clausal objects, etc.: they must be embedded into a DP projection due to the fact that the category selecting the clausal argument requires a DP and, therefore, a CP would be incompatible with such requirements. What is more, if we also consider the instances where *það* is ungrammatical, as with verbs like *þvinga* 'force' or *hjálpa* 'help', the DP selection hypothesis entails that *það* cannot surface due to the fact that these verbs do not select a DP as complement in the same position where the clause appears, but only a CP (similarly to the *hope* class in Takahashi (2010)).

However, if we focus on those instances where both NomCIs and non-NomCIs are acceptable, this analysis becomes more problematic. Assuming that, from a structural perspective, NomCIs emerge as DPs and non-NomCIs as CPs, the fact that both variants are acceptable indicates that a CP in Icelandic can be a valid complement of a certain lexical item like a verb or a preposition. But, at that point, since the relevant lexical item is already able to subcategorize for a verbal category, what is the motivation for nominalizing? If we consider Kim and Sag's approach here, where we find an instance of optional nominalization with the *expect* class, the explanation offered by Kim and Sag (2005) is that the PP involved in the structure before the CP is non-predicative, unlike the *blame* class. However, we cannot use this argument to account for the difference between obligatory and optional *það* in Icelandic. Consider, for instance, the case of clausal subjects before and after the finite verb. The matrix verb involved is the same as well as the subject. But the position where it surfaces affects the distribution of clausal nominalization. The same can also be said of the contrast presented by Thráinsson (1979) between clausal objects of *spá* 'predict' and *fagna* 'rejoice' (as illustrated in the examples below). These two verbs do not have a predicative PP preceding the CP, but they still differ as to whether *það* is optional or mandatory. It is difficult, therefore, to interpret this contrast in light of Kim and Sag's approach:

- (97) a. Jón spáði **(því)** að María væri farin  
 Jón predicted that<sub>DAT</sub> that María was gone  
 'Jón predicted that María would have left.' (Thráinsson 1979:216)
- b. Ég fagna **\*(því)** að þú skulir vera kominn  
 I rejoice that<sub>DAT</sub> that you shall be come  
 'I am happy that you came.' (Thráinsson 1979:230)

In sum, all these data tell us that the DP selection hypothesis by itself can account for obligatory and ungrammatical nominalization, but not for pronominal optionality.

### 2.5.5 Case-Marking hypothesis

Let us now move to the next hypothesis, which I call the *Case-marking hypothesis*. The Case-marking hypothesis assumes a correlation between case and the use of a CN and attributes two possible causes to the existence of clausal nominalization: either the necessity for case to be assigned (e.g. Roussou (1991)) or the necessity for case to be visible (e.g. Knyazev (2016)). These two proposals have so far been presented for clausal nominalization in other languages in the linguistic literature. Therefore, we will first observe some cross-linguistic accounts where these two options have been proposed and then we will focus on Icelandic in order to see whether one of them is valid to explain clausal nominalization.

Let us start with the former option, which proposes that the reason why a CN surfaces (and, therefore, also a DP shell) before a clausal complement is that the relevant item that selects the clause needs to assign case and the clausal complement itself is unable to check it. This can be explained with Stowell's (1981) *Case Resistance Principle* (henceforth CRP), which is presented below:

(98) *Case Resistance Principle*

Case may not be assigned to a category bearing a Case assigning feature

(Stowell 1981:146)

An important point made by Stowell, however, is that if a CP, which cannot check case by itself, is moved from a case position via  $\bar{A}$ -movement, its trace can check case.<sup>47</sup>

Following CRP, Roussou (1991) pointed out that structural case, in particular nominative, triggers clausal nominalization in clausal subjects. By contrast, if case is not assigned, a determiner is disallowed. This contrast can be observed in Modern Greek with clausal subjects and objects:

(99) a. [\***To**] oti ehis filus] simeni pola  
the<sub>NOM</sub> that have-2sg friends mean-3sg much  
'That you have friends means a lot.' (Roussou 1991:78)

b. Ksero (**\*to**) [oti efighe]  
know-1sg the<sub>ACC</sub> that left-3sg  
'I know that he left.' (Roussou 1991:93)

M. GREEK

Let us move to the second option proposed within the case-marking hypothesis, i.e. case visibility. One of the languages that has been analyzed in light of this proposal is Russian. Knyazev (2016:22) in particular showed that when inherent case is assigned by a preposition in Russian, a CN must surface:

<sup>47</sup>I will propose in Chapter 4, in light of the empirical data that will be presented in Chapter 3, that the trace left by a non-NomCI is a DP trace due to the fact that the clause is originally merged as a DP and undergoes structural removal in a position in which *pað* does not have to check features. This aligns to a great extent with Stowell's insights.

- (100) Maša nadeetsja na **\*(to)**, što u nee budet mnogo svobodnogo vremeni  
 Masha<sub>NOM</sub> hopes on that<sub>ACC</sub> that at her<sub>GEN</sub> will be a lot free time<sub>GEN</sub>  
 ‘Masha hopes that she will have a lot of free time.’ RUSSIAN

As for Russian clausal direct objects (which we have already discussed in relation to GDPH), when a certain verb is agentive and the object of such verb is a clausal complement, it requires accusative case licensing. However, the pronoun should not be phonetically realized, otherwise the sentence is degraded:

- (101) Maša skazala (?\***to**), što ona ustala  
 Masha<sub>NOM</sub> said that<sub>ACC</sub> that she<sub>NOM</sub> tired  
 ‘Masha said that she is tired.’ (Knyazev 2016:25) RUSSIAN

Considering the fact that both variants of clausal complements are DPs in Russian, according to Knyazev (see §2.4.2), why is there a difference between examples with a prepositional object and examples with a direct object in accusative case? Knyazev proposed an economy condition based on which CN is less likely to surface in instances of structural case:

- (102) Case is preferred to remain unrealized (whenever possible)  
 (Knyazev 2016:39)

In other words, *to* must be visible in instances of inherent case in Russian, while it should remain covert in instances of structural case. As for clausal subjects, however, Knyazev, who followed Landau (2007) on the idea that the D-head that checks the EPP-feature in T must be overt, proposed that the obligatory presence of *to* in examples like the following is due to the fact that the D-head must be realized, unlike Roussou (1991), who attributed pronominal obligatoriness to structural case assignment.<sup>48</sup>

- (103) **\*(To)**, što on èto skazal, dokazyvaet ego nevinovnost  
 that<sub>NOM</sub> that he<sub>NOM</sub> this<sub>ACC</sub> said proves his innocence<sub>ACC</sub>  
 ‘The fact that he said this proves his innocence.’ RUSSIAN

Let us now take a look at Icelandic in light of the Case-Marking Hypothesis. Similar to the DP selection hypothesis, this proposal can be applicable in instances where *það* is either mandatory or ungrammatical. But pronominal optionality remains problematic. Assuming that NomCIs are DPs while non-NomCIs are CPs, the case visibility option proposed by Knyazev (2016) already appears incompatible with the Icelandic data, in particular his economy condition on structural case. Recall that *það* is optional before accusative clausal direct objects, differently from the example from Russian in (101):<sup>49</sup>

<sup>48</sup>Further discussion on NomCIs and non-NomCIs in Russian in different syntactic positions is offered in §4.5.5.

<sup>49</sup>However, it is also true that the empirical data from Icelandic presented in Chapter 3 align to some extent to Knyazev’s view, albeit not completely. Both NomCIs and non-NomCIs are acceptable in structural case positions (see, in particular, clausal direct objects with nominative and accusative case), but NomCIs are systematically the least frequent.

- (104) a. Ég harma (**það**) að ég sé ekki búinn að senda umsóknina  
 I regret that<sub>ACC</sub> that I am not finished to send application.the  
 ‘I regret the fact that I did not send my application.’
- b. Karl samþykkir (**það**) að einhver annar stjórnir fundinum  
 Karl accepts that<sub>ACC</sub> that someone else leads meeting.the  
 ‘Karl accepts the fact that someone else will lead the meeting.’
- c. Ég held (**það**) að góður liðsandi sé mikilvægur fyrir fótboltalið  
 I think that<sub>ACC</sub> that good team.spirit is important for football.team  
 ‘I think a good team spirit is important for a football team.’

But let us consider Roussou’s (1991) approach. Pronominal obligatoriness in Icelandic can be explained by the fact that case must be assigned and, since the clausal argument itself cannot do so, *það* intervenes so that case can be assigned. But pronominal optionality remains a problem, in particular with accusative clausal direct objects. The fact that *það* can introduce them indicates that case is assigned. However, a CP can still surface in that position despite the fact that CRP excludes the possibility for a CP to be assigned case. How do we solve this dilemma? In this regard, we find an interesting account on Icelandic in Ingason (2018), who proposed, challenging CRP (which is at the basis of Roussou’s proposal), that CPs in Icelandic can be assigned case-marking directly. He presented two arguments in favor of this hypothesis. Firstly, floating quantifiers associated with clausal arguments are assigned case, as shown in (105) where the clausal object is passivized and moved to the first position. Secondly, direct objects surface in accusative case when the subject is sentential (see (106)), which implies that nominative case has also been assigned to the subject if we were to assume a configurational theory of case like *Dependent Case Theory* (Yip et al. (1987); Marantz (2000) among many others; cf. also Sigurðsson (2003:249); see further discussion in §4.4):

- (105) [Að hann hefði sigrað] var **einu** haldið fram  
 that he had won was one<sub>DAT</sub> claimed  
 ‘It was only claimed that he had won.’ (Ingason 2018:448)
- (106) a. [<sub>CP</sub> Að Gunnar skyldi ekki hafa mætt] drap alveg **stemninguna**  
 that Gunnar should not have attended killed totally mood.the<sub>ACC</sub>  
 ‘It killed the mood that Gunnar did not attend.’
- b. [<sub>CP</sub> Að syngja of hátt] gæti angrað **hana**  
 to sing too loudly could irritate her<sub>ACC</sub>  
 ‘Singing too loudly could irritate her.’ (Ingason 2018:452)

If we assume for the moment that Icelandic clausal arguments can be assigned case directly, an important consequence is that we are able to explain the issue of pronominal optionality. If CN and the CP are both capable of being assigned case, we have a valid reason why *það* is optional with clausal objects. However, we are unable to explain two further issues. The first one is the existence of clausal nominalization itself. If a CP can already receive case, there is no motivation

for nominalizing the clause and making the structure of the clausal argument more complex.<sup>50</sup> The second problem we cannot explain is pronominal obligatoriness. If CPs can receive case-marking directly, then we should at least expect to see an optional pronoun in all complement positions. However, we have seen multiple times that certain CNs like the following are mandatory:

- (107) *Ég fagna \*(því) að ég skuli hafa hætt við flugið*  
 I rejoice that<sub>DAT</sub> that I shall have cancelled flight.the  
 'I am happy that I cancelled my flight.' (Garofalo 2020:19)

But if direct case marking for CPs is not a correct analysis, how do we account for the examples in (105) and (106)? As for (106), I believe a CRP-compliant analysis can be a valid alternative interpretation. If the sentential subject is moved to Spec,C, it is possible, following CRP, that the trace left by the CP (e.g. in Spec,T, see Chapter 4) is the one that is assigned nominative case rather than the overt CP itself (further discussion on this issue is offered in §4.4). As for (105), the clause can be nominalized, but it is not necessarily assigned dative case, according to native speakers I consulted, as the pronoun can surface as *það*. By contrast, the floating quantifier cannot appear in any other case than dative (the example is adapted from Ingason (2018:451); see also Garofalo (2020:4)):

- (108) [*því/Það* að hann hefði sigrað] var *einu/\*eitt* haldið fram  
 that<sub>DAT/?</sub> that he had won was one<sub>DAT/?</sub> claimed  
 'The only thing that was claimed was that he had won.'

In sum, the approaches presented here are unable to fully explain the distribution of NomCIs and non-NomCIs, in particular due to pronominal optionality, which remains a puzzling dilemma for the moment. In order to solve it, we have to keep in mind that we need to find a criterion based on which a CN surfaces even if the CP can already follow a certain lexical item as complement.

## 2.5.6 Less common hypotheses

### 2.5.6.1 Placeholder hypothesis

We can find in the literature other accounts on clausal nominalization which are much less common. One of them is what I call the placeholder hypothesis and is found in Garofalo (2015). This analysis focuses on prepositional clausal complements and proposes: a) that the role of *það* is to be a placeholder that is adopted for the sake of completing the structure of the matrix clause before the subordinate clause is processed; and b) that the clause type that follows *það* determines whether *það* is more or less likely to surface. In particular, this study observed that indirect questions are correlated to a higher frequency of *það* while infinitives are less likely to be nominalized. As for *that*-clauses, neither the NomCI variant nor the non-NomCI one is more frequent. In order

<sup>50</sup>To be clear, this motivation for nominalizing cannot be found in the semantics of the matrix verb, as we have already seen that both clausal variants are acceptable in the direct object position with accusative case independently of the verb involved.

to explain this contrast, the placeholder hypothesis proposed that co-referentiality between the subject of the embedded clause and the subject or the object of the matrix clause (as in control infinitives) makes *það* less likely to surface. In indirect questions, *það* is more likely to emerge due to a possible linear mismatch between the case assigned by the relevant subcategorizing preposition and the case of the interrogative pronoun in the embedded clause, which is assigned by a different item:

- (109) a. *Ég veit ekki neitt um [hverju hann bjóst við]*  
 I know not anything about<sub>→ACC</sub> what<sub>DAT</sub> he expected  
 ‘I don’t know anything about what he expected.’
- b. *Ég veit ekki neitt um það [hverju hann bjóst við]*  
 I know not anything about<sub>→ACC</sub> that<sub>ACC</sub> what<sub>DAT</sub> he expected

However, the placeholder hypothesis is built upon a narrow data sample and focuses specifically on prepositional clausal complements. What is more, we know for a fact that a NomCI can be nested inside the matrix clause, e.g. clausal subjects in Spec,T:

- (110) *Er [\*(það) að læra á hljóðfæri] eins og að læra að hjóla?*  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 ‘Is learning to play an instrument like learning to ride a bike?’ (Garofalo 2023:106)

In this regard, the placeholder hypothesis is problematic because it assumes that the matrix clause and the subordinate clause are two separate entities, but it is actually not the case if we observe the example above.

### 2.5.6.2 Movement hypothesis

The movement hypothesis is found in Shahar (2008) and posits that whenever a clausal argument is extraposed, it leaves behind an underspecified copy of itself that is phonetically realized as a pronoun. This indicates that movement is the main trigger behind clausal nominalization. Shahar mainly applied this hypothesis on English, where the pronoun *it* surfaces whenever clausal subjects or objects are extraposed. His main argument is built upon the fact that *it* and the clausal argument are part of the same chain. Contrary to Bennis (2010), who claimed that the pronoun *it* was the only element to control PRO, Shahar claimed that both *it* and the embedded clause control PRO, which suggests that they are part of the same chain, as shown in the following examples (see Shahar (2008:31)):

- (111) a. \*Before PRO seeming that John was thirsty, it was likely that he was hungry  
 b. Before PRO seeming/appearing likely, it was unlikely that John would get the job

In the first example, if *it* and the clausal argument were two different entities, the pronoun should control PRO (as in Bennis (2010)) and the result of the first example should therefore be grammatical, but it is not. This suggests that the pronoun is part of a chain associated with the embedded clause. The ungrammaticality of the first example can be explained if PRO is replaced with the

clausal argument, as shown in (112). By contrast, the second example remains grammatical if PRO is replaced with the clausal argument (see Shahar (2008:32)):

- (112) a. \*That he was hungry seemed that John was thirsty  
b. That John would get the job seemed/appeared likely

Of course, it is difficult to apply his approach to clausal nominalization in Icelandic. Shahar's proposal works well with English because *it* always surfaces when the CP is extraposed, but the same cannot be said for Icelandic, as *það* and the clausal argument can co-occur e.g. in the first position. If we assumed Shahar's hypothesis, it would entail that the CP and a copy of itself can surface together in the same position. Therefore, we cannot attribute to extraposition the role of trigger of clausal nominalization in Icelandic.

### 2.5.7 The contribution of this dissertation: a preliminary hypothesis

As we have seen, the various hypotheses I have presented in this section, especially the ones based on data from Icelandic, were to some extent problematic and could not map by themselves the complex distribution of *það* across syntactic positions. First of all, it has come to light that semantic factors have little effect on the distribution of NomCIs and non-NomCIs, especially due to the fact that accusative clausal objects neutralize any semantic distinction among matrix verbs. Secondly, pronominal optionality also constitutes a major issue, as the syntactic hypotheses we have mentioned, like the DP selection hypothesis and the Case-marking hypothesis, cannot easily explain why both NomCIs and non-NomCIs are acceptable.

The question is, at this point, how this dissertation is going to contribute to the discussion on clausal nominalization in Icelandic, in particular about its distribution and purpose. I will broadly delineate here the hypothesis that I propose for Icelandic clausal nominalization in light of the discussion in the current chapter (I will present my hypothesis in further detail, as well as certain problematic exceptions and theoretical implications, in Chapters 3 and 4).

Based on the Icelandic data observed so far, pronominal optionality appears to be a common trait of clausal nominalization, e.g. in clausal subjects in the first position, accusative clausal direct objects, postcopular clauses, prepositional objects, etc. In a way, we can assume optionality to be a default trait of Icelandic clausal nominalization. Now, provided that this assumption is correct, how do we consider pronominal ungrammaticality and pronominal obligatoriness? Regarding the former, since the distribution of NomCIs appears to be subsumed within that of DPs, we have already inferred in §2.4.4 that non-DP positions are unable to host NomCIs. Moreover, extraction from the clausal argument causes a NomCI to be impossible (as shown by Thráinsson (1979)). These are the two domains where pronominal ungrammaticality is accounted for in light of what we have observed so far.

As for pronominal obligatoriness, if both clausal variants are acceptable by default yielding pronominal optionality, then pronominal obligatoriness consists in an exceptional instance where

a non-NomCI is disallowed.<sup>51</sup> This interpretation has two important implications: a) there must be a syntactic mechanism in Icelandic that allows NomCIs to surface alongside non-NomCIs by default; and b) there must be a restriction rule applied to specific syntactic positions which rules out non-NomCIs. I propose, in relation to implication (a), that optional *það* is the reflex of a DP selection mechanism:

(113) *NomCI selection mechanism*

A NomCI is always allowed if it merges in a DP position

If a DP is licensed in a certain position, a NomCI is automatically allowed to surface. In this way, even if a CP can be selected by a certain item, the presence of a NomCI is automatically granted by subcategorization for a DP. This mechanism, however, poses another problem, which is case assignment. How are non-NomCIs allowed as clausal direct objects if accusative case is assigned, since CPs cannot check case based on CRP (as discussed in the case-marking hypothesis)? My proposal in this regard is that accusative case and, by extension, structural case, does not correspond to a case feature to check in syntax (I made a similar claim in this regard in Garofalo (2023)).<sup>52</sup> It is possible, for example, that structural case is rather assigned through a configurational case mechanism as in *Dependent Case Theory* (Yip et al. (1987); Marantz (2000) among others), which I have mentioned in relation to the Case-marking hypothesis. For now, we do not need to delve into the possible theories that would explain case marking as I am just offering a preliminary description of my hypothesis in light of what has been discussed so far. However, after presenting the empirical data in Chapter 3, I will discuss the problem of structural case in more detail from a theoretical perspective.

Let us now focus on implication (b) mentioned above, i.e. there must be a restriction rule applied to specific syntactic positions that rules out non-NomCIs. I propose in this case that two possible conditions cause non-NomCIs to be disallowed, as described below:

(114) *Non-NomCI filter*

A NomCI is required if (it is enough to fulfill one condition):

- a) a case feature must be checked by the relevant clause
- b) the relevant clause is in a position in which a DP is required

If it is correct to assume that structural case does not correspond to case features to check in syntax, lexical case, by contrast, tends to require clausal nominalization, which suggests that lexical case corresponds to a case feature to check as a general rule, as described in the first

---

<sup>51</sup>An alternative view would be to consider pronominal obligatoriness as the default trait of clausal nominalization, but this would entail that a non-NomCI is exceptionally allowed in those instances where *það* is optional. I find this alternative interpretation problematic, especially due to the fact that both variants are systematically acceptable in structural case positions, in particular in direct object position (clausal subjects after the finite verb represent an exception to this rule; I will discuss this in Chapters 3 and 4).

<sup>52</sup>This hypothesis will be confirmed by the data in Chapter 3, where, although both clausal variants are acceptable in structural case positions, a non-NomCI is generally more preferable. This pattern also extends beyond structural case positions, as accusative case displays similar results in prepositional objects.

condition in (114). We have seen this, for instance, in the case of dative and genitive case in canonical argument positions:

- (115) a. *Ég fagna \*(því) að ég skuli hafa hætt við flugið*  
 I rejoice that<sub>DAT</sub> that I shall have cancelled flight.the  
 ‘I am happy that I cancelled my flight.’ (Garofalo 2020:19)
- b. *Verkefnastjórinn krafðist \*(þess) að skýrslunni yrði skilað*  
 project-manager.the demanded that<sub>GEN</sub> that report.the became sent  
 strax  
 immediately  
 ‘The project manager demanded that the report be sent immediately.’  
 (Garofalo 2020:21)

As for the second condition, there are certain syntactic positions that require a DP independently of which case is assigned.<sup>53</sup> For instance, recall the contrast between nominative clausal subjects before and after the finite verb:

- (116) a. *Er [(það) að læra á hljóðfæri] eins og að læra að hjóla?*  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 ‘Is learning to play an instrument like learning to ride a bike?’ (Garofalo 2023:106)
- b. *[(Það) að læra á hljóðfæri] er eins og að læra að hjóla*  
 that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
 ‘Learning to play an instrument is like learning to ride a bike.’ (Garofalo 2023:106)

The fact itself that this contrast exists suggests that a DP is required in positions like Spec,T (for subjects) independently of case, most probably due to the presence of a D-feature in T (see Chomsky (1995) and subsequent work within the *Minimalist Program*; see Chapter 4 in particular for a detailed discussion on this issue). Provided that this hypothesis is correct, then we can infer that the purpose of clausal nominalization consists specifically in checking features that a CP would be unable to check.<sup>54</sup> However, clausal nominalization in structural case positions does not have such a purpose, but it is rather a mere reflex of DP selection, as shown in (113).

My proposal, as it is presented here, is broadly a mixture of the DP selection hypothesis and the Case-marking hypothesis, but it also differs from them to some extent. Firstly, it diverges from both hypotheses because it proposes a rule that encompasses and explains pronominal optionality, which appears to be the most common trait of Icelandic clausal nominalization (and which the other hypotheses<sup>55</sup> could not account for). Secondly, even if my hypothesis follows

<sup>53</sup>These positions also include e.g. complements of nouns and adjectives as well as complements of lexical prepositions. We will discuss them in Chapter 3.

<sup>54</sup>After we will observe the data in Chapter 3, I will also claim, in light of the empirical data in Chapter 3 on postcopular clauses, that  $\varphi$ -features (gender and number) also trigger clausal nominalization. The data will, in fact, show that a non-NomCI is more frequent when the gender of the subject is neuter singular, whereas a NomCI becomes more frequent as a general rule as soon as the gender or the number of the subject with which the CN agrees is non-default.

<sup>55</sup>It is true, though, that Thráinsson (1979) tried to explain pronominal optionality with verbs like *spá* ‘predict’ in contrast with verbs like *fagna* ‘rejoice’, and correlated it to the fact that *spá* is a non-factive. However, the factivity hypothesis, as we have already seen, does not cover the whole distribution of Icelandic clausal nominalization,

Roussou (1991) regarding the idea that *það* checks case due to CRP, it still diverges from it by viewing structural case as a case that does not correspond to a formal feature to check in syntax. In other words, my hypothesis follows CRP only when a case feature needs to be checked, e.g. when lexical case is assigned.

It is true, however, that, as such, this hypothesis still has to account for other issues, in particular the contrast presented by Thráinsson (1979) between verbs like *fagna* ‘rejoice’ and *spá* ‘predict’, where the latter verb is followed by an optional *það* even if lexical case is assigned. We will discuss these issues and further implications in the following chapters in light of the empirical data in Chapter 3. For the moment, it is important to present to the reader a preliminary view of my hypothesis on Icelandic clausal nominalization on the basis of our discussion in the current chapter.

## 2.6 Summary

In this chapter, we have looked at the main facts about Icelandic clausal nominalization across syntactic positions. The distal demonstrative *það* has a very extended distribution as CN and it is found in various types of clauses, i.e. *that*-clauses, infinitives and indirect questions. The distribution of clausal nominalization is complex, as *það* is sometimes optional and sometimes mandatory on the basis of its position and also on the basis of the item selecting a clausal complement. I have proposed a new preliminary hypothesis to explain the distribution of NomCIs and non-NomCIs, in which pronominal optionality is specifically accounted for. Pronominal obligatoriness is explained by virtue of the fact that a) the clausal complement has to check case, or b) it surfaces in a position where a DP is required. Pronominal optionality, on the other hand, is caused by a mechanism based on which a NomCI is automatically allowed in a DP position. In addition, *það* is ungrammatical when extraction takes place (see Thráinsson (1979)), or in positions where a DP is not allowed (as we have seen with clausal complements of verbs like *þvinga* ‘force’ and *hjálpa* ‘help’).

Icelandic aligns with other cross-linguistic analyses as NomCIs surface as DPs and non-NomCIs as CPs (see e.g. Roussou (1991); Delicado Cantero (2013); Rosenbaum (1967); Borsley and Kornfilt (2000) among others). The main arguments for this are that a) *það* is disallowed when an item is extracted from a clausal argument, and b) *það* is unable to surface in non-DP positions. As for the distribution of NomCIs and non-NomCIs, various key facts show that Icelandic clausal nominalization is incompatible with different analyses that have been presented in the literature (further theoretical discussion about the distribution as well as the derivation of NomCIs and non-NomCIs is offered in Chapter 4). For instance, the CN neutralizes all possible semantic differences among verbs selecting a clausal complement, which contrasts with other languages like English (see e.g. Kastner (2015)). Factivity (see Thráinsson (1979)) cannot predict how NomCIs

---

especially due to the presence of different exceptions that we have already examined. Therefore, pronominal optionality in Icelandic has not been successfully accounted for in the linguistic literature so far.

and non-NomCIs are distributed across syntactic positions. Pronominal optionality, on the other hand, cannot be explained by hypotheses that consider DP selection or case-marking as the trigger of clausal nominalization (consider Roussou (1991); Knyazev (2016); Borsley and Kornfilt (2000); Kornfilt and Whitman (2011b); Kim and Sag (2005) among others), either because these cross-linguistic accounts rather focus on instances where clausal nominalization is mandatory or ungrammatical (see e.g. Roussou (1991)), or because any explanation offered for pronominal optionality is incompatible with the Icelandic data (cf. e.g. Kim and Sag (2005)).

Stronger empirical underpinnings, however, are still needed to develop my preliminary hypothesis further. One of the main reasons is that, with the advent of corpora, it has become possible, and thus highly desirable, to gather naturalistic data for all sorts of linguistic inquiries. In the case of our analysis of clausal nominalization, the *Icelandic Gigaword Corpus* (see Steingrímsson (2019); also Chapter 3) offers a wide sample of data that can be used to determine whether *það* is mandatory or not in a certain position and after a certain lexical item. Relatedly, due to the fact that I am not a native speaker of Icelandic, gathering data from native speakers on the distribution of NomCIs and non-NomCIs, whether through corpora or judgment tasks, is necessary to ensure the validity of my hypothesis or to refine it.

What is more, the distributional puzzle of *það* summarized in Table 2.3 calls for further empirical data, as optional as well as obligatory nominalization are found with lexical items of the same category: consider, for instance, the contrast mentioned by Thráinsson (1979) about *fagna* ‘rejoice’ and *spá* ‘predict’, or the one between prepositions like *hjá* ‘by’ and *yfir* ‘over’ where only the latter selects *það* optionally. It is not clear at this stage why these items differ in selecting a NomCI optionally or obligatorily and, apart from analyses like Garofalo (2020; 2023), which are preliminary studies at the basis of this doctoral dissertation, the various contributions on Icelandic clausal nominalization in the literature (Thráinsson (1979), Garofalo (2015) and Ingason (2018)) are not based on enough empirical data to address these issues about the distribution of *það*. The same can also be said about the mysterious contrast we observed between extraposed and non-extraposed sentential subjects after the finite verb, where *það* can only be optional with the former. Therefore, it is necessary to collect a great number of new empirical data from all syntactic positions where *það* can emerge, so that we can develop a strong and extensive basis to correctly map the distribution of NomCIs and non-NomCIs in Icelandic and explain why only certain lexical items appear to select a NomCI obligatorily while others allow *það* optionally, as well as why extraposition causes *það* to be optional with clausal subjects after the finite verb. These empirical data will be discussed in Chapter 3.



# 3 Empirical analysis

## 3.1 Introduction

This chapter lays the empirical foundation for this dissertation, where we will observe and discuss naturalistic data on clausal nominalization as well as judgments from native speakers of Icelandic. So far, we have explored from a theoretical perspective the properties of clausal nominalization and we have taken a closer look at the previous hypotheses that have been presented to explain the phenomenon, both in relation to its distribution as well as the structure of NomCIs and non-NomCIs (which emerge in Icelandic as DPs and CPs respectively). In the literature, we have not seen any hypothesis that can clearly explain by itself the whole distribution of Icelandic clausal complements with and without *það*. Consequently, I have broadly outlined a new possibility of explaining clausal nominalization in Icelandic, which consists of two main rules (they have already been presented in Chapter 2):

- (1) *NomCI selection mechanism*  
A NomCI is always allowed if it is in a DP position
- (2) *Non-NomCI filter*  
A NomCI is required if (it is enough to fulfill one condition):
  - a) a case feature must be checked by the relevant clause
  - b) the relevant clause is in a position in which a DP is required

In this chapter, I will elaborate on (1) and (2) with the support of empirical data and present a more complete hypothesis (albeit not final; this hypothesis will be revisited in Chapter 4) on the role and distribution of *það*, which considers feature checking as the main trigger of clausal nominalization as well as DP selection (cf. Garofalo (2023)):<sup>1</sup>

- (3) *Feature Checking Hypothesis (FCH)*  
A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and it is mandatory or at least preferable if:
  - a. it checks at least one D-,  $\varphi$ - or case feature
  - b. a DP complement is required

If conditions (a) and (b) are not met, CN is preferably absent

In support of this hypothesis, the empirical data will display, first of all, recurrent preferability

---

<sup>1</sup>Although DP selection is also a triggering factor of clausal nominalization, I called the hypothesis *Feature Checking Hypothesis* to make things easier for the reader. But there is also another reason for this choice. We will see in Chapter 4 that DP selection can also be considered a form of feature checking if we assume, following Heck and Müller (2007), that certain features of an item can trigger Merge of a specific category, like D.

patterns across syntactic positions: in nominative and accusative case positions (except Spec,T and Spec,Appl), a non-NomCI is generally the most preferable variant even if both are acceptable, while a NomCI tends to be more preferable when the NomCI receives dative or genitive case, as well as in contexts where both variants are possible. This, interestingly, also applies to complements of verbs like *spá* ‘predict’ (as well as *fresta* ‘postpone’) which displayed mysterious characteristics in Chapter 2: in fact, on the basis of the data from Thráinsson (1979), this verb is supposed to be followed by *það* optionally despite the fact that it assigns dative case. The data presented here will show that *það* is highly preferred with *spá*.

Secondly, the data from post-copular clauses will show that gender and number features also affect the distribution of *það*. If the subject of the matrix clause is neuter singular, a non-NomCI is frequent. But, interestingly, if the gender or number of the subject differs, a NomCI will be more likely to surface. This indicates that  $\varphi$ -features also need to be included into the set of conditions that determine the distribution of *það* (they were not accounted for in (1) and (2)).

Thirdly, apart from Spec,Appl and Spec,T, where NomCIs are mandatory (except those copula+adjective constructions discussed in §2.5.2 where the CP looks extraposed), the empirical data will show that complements of nouns and adjectives as well as complements of lexical prepositions are generally characterized by mandatory nominalization, independently of case. This indicates that these lexical items can only subcategorize for DPs. There are, however, exceptions, like the emotive adjective *feginn* ‘satisfied’, which only optionally subcategorizes for a DP complement in a default context, unlike other adjectives that require a DP complement.

This chapter is divided into three main subsections. In the first subsection, I will briefly present the main methodology applied to the empirical data (more technical information is also provided in Appendices A and B). In the second subsection, we will discuss the data I retrieved on clausal nominalization and argue for FCH in light of the empirical results. In the third, the main findings of this chapter will be summarized.

## 3.2 Methodology

### 3.2.1 Clausal nominalization as a syntactic variable

On the basis of the data presented in the previous chapter, we have often seen that NomCIs and non-NomCIs are able to occupy the same syntactic position (in particular in instances of pronominal optionality) and that they are also characterized by semantic equivalence, as shown in the following examples:

- (4) a. Nú stýttist í (**það**) að flautað verði til leiks á Laugardalsvelli  
 now shortens in that<sub>ACC</sub> that whistled is for game on Laugardalsvöllur  
 ‘It is not long now until the starting whistle for the game in Laugardalsvöllur.’
- b. Allardyce er svekktur yfir (**því**) að hafa fengið sparkið  
 Allardyce is annoyed over that<sub>DAT</sub> to have got kick.the  
 ‘Allardyce was annoyed about the fact that he was fired.’

- c. Ég harma (**það**) að ég sé ekki búinn að senda umsóknina  
 I regret that<sub>ACC</sub> that I am not finished to send application.the  
 'I regret the fact that I did not send my application.'

Assuming the analysis in Labov (1966) on syntactic variation, if we consider syntactic variants within a specific syntactic variable as different ways of saying the same thing (Chambers and Trudgill (1980); also see Pütz et al. (2014) and Cornips and Gregersen (2016) in relation to the concept of the *synonymy principle*), then NomCIs and non-NomCIs represent two variants of the same syntactic variable.<sup>2</sup>

The context in which the variants involved emerge is defined as the *envelope of variation* (cf. Aaron (2010)). Now, before gathering data on NomCIs and non-NomCIs, we need to have a clear idea of what to include in the envelope of variation and what to exclude from it. If, on the one hand, we can easily include cases of pronominal optionality due to the fact that NomCIs and non-NomCIs in examples like (4) are characterized by semantic equivalence and are both acceptable, instances of pronominal obligatoriness, on the other hand, should not have to be included, as there is only one structure available that can surface. However, in this regard, it is also true that almost no empirical study of the distribution of clausal nominalization in Icelandic (whether with judgment tasks or naturalistic data) exists in the linguistic literature.<sup>3</sup> In addition, sociolinguistic studies have pointed out the importance of looking beyond the envelope of variation in order to obtain a more complete picture of a specific linguistic phenomenon (see e.g. Aaron (2010); Dinkin (2016) among others). Therefore, I believe it is important to empirically verify the obligatoriness of clausal nominalization with e.g. clausal direct objects that are assigned lexical case or clausal complements of nouns and adjectives, by testing their non-NomCI counterparts as well. If the results exclude the possible existence of non-NomCIs in these cases, then we have at least a wide sample of empirical data that confirms the obligatoriness of *það* in certain syntactic positions. But if, by contrast, pronominal obligatoriness is not confirmed by the empirical data in certain positions where *það* is expected to emerge, then we would discover something new that was not captured by the data presented in the linguistic literature on Icelandic.

As for pronominal ungrammaticality, we have already seen in Chapter 2 that it occurs: a) when an item is extracted from a CP complement; and b) when a clausal complement surfaces in a non-DP position. These two contexts cannot be included in our envelope of variation as they are caused by constraints related to movement as well as ungrammatical DP selection (recall that the distribution of NomCIs is subsumed within the distribution of DPs, see §2.4.4). However, we

<sup>2</sup>However, it is true, if we follow Thráinsson (1979), that *það* can play the role of a discourse anaphor, as well as with matrix verbs which, based on Thráinsson's judgment, do not require a context for CN *það* to appear. This means, by extension, that every instance of clausal nominalization could potentially become an instance of anaphoricity, given the presence of a previous context *það* refers to. To the reader, this might perhaps represent a problem for identifying actual instances of clausal nominalization. But, as we will see, at least in the case of judgment tasks and interviews, the instances of *það* represent actual examples of clausal nominalization, as no context has been given to participants.

<sup>3</sup>The only exception in this regard is Garofalo (2023), but it just contains a preliminary version of the data presented in this dissertation.

can still verify within our envelope of variation whether there are any other instances of NomCIs in DP positions that are considered unacceptable by speakers. If such findings emerged, they would indicate that we need to refine or reject the *NomCI selection mechanism* described in (1). But if the data showed that NomCIs are systematically grammatical in DP positions, then these results would corroborate the validity of (1) for Icelandic.

Now that we have broadly defined our envelope of variation for our analysis of NomCIs and non-NomCIs, let us take a step forward in our discussion on syntactic variation. From a language acquisition perspective, the variants that we can find in a syntactic variable can be the result of the existence of competing grammars that speakers develop during the process of first language acquisition (I assume here the analysis proposed by Kroch (1989) and also Yang (2000)). More specifically, first language learners try to analyze the linguistic input through a set of grammars that constitute their learning space in the process of language acquisition. Grammars that are able to analyze those sentences that emerge from the linguistic input are rewarded and tend to be more frequent in the linguistic production of children, while grammars that are unable to analyze the sentences produced in the linguistic input are punished and become less frequent (see Yang (2000) for a more detailed discussion on this model as well as a review of other approaches that analyze linguistic variation). These competing grammars can be found in a single individual as a form of *intra-speaker variation*, as the speaker can choose a specific grammar over other ones during linguistic performance. But if we transpose the same principle on a wider scale, within a group of individuals or in society in general, grammars can also compete at this level as well, i.e. as a form of *inter-speaker variation*. Now, although it is a very interesting task to explore variation in clausal nominalization in the grammar of single individuals, given the extension of this variable in Icelandic across syntactic positions (see Chapter 2), it can be difficult to gather data on the whole distribution of NomCIs and non-NomCIs in the grammar of single individuals, as it would require the administration of various tests to the same sample of native speakers and it is not guaranteed that participants would be always willing or available to do that. Therefore, I found it more practical for this analysis to investigate inter-speaker variation, as one can at least gather a greater number of native speakers and test different syntactic positions with different samples of participants.

### 3.2.2 Two types of data

In order to gather clues about clausal nominalization in the grammar(s) of speakers, I collected two different types of data. On the one hand, I retrieved examples of NomCIs and non-NomCIs from the *Icelandic Gigaword Corpus* or *Risamálheild* (henceforth RMH; see Steingrímsson (2019)).<sup>4</sup> On the other hand, I collected judgments on examples of NomCIs and non-NomCIs obtained through online questionnaires that have been administered to various samples of native speakers of Icelandic.

---

<sup>4</sup>The corpus is accessible at <http://malheildir.arnastofnun.is/>. I gathered data using the 2019 version of RMH, before the 2022 version was released.

There are various reasons why I chose to gather two main different types of data. The first reason is related to what we can observe in corpora and judgment tasks. The data from questionnaires, on the one hand, can reveal the degree of acceptability of the relevant variants and offer some insight into linguistic competence (cf. Lau et al. (2017)). By searching for naturalistic data, on the other hand, we can compare the frequency of NomCIs and non-NomCIs in linguistic performance and assess how common these variants are across syntactic positions. Although it is true that the data from judgment tasks are mainly those we can observe in order to gather clues about the grammar of speakers, naturalistic data can complement our analysis and establish a higher empirical validity of the results (this idea is also supported by other studies like e.g. Bever (1970) and Hoffmann (2006) among others). In addition, it has also been observed in certain studies that there is a correlation between data from corpora, provided that the proportional frequency of variants is observed,<sup>5</sup> and judgment tasks, although this relationship is not straightforward due to the fact that we cannot expect proportions in the data from corpora to correspond precisely to values derived from acceptability judgments (see in more detail Bermel and Knittl (2012)). In the case of our study, the choice of two different types of data has been particularly useful to investigate the distribution of NomCIs and non-NomCIs, as we will see in the following subsections, due to the fact that naturalistic data and judgments from native speakers display differences that are recurrent across syntactic positions. I believe these contrasts can cast better light on the purpose and distribution of clausal nominalization.

Moreover, a further reason in support of this methodology is that both sources of data have some issues that emerge when taken on their own. It is commonly known that judgment tasks are problematic in several respects as a method for testing the grammatical competence of native speakers, e.g. because the speakers may be biased towards giving the ‘correct’ answers or answers that they think the researcher is hoping for (cf. e.g. Schütze (1996); Culbertson and Gross (2009); Dąbrowska (2010); Práinsson et al. (2013) among many others). On the other hand, naturalistic data are also insufficient on their own as it is known that corpora cannot provide negative data (see e.g. Horsch (2023)) and may not contain enough examples of the relevant type (take the case, for example, of clausal indirect objects, which are uncommon in Icelandic), and also because certain examples that emerge from corpora might not necessarily be considered acceptable in judgment tasks (cf. Hoffmann (2006)). Thirdly, the items that compose the questionnaires have to appear as natural as possible to native speakers. Therefore, naturalistic data from corpora are also a precious source of information to create the sentences that need to be judged. In sum, using both types of data is pivotal to ensure that our linguistic analysis is empirically valid.

### 3.2.3 Corpora

Now that we have discussed the importance of having two types of data, let us first discuss briefly the set of naturalistic data used in this study. RMH is a linguistic corpus of Modern Icelandic

---

<sup>5</sup>Absolute frequency of variants in corpora appears to have a less strong effect on predicting acceptability judgments (see in particular Bermel et al. (2018)).

that contained around 1.67 billion words at the time when I collected all the data,<sup>6</sup> and is made up of texts from Parliament speeches, novels, local and national news as well as blogs and other websites. Each word of the corpus has been grammatically tagged, so that it is possible to look for words and also sequences of words with specific grammatical features. The interface available on the RMH website allows the user to look for sequences of lemmas, word forms and/or grammatical features through the *Extended* tab, which is illustrated in Figure 3.1.<sup>7</sup>

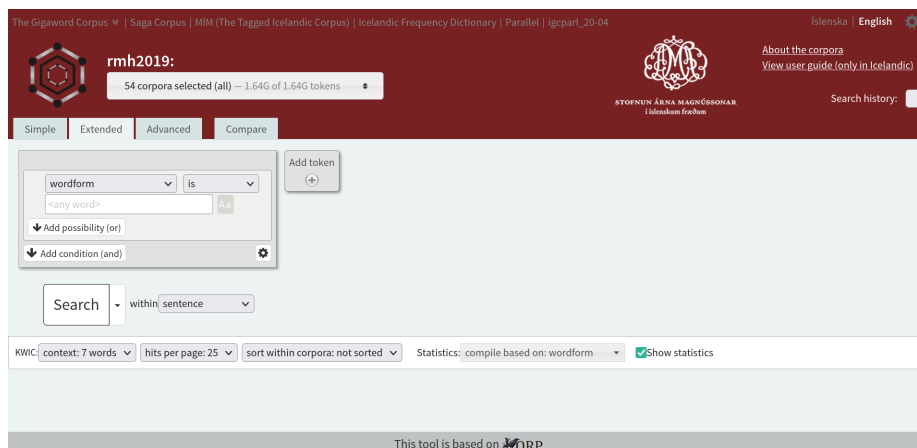


Figure 3.1: Extended tab in RMH (2019)

In RMH, I looked for examples of NomCIs and non-NomCIs preceded by specific lexical items, like selected verbs, prepositions, adjectives, nouns, etc. In general terms, most RMH queries targeting examples of NomCIs and non-NomCIs have been composed as in the example shown in Figure 3.2:

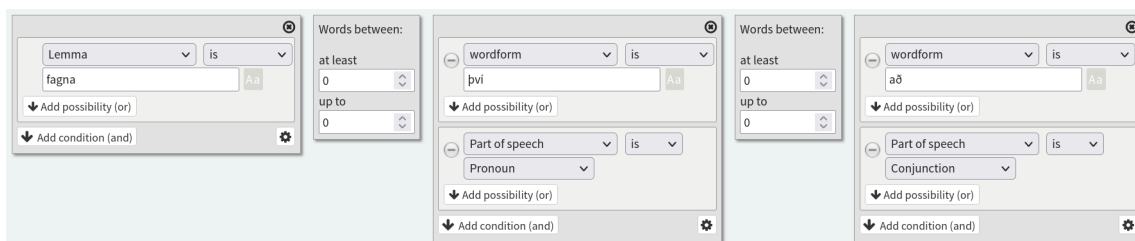


Figure 3.2: Example of query (*fagna* + dative *það* + *that*-clause)

The first element I added into the query was generally the item that selects the clausal complement, which can be, for example, a verb or an adjective.<sup>8</sup> At times, there were more lexical items to specify, as in the case of prepositional complements, where I looked for specific verbs or adjectives which are followed by a specific preposition. When taking a look at clausal complements of verbs, nouns or adjectives, I searched the lemma of the relevant lexical item, so that, with

<sup>6</sup>The current version of RMH contains around 2.5 billion words.

<sup>7</sup>The interface of RMH was updated and modified in April 2023. Figure 3.1 shows what it looked like before the update.

<sup>8</sup>In the case of sentential subjects in the first position, I could not rely on the presence of a specific item preceding the clausal argument, so I needed to construct the query in a different way (this is discussed in detail in §A.2.1).

the query, all the tokens containing the relevant word forms of the item itself could be found. In the example shown in Figure 3.2, the lemma specified for the first lexical item is the verb *fagna* ‘rejoice’.

The next element I added to the query is *það* in its relevant word form whenever I had to observe instances of clausal nominalization, otherwise I did not add it. I tried to make sure *það* is recognized as a pronoun, and not, for instance, as an adverb, by simply adding as a condition that *það* must be a pronoun only. In fact, the word form *því* is not restricted to being a CN in dative case but also an adverb (with the meaning ‘therefore’; see Figure 3.2).

The next element I added to the query is the first word of the CP such as a conjunction, an infinitive marker or an interrogative pronoun. Considering, for instance, that *að* functions both as an infinitive marker and as a conjunction, we need to specify its function accordingly.<sup>9</sup> For every query, I increased up to 20 the number of words before and after the tokens we are looking for (see Figure 3.1, in particular the KWIC bar). In this way, it was easier to take a look at more context in the tokens found by the query.

As a general rule, for each item selecting clausal complements that I investigated, I applied 6 queries on the basis of the clause type. Here is an example with the verb *krefjast* ‘demand’, which is followed by a genitive CN. For each query, the raw number of hits is shown, followed by the number of valid examples:

- (5) *krefjast* + *þess* + *að* (conjunction) > NomCl *that*-clauses (32667 hits, 28563 valid)  
*krefjast* + *þess* + *að* (infinitive marker) > NomCl infinitives (4213 hits, 3790 valid)  
*krefjast* + *þess* + interrogative pronoun > NomCl embedded questions (0 hits, 0 valid)  
*krefjast* + *að* (conjunction) > non-NomCl *that*-clauses (8118 hits, 6677 valid)  
*krefjast* + *að* (infinitive marker) > non-NomCl infinitives (143 hits, 90 valid)  
*krefjast* + interrogative pronoun > non-NomCl embedded questions (13 hits, 0 valid)

The syntactic positions I tested in RMH are those where a DP can appear (as we said in Chapter 2, non-DP positions cannot host NomCl as the distribution of *það* is subsumed within the distribution of DPs). However, I could not test clausal indirect objects due to the fact that they are quite rare and difficult to look for in RMH.

In Appendix A, the reader can find a detailed description of specific queries that were more difficult to construct to find examples (in particular, the ones for clausal subjects) as well as the criteria applied in order to exclude tokens that cannot be considered valid (e.g. duplicates, tokens in which no actual clausal complement emerges, etc.). All the data obtained from RMH have also been saved online in spreadsheet documents, which the reader can consult. The link to access them as well as a detailed description of these spreadsheets is also offered in Appendix A.

<sup>9</sup>The clausal complements of *spá* ‘predict’ were the very first I analyzed at the beginning of my research. While looking for items, I did not apply there the distinction between *að* as a conjunction or infinitive marker, so that the set of data retrieved was actually mixed. However, I went over all the items I downloaded from RMH so that they are classified based on the correct clause type following the verb.

### 3.2.4 Judgment tasks

Let us move to the second type of data, which consists of judgments on various sentences from native speakers of Icelandic, obtained through online questionnaires. The questionnaires were administered to speakers of various age groups. Each questionnaire was advertised in selected Facebook groups, other social networks, on the intranet of the University of Iceland (*UGLA*) as well as through the Student Registration (*Nemendaskrá*) of the University of Iceland, which shared the links of the questionnaires with different departments.

Similarly to the data from RMH, in the judgment tasks I tested DP positions specifically, but, this time, clausal indirect objects were also included (although clausal indirect objects were also tested with oral interviews, as both variants received few positive judgments in the online questionnaires; see §3.3.7 for more details). Moreover, given that the relevant sentences I needed to test had to be as natural as possible for participants, I decided to use (and at times adapt) valid tokens I obtained from the multitude of data I found in RMH on NomCIs. A further reason behind this decision is the fact that, as a non-native speaker of Icelandic, I could not create examples by myself exclusively (for clausal indirect objects, I requested assistance from native speakers to construct the examples).

Constructing these questionnaires presented some challenges due to various factors. Firstly, the significant number of items to test posed a problem, as we have seen in Chapter 2 that clausal nominalization can take place in various syntactic positions and that *það* can precede *that*-clauses, infinitives and embedded questions. Secondly, once again in relation to the number of items to test, I had to make sure that the questionnaires did not contain too many items, so that it would be easy for participants to respond in the span of few minutes. Thirdly, for the sake of ensuring empirical validity, every sentence that needed to be tested in the questionnaires had to appear in two variants, i.e. with and without *það* respectively. But the difference between these two versions of the same sentence can be easily seen by participants if both items appear in the same questionnaire and, consequently, they might figure out that the CN is exactly the element I am investigating, independently of the fact that there might be filler sentences in between.

Considering these issues, I decided to administer 10 questionnaires, each of them in two versions, A and B, i.e. a total of 20 short questionnaires. I made sure that the two variants of a specific sentence never emerged in the same questionnaire, so that it would be difficult for participants to figure out that *það* is the element I am researching. In other words, every single participant judged only one variant (which is not highly problematic if we specifically observe inter-speaker variation). The relevant tokens with and without *það* covered all argument and complement positions discussed in Chapter 2, with a maximum of 6 tokens for each category (e.g. accusative direct objects, dative prepositional objects, postcopular clauses, etc.). Each of the questionnaires contained 25 sentences to be evaluated on a scale from 1 to 7, where two thirds were filler sentences. The sequence of relevant tokens displayed a constant alternation of NomCI and non-NomCI tokens. More information on the structure of the questionnaires as well as

some general data on participants and statistical tests that were performed is offered in Appendix B.

### 3.2.5 Preliminary guide to the analysis of results

Before we start looking at the results, it is important to briefly discuss how we will interpret the data from the questionnaires and corpora. Let us focus first on the data from the questionnaires, as they offer insight into the grammar of speakers. In order to study how *pað* is distributed, I compared the judgments attributed to the NomCI and non-NomCI variants in a certain syntactic position, and took into account two factors in particular: a) whether the weighted averages of the relevant judgments are similar or not; and b) whether there is a significant statistical difference between the judgments attributed to the relevant variants. By following this approach, I tried to map pronominal obligatoriness, optionality and ungrammaticality in the following way:

(6) *Hints suggesting pronominal obligatoriness*

Data from questionnaires: as a main trend, the weighted average of the relevant judgments on the NomCI variant are more positive than the one of the non-NomCI variant and there is, at least for the majority of the relevant pairs of examples, a significant statistical difference (i.e.  $p < .05$ ) between the sets of judgments on the NomCI and the non-NomCI variant

(7) *Hints suggesting pronominal ungrammaticality*

Data from questionnaires: as a main trend, the weighted average of the relevant judgments on the non-NomCI variant are more positive than the one of the NomCI variant and there is, at least for the majority of the relevant pairs of examples, a significant statistical difference (i.e.  $p < .05$ ) between the sets of judgments on the NomCI and the non-NomCI variant

(8) *Hints suggesting pronominal optionality*

Data from questionnaires: as a main trend, there is, at least for the majority of the relevant pairs of examples, no significant statistical difference (i.e.  $p > .05$ ) between the sets of judgments on the NomCI and the non-NomCI variant, and the weighted averages of the judgments on NomCIs and non-NomCIs respectively are relatively similar

For the sake of clarification, let us observe an example of how the results from the questionnaires will be presented in this chapter:

| No.   | Sentence              | Type | - <i>pað</i> | + <i>pað</i> | p     |
|-------|-----------------------|------|--------------|--------------|-------|
| 1A-XX | <i>Item 1</i>         | inf  | 2.10         | 6.30         | <.001 |
| 1B-XX | 'English translation' |      | (n)          | (n)          |       |
| 2A-XX | <i>Item 2</i>         | tht  | 2.20         | 6.90         | <.001 |
| 2B-XX | 'English translation' |      | (n)          | (n)          |       |
| 3A-XX | <i>Item 3</i>         | que  | 1.60         | 5.90         | <.001 |
| 3B-XX | 'English translation' |      | (n)          | (n)          |       |

Table 3.1: Example of results from the questionnaires

Table 3.1 shows, starting from left, the ID number of a pair of examples (the number of the questionnaire is shown first, e.g. 1A, and then the number of the item, which can be from 01 to 25; see Appendix B), the sentence tested as well as its English translation, the type of clause that the sentence tested contains (*inf* for infinitives, *tht* for *that*-clauses and *que* for indirect questions) and then the weighted average of the judgments on the non-NomCI variant (*-pađ*) as well as the weighted average of the judgments on the NomCI variant (*+pađ*). Note that the minimum possible value for the weighted average is 1.00 and the maximum 7.00, as the judgment values, as already mentioned, are on a scale from 1 to 7. Moreover, under the weighted averages, the number of participants who answered is specified between brackets (see (*n*)). Lastly, in the *p* column it is shown whether there is a significant statistical difference between the judgments given to the non-NomCI and NomCI variant respectively. If we imagine the results presented in the table to be actual results, the weighted average of judgments on NomCIs is, as a general trend, much higher than the one for non-NomCIs, and the *p*-values are also  $< .01$ ,<sup>10</sup> which support the idea that *pađ* is mandatory in the set of sentences involved.

In relation to the degree of acceptability of both variants, an important question is how we can determine whether a certain example is broadly acceptable or not. What we can do is to set a minimum threshold beyond which an example can easily be considered acceptable. For our study, I followed the guidelines on Likert scale analysis presented in Pimentel (2019), who set 4.44 as the minimum average for an item to be part of the positive side of the spectrum on a scale from 1 to 7. Therefore, I will consider all examples that display an average beyond 4.44 as acceptable.<sup>11</sup> However, this does not necessarily mean that examples with an average below this threshold have no information to provide or that we have to ignore them completely.<sup>12</sup> Degraded examples can still offer us some useful information to determine the pattern behind the distribution of NomCIs and non-NomCIs (as we will also see in the following paragraphs).

Now that we have discussed the data from questionnaires, let us focus on what role the data from corpora can play in our analysis and how they can be correlated to the data from questionnaires. In this regard, I followed a similar approach to the one presented in Bermel and Knittl (2012),<sup>13</sup> who have attempted to compare naturalistic data and judgments from native speakers, and have shown that there is some correlation between these two types of data.<sup>14</sup> In particular, they mapped proportionally the frequency of variants in corpora with seven frequency bands, as shown in Table 3.2:

<sup>10</sup>Note that, whenever a *p*-value is lower than .001, as is the case in the fictitious example I show in Figure 3.4, I will type it as  $< .001$  for simplicity.

<sup>11</sup>The results on nominative clausal subjects, direct objects (in active voice), postcopular clauses and indirect objects have also been published in Garofalo (2023). The data presented in this dissertation on these syntactic positions, however, have been updated and also take into account those examples where the weighted average is below 4.44.

<sup>12</sup>I differ here from the more conservative position I took in Garofalo (2023), where I considered the examples that got an average below Pimentel's minimum threshold as not acceptable.

<sup>13</sup>Note that, in this chapter, I have revisited my position presented in Garofalo (2023) by adopting the frequency bands proposed by Bermel and Knittl (2012).

<sup>14</sup>Relative frequency of variants expressed in percentage in naturalistic data appears to have a stronger correlation to the results from judgment tasks compared to absolute frequency (see Bermel et al. (2018)).

| <b>Isolated</b> | <b>Marked</b> | <b>Minority</b> | <b>Equipollent</b> | <b>Majority</b> | <b>Unmarked</b> | <b>Dominant</b> |
|-----------------|---------------|-----------------|--------------------|-----------------|-----------------|-----------------|
| 0–1%            | 1–9%          | 10–29%          | 30–69%             | 70–89%          | 90–99%          | 99–100%         |

Table 3.2: Frequency bands for competition between variants in Bermel and Knittl (2012)

The area that, in their proposal, can be considered close to the one of optionality is the *Equipollent* band, which is the broadest and extended to 30–69%. The bands *Minority/Majority*, which are narrower, suggest that one variant is more frequent. *Marked* and *Unmarked* are labels based on Halliday’s (1991) idea that the ratio 9:1 can account for markedness or unmarkedness of a certain form, while *Isolated* and *Dominant* represent the extremes of the scale and display a very narrow extension.

In the case of our study, I integrated the results from corpora with the ones from the questionnaires in the following way. In relation to pronominal obligatoriness, if the frequency of *pað* reaches, say, 95% while the results from the questionnaires indicate that the NomCl is the only acceptable variant, the data from corpora will corroborate the idea that *pað* is mandatory in a certain syntactic position. Of course, a similar reasoning can be applied to pronominal ungrammaticality where a non-NomCl is the only acceptable variant. As for pronominal optionality, despite the fact that both variants can be considered equally acceptable, we cannot expect the data from corpora to do the same and constantly display a 50% frequency for every subset of data. Rather, as we are going to see, the data from corpora display a great variety of frequencies, but, interestingly, they tend to form a pattern in which NomCl’s are preferable in specific syntactic positions, while non-NomCl’s are preferable in others. Revisiting the frequency bands proposed by Bermel and Knittl (2012), I propose the following analysis of the data from corpora, in which I grouped the frequency bands into three main groups:

| <b>UNGRAMMATICALITY/LOW FREQUENCY</b> |               |                 | <b>OPTIONALITY</b> | <b>HIGH FREQUENCY/OBLIGATORINESS</b> |                 |                 |
|---------------------------------------|---------------|-----------------|--------------------|--------------------------------------|-----------------|-----------------|
| <b>Isolated</b>                       | <b>Marked</b> | <b>Minority</b> | <b>Equipollent</b> | <b>Majority</b>                      | <b>Unmarked</b> | <b>Dominant</b> |
| 0–1%                                  | 1–9%          | 10–29%          | 30–69%             | 70–89%                               | 90–99%          | 99–100%         |

Table 3.3: Revisitation of the frequency bands proposed in Bermel and Knittl (2012)

These groupings are a simple general guideline for us to corroborate the data from questionnaires and do not aim to strictly define whether *pað* is mandatory, ungrammatical or optional, as naturalistic data offer insight into what variant is more common in linguistic performance. To be clear, Table 3.4 shows an example of how naturalistic data are going to be presented:

| <b>GENERAL</b> |                  |                  |     |
|----------------|------------------|------------------|-----|
|                | <i>Dataset 1</i> | <i>Dataset 2</i> | ... |
| <b>Valid</b>   | 1000             | 1000             | ... |
| <b>-pað</b>    | 20               | 40               | ... |
| <b>+pað</b>    | 980              | 960              | ... |
| <b>%pað</b>    | <b>98.00%</b>    | <b>96.00%</b>    | ... |

Table 3.4: Example of results from corpora

The results from corpora include all the relevant datasets for one specific syntactic position, and for each dataset (which varies based on the lexical item that subcategorizes for the clausal complement) I illustrate the total number of valid examples (*Valid*) and then the ones containing non-NomCIs (*-pað*) and NomCIs (*+pað*). Then, I will also show the frequency of *pað* in percentage for each dataset. According to the example in Table 3.4, if we imagine the numbers to be actual results, the frequency of *pað* is above 90% in the datasets involved, which would suggest that *pað* is probably mandatory in the relevant syntactic position.

Lastly, it is important to keep in mind that the overall data that are going to be presented are quite complex and, regardless of the guidelines described above for the results from the judgment tasks and corpora, there will still be problematic cases that have been difficult to map. I will discuss these instances in more detail in the following subsections.

### 3.3 Data

#### 3.3.1 Subjects

##### 3.3.1.1 Nominative subjects

In Chapter 2, we discussed the contrast that we find between nominative subjects in the first position and after the finite verb. Recall that *pað* is generally optional when the whole clausal argument precedes the finite verb, while it is mandatory after it. The contrast is shown once more below for convenience:

- (9) a. (**pað**) að læra á hljóðfæri er eins og að læra að hjóla  
 that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
 ‘Learning to play an instrument is like learning to ride a bike.’ (Garofalo 2023:106)
- b. (**pað**) að hann sé farinn sannar að við vitum ekki enn þá allan sannleikann  
 that<sub>NOM</sub> that he is gone proves that we know not yet all truth.the  
 ‘The fact that he left proves that we do not know all the truth.’
- (10) a. Er **\*(pað)** að læra á hljóðfæri eins og að læra að hjóla?  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 ‘Is learning to play an instrument like learning to ride a bike?’ (Garofalo 2023:106)
- b. Sannar **\*(pað)** að hann sé farinn að við vitum ekki enn þá allan sannleikann?  
 proves that<sub>NOM</sub> that he is gone that we know not yet all truth.the  
 ‘Does the fact that he left prove that we do not know all the truth?’

The data from the questionnaires in Table 3.5 show that the NomCI variant is significantly more acceptable than a non-NomCI subject when it surfaces after the finite verb, although *that*-clauses appear to be more degraded compared to infinitives. I tested *that*-clauses twice to make sure that it was not a coincidence:

| No.   | Sentence                                                                         | Type | -það  | +það | p               |
|-------|----------------------------------------------------------------------------------|------|-------|------|-----------------|
| 4A-25 | <i>Er [það að læra á hljóðfæri] eins og að læra að hjóla?</i>                    | inf  | 3.64  | 5.62 | <b>.003</b>     |
| 4B-01 | 'Is learning to play an instrument like learning to ride a bike?'                |      | (36)  | (21) |                 |
| 3A-20 | <i>Breytir [það að ég sé ríkur] öllu því sem þú hélt um mig?</i>                 | tht  | 2.47  | 2.54 | .818            |
| 3B-06 | 'Does the fact that I am rich change everything you thought about me?'           |      | (126) | (35) |                 |
| 8A-23 | <i>Sýnir [það að Karl keypti nýjan síma] að hann er búinn að fá launahækkun?</i> | tht  | 1.60  | 2.50 | <b>&lt;.001</b> |
| 8B-03 | 'Does the fact that Karl bought a new phone that he got a raise?'                |      | (132) | (76) |                 |

Table 3.5: Questionnaires - Nominative subjects (not extraposed) after the finite verb

I am not quite sure why there is a difference between infinitives and *that*-clauses here, but it is a fact that the sentence without *það* is significantly more degraded (which is also demonstrated among *that*-clauses by the contrast between 8A-23 and 8B-03, where  $p = <.001$ ).<sup>15</sup> In the majority of cases, the items presented in Table 3.5 show that the weighted average of judgments on NomCIs is higher than the one related to non-NomCIs and  $p < 0.01$ . These results suggest that nominative clausal subjects after the finite verb are obligatorily nominalized.

The results from RMH on nominative subjects after the finite verb corroborate the idea that *það* is mandatory, as shown in Table 3.6, where the frequency of *það* almost reaches 93%:

| Valid | -það | +það | %það          |
|-------|------|------|---------------|
| 71    | 5    | 66   | <b>92.96%</b> |

Table 3.6: RMH - Nominative subjects (not extraposed) after the finite verb

Let us now compare these results with the ones on clausal subjects in the first position. Judgments from speakers are mostly in line with the idea that both variants are acceptable in the first position, as shown in Table 3.7:<sup>16</sup>

| No.   | Sentence                                                        | Type | -það  | +það | p           |
|-------|-----------------------------------------------------------------|------|-------|------|-------------|
| 2A-16 | <i>[það að lifa í núvitund] snýst um að njóta augnabliksins</i> | inf  | 6.41  | 6.08 | .522        |
| 2B-10 | 'Living in mindfulness is about enjoying the moment.'           |      | (107) | (26) |             |
| 3A-08 | <i>[það að vörurnar séu ekki enn komnar] skiptir engu máli</i>  | tht  | 4.51  | 5.34 | <b>.038</b> |
| 3B-18 | 'It does not matter that the products have not arrived yet.'    |      | (126) | (35) |             |
| 5A-11 | <i>[það hvort hann komi eða ekki] skiptir engu máli</i>         | que  | 4.20  | 4.91 | .423        |
| 5B-15 | 'It does not matter whether he comes or not.'                   |      | (20)  | (23) |             |

Table 3.7: Questionnaires - Nominative subjects (not extraposed) in first position

The results from corpora, on the other hand, reveal that NomCIs are less frequent in the first position than non-NomCIs. This is an important fact that we need to keep in mind, as this pattern

<sup>15</sup>Note that all the  $p$ -values where there is a significant statistical difference are marked in bold in all the tables showing the results from the judgment tasks.

<sup>16</sup>In Table 3.7, only in the case of 3A-08 and 3B-18, which contains a *that*-clause, is the non-NomCI variant slightly more degraded than NomCI, but it is still somehow acceptable if we consider the examples in 5A-11 and 5B-15, which have a similar weighted average.

(i.e. both variants are acceptable, but NomCIs are less frequent in language performance) is recurrent in nominative/accusative case positions:

| Valid | -það | +það | %það   |
|-------|------|------|--------|
| 9478  | 7047 | 2431 | 25.65% |

Table 3.8: RMH - Nominative subjects (not extraposed) in first position

Incidentally, I also found some other examples of nominative clausal subjects in other datasets from RMH, namely the ones on accusative clausal objects (see §3.3.2), where the clausal object is passivized and moved either to the first position or after the finite verb.<sup>17</sup> The datasets are *gagnrýna* ‘criticize’, *harma* ‘regret, be sorry’ and *opinbera* ‘disclose, reveal’. These passivized objects I collected, however, are generally extraposed. Therefore, I could not take into account passivized objects where *það* surfaces in the first position because it would be impossible to determine whether *það* is a CN or an expletive (recall our discussion on extraposed clausal subjects in Thráinsson (1979) in Chapter 2). As for Spec,T, I only considered those examples where *það* can surface after the finite verb. Since we know that the expletive can only occupy the first position of the sentence, we can consider *það* as a CN when we find it after the finite verb, as shown in the following examples from RMH:<sup>18</sup>

- (11) a. Á árunum fyrir hrun var það gagnrýnt að bankarnir réðu til sín  
 On years.the before crisis was that<sub>NOM</sub> criticized that banks.the hired to themselves  
 besta fólkið úr eftirlitinu  
 best people.the from supervision.the  
 ‘In the years before the crisis, banks were criticized for hiring the best people from financial supervision.’
- b. Eftir leik var það opinberað að Andri var kjörinn leikmaður ársins af  
 after game was that<sub>NOM</sub> revealed that Andri was elected player year.the from  
 þjálfurum  
 trainers  
 ‘After the game, it was revealed that Andri was elected player of the year by the trainers.’

Now, recall, based on what we have discussed in §2.3.1, that there is a contrast between examples of clausal subjects in Spec,T where *það* and its associate CP occupy the same position together and those where *það* and the CP are separated. In the case of the former, *það* is mandatory, but it is optional in the case of the latter:

<sup>17</sup>There is also one example in the *gagnrýna* dataset of the so-called Icelandic new passive (see e.g. Sigurðsson (2011) for some discussion on this construction). This example has not been counted in Table 3.9 as *það* is not moved from the direct object position.

<sup>18</sup>It is true, however, that Thráinsson (1979) has considered the possibility (see §2.5.2) that *það* is a discourse anaphor in subject position with predicates like *vera líklegt* ‘be likely’. But the situation here is different because we are dealing with passivized clausal objects where *það* is not necessarily interpreted by Thráinsson as a discourse anaphor (see in particular the case of *harma*, which is an emotive factive).

- (12) a. Er [**það**] að læra á hljóðfæri] eins og að læra að hjóla?  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 ‘Is learning to play an instrument like learning to go ride a bike?’ (Garofalo 2023:106)
- b. Er (**það**) hörmulegt [að Jón skuli hafa barið Maríu]?  
 is it deplorable that Jón shall have beaten María  
 ‘Is it deplorable that Jón has beaten María?’ (Thráinsson 1979:192, 194)

We expect this contrast to emerge in our empirical data as well. Our expectations are confirmed in Table 3.9, which displays the results from RMH on passivized clausal objects after the finite verb:<sup>19</sup>

| GENERAL      |                         |                      |                         |
|--------------|-------------------------|----------------------|-------------------------|
|              | <i>gagnrýna (pass.)</i> | <i>harma (pass.)</i> | <i>opinbera (pass.)</i> |
| <b>Valid</b> | 1232                    | 140                  | 174                     |
| <b>-það</b>  | 1011                    | 96                   | 162                     |
| <b>+það</b>  | 221                     | 44                   | 12                      |
| <b>%það</b>  | <b>17.94%</b>           | <b>31.43%</b>        | <b>6.90%</b>            |

Table 3.9: Passivized clausal objects in Spec,T - Tokens with extraposition

These results suggest that NomCIs are in general less common than non-NomCIs (the average of the relevant frequencies of *það* is 18.75%), similarly to what we observed with non-extraposed nominative clausal subjects in the first position (see Table 3.8).<sup>20</sup> To sum up, NomCIs as nominative clausal subjects tend to be less frequent than non-NomCIs, except when the whole clausal subject emerges in Spec,T. In that case, *það* is mandatory.

Further data on nominative clausal subjects can be found in Appendix A, where I observed in particular whether there is any specific difference in the frequency of NomCIs based on the matrix verb involved (recall that Thráinsson (1979) claimed that the meaning of the matrix verb can affect how subject and object NomCIs are distributed). None of the most common matrix verbs found in the data on clausal subjects in the first position has a frequency of *það* over 50%. The results, therefore, suggest that non-NomCIs are preferable as a general trend. As for non-extraposed clausal subjects after the finite verb, the data are mainly built upon the matrix verb *vera* ‘be’, while there are very few tokens where different matrix verbs emerge.

### 3.3.1.2 Subjects with lexical case

Now that we have observed nominative subjects, we can also take a closer look at clausal subjects with lexical case. I tested them in order to verify whether *það* is mandatory both before and after

<sup>19</sup>Among the data from RMH, I also gathered some examples of nominative clausal subjects where the matrix verb is in middle voice. They confirm the pattern we have seen, as the frequency of *það* is 31.03% with clausal subjects in first position and 100% with clausal subjects in Spec,T (see Appendix A).

<sup>20</sup>The *opinbera* dataset, however, displays a very low frequency of *það*, which can make us wonder whether nominalization can even be considered ungrammatical with this matrix verb. The data on active voice with the verb *opinbera*, however, are not problematic, as the frequency of *það* reaches 21.21% (see §3.3.2). I believe, therefore, that these results on passive voice only incidentally display a very low frequency of *það*.

the finite verb or whether there is a difference between them, as with nominative subjects. If clausal subjects with lexical case are obligatorily NomCIs in first position, we can infer that there is a difference between structural and lexical case as to whether they trigger nominalization or not, in line with the *non-NomCI filter* I proposed in (2).

I focused on a specific predicate that assigns genitive case to its subject, i.e. *verða vart* ‘be noticed’ (with this verb, I could only test *that*-clauses as I could not find examples of infinitives and indirect questions). Unfortunately, the only examples I could find in RMH of this predicate featured clausal extraposition (compare the examples below):

- (13) a. **Þess** varð ekki vart að hann væri farinn  
 that<sub>GEN</sub> became not noticed that he was gone  
 ‘It went unnoticed that he left.’
- b. Varð **þess** ekki vart að hann væri farinn?  
 became that<sub>GEN</sub> not noticed that he was gone  
 ‘Did it go unnoticed that he left?’

Hence, for judgment tasks, I decided to test examples characterized by extraposition as well. The results are shown in Table 3.10, where the sentence is significantly degraded if *þess* does not surface (although the NomCI variants are also not highly acceptable as their weighted averages are below the minimum threshold; nevertheless, they are deemed better than their non-NomCI counterparts). Also notice that there is always a highly significant statistical difference between the judgments given to both variants:<sup>21</sup>

| No.   | Sentence                                                                          | Type | -það  | +það  | p     |
|-------|-----------------------------------------------------------------------------------|------|-------|-------|-------|
| 8A-17 | <b>Þess/Það</b> <i>hefur orðið vart að ökumenn keyra utan leyfilegra svæða</i>    | tht  | 1.97  | 4.00  | <.001 |
| 8B-09 | ‘It was noticed that drivers drive outside permissible areas.’                    |      | (132) | (76)  |       |
| 7A-17 | <b>Þess/Það</b> <i>varð ekki vart að hann væri farinn</i>                         | tht  | 1.57  | 3.81  | <.001 |
| 7B-09 | ‘The fact that he left went unnoticed.’                                           |      | (23)  | (26)  |       |
| 8A-02 | <i>Líklega</i> varð <b>þess</b> <i>ekki vart að vefsíðan var ekki lengur virk</i> | tht  | 1.42  | 2.99  | <.001 |
| 8B-24 | ‘Probably it went unnoticed that the website was not working anymore.’            |      | (76)  | (132) |       |

Table 3.10: Questionnaires - Clausal subjects with lexical case

Let us now take a closer look at the data from RMH. Apart from the data on *verða vart*, I incidentally obtained other examples of clausal subjects with lexical case from other datasets I gathered from RMH, exactly like in the case of nominative subjects. This occurred in five datasets, i.e. *spá* ‘predict’, *fresta* ‘postpone’ and *fagna* ‘rejoice, celebrate’ for dative case, and *krefjast* ‘demand’ and *sakna* ‘miss’ for genitive case, where I also checked whether the tokens with passivized clausal objects had their argument in the first position or after the finite verb.<sup>22</sup> These tokens generally feature clausal extraposition as well, which is simply a by-product of how the query for non-NomCIs

<sup>21</sup>Note that, in the first two examples, the *það* that is alternated with *þess* is an expletive and not a CN. I constructed the non-NomCI variant in this way because, when *þess* is dropped, the first position of the sentence cannot remain empty. Therefore, it has to be occupied by an expletive.

<sup>22</sup>Also in these datasets, I found some examples of the new passive construction. However, these examples are not counted in Table 3.11.

has been constructed (as the CN was found before the sequence of words captured by the query; see §3.2.3).

Regarding clausal subjects in the first position, from the five datasets I just mentioned, I could only take into account passivized genitive objects in order to observe more examples of clausal subjects with lexical case, as the dative CN *því* is homophonous with the adverb *því* ‘therefore’ and can become ambiguous when extraposition occurs (see §3.2.3). In the case of passivized genitive subjects, by contrast, *þess* is never ambiguous. As for clausal subjects with lexical case after the finite verb, I included passivized dative clausal objects as long as *því* was unambiguously a pronoun, as for instance in (14), where *því* comes after the conjunction *ef* ‘if’, which occupies the C-head, so that the only position that *því* can be in is Spec,T, which cannot be occupied by the adverb *því*:

- (14) Hann fullyrti einnig að öryggi Bandaríkjanna yrði stefnt í hættu ef því  
 he declared also that safety United States.the became directed to danger if that<sub>DAT</sub>  
 yrði frestað að afvopna Íraka  
 became postponed to disarm Iraqis  
 ‘He also declared that the safety of the United States would be endangered if disarming  
 the Iraqis were postponed.’

The results from RMH, shown in Table 3.11, corroborate the data from the questionnaires and suggest that *það* is highly frequent and could be even mandatory both before and after the finite verb as well as both with dative and genitive case. The great majority of percentages are above 90%:

| FIRST POSITION        |       |      |      |         |
|-----------------------|-------|------|------|---------|
| Verb                  | Valid | -það | +það | %það    |
| verða vart            | 21    | 0    | 21   | 100.00% |
| krefjast (pass.)      | 1422  | 5    | 1417 | 99.64%  |
| sakna (pass.)         | 16    | 0    | 16   | 100.00% |
| AFTER THE FINITE VERB |       |      |      |         |
| Verb                  | Valid | -það | +það | %það    |
| verða vart            | 23    | 0    | 23   | 100.00% |
| krefjast (pass.)      | 4577  | 115  | 4462 | 97.48%  |
| sakna (pass.)         | 13    | 1    | 12   | 92.30%  |
| spá (pass.)           | 1772  | 258  | 1514 | 85.44%  |
| fresta (pass.)        | 39    | 7    | 32   | 82.05%  |
| fagna (pass.)         | 498   | 4    | 494  | 99.19%  |

Table 3.11: RMH - Clausal subjects with lexical case

The data overall suggest that a clausal subject with lexical case must be generally nominalized, independently of whether it emerges before or after the finite verb.

The results we have discussed so far suggest that case plays a pivotal role in clausal nominalization, since a clear distinction emerges between clausal subjects with lexical case, where the

CN is generally mandatory, and nominative clausal subjects (including passivized clausal objects of verbs like *harma* or *gagnrýna*), where the pronoun is less common in the first position and mandatory after the finite verb. In other words, Spec,T position only hosts NomCIs independently of case, provided that the clausal subject is not extraposed. This can be easily linked to mandatory DP selection, since there is no distinction between structural and lexical case (we will discuss this issue further in §3.3.7 while observing clausal indirect objects). But this means that, given that NomCIs are not preferable in Spec,T when clausal extraposition occurs, this discrepancy between extraposed and non-extraposed clauses in Spec,T must be due to how they are derived. I will argue in Chapter 4 for the possibility that, in the case of clausal subjects that emerge after the finite verb, NomCIs compete with two possible different derivations which yield a non-NomCI.

### 3.3.2 Clausal direct objects

#### 3.3.2.1 Verbs involved

Let us now focus on clausal direct objects. The results presented here indicate that the distribution of NomCIs and non-NomCIs differs when the clausal complement is assigned nominative/accusative case on the one hand and dative/genitive on the other hand. The majority of the verbs I selected for this study are also the ones used by Thráinsson (1979) in support of his factivity hypothesis (in particular *harma*, *spá*, *fagna*, *fresta* and *sakna*; see Chapter 2). As for the other verbs, I generally tried to select verbs that had a fair number of tokens in RMH.<sup>23</sup>

- nominative: *líka* 'like', *leiðast* 'be sorry'
- accusative: *gagnrýna* 'criticize', *harma* 'regret, be sorry', *opinbera* 'disclose, reveal'
- dative: *spá* 'predict', *fresta* 'postpone', *fagna* 'rejoice, celebrate'
- genitive: *krefjast* 'demand', *sakna* 'miss'

Let us take a closer look at the results from questionnaires and naturalistic data. We will focus first on clausal direct objects with structural case and then the ones with lexical case.

#### 3.3.2.2 Nominative and accusative

As one can see in Table 3.12, I tested various examples of sentences in judgment tasks, the majority of which point to the fact that both variants are generally acceptable in accusative and nominative clausal objects.<sup>24</sup>

<sup>23</sup>As for *krefjast*, although the number of tokens in RMH was very high (see Table 3.15), I still wanted to test it as a non-factive genitive-assigning verb in contrast with the emotive factive *sakna*.

<sup>24</sup>I also additionally tested a variant of examples 2A-10 and 2B-16 where the verb of the subordinate clause is in subjunctive mood (see 4A-07 and 4B-19 in Appendix B), but it will not be shown on Table 3.12, as it seems that subjunctive mood is more degraded than indicative in these kinds of examples. It appears that, while *that*-clauses are generally acceptable in subjunctive mood, as shown in many of the examples presented in the questionnaires, embedded questions are more acceptable in indicative mood (consider e.g. 3A-02 and 3B-24, 4A-16 and 4B-10,

| NOMINATIVE |                                                                                           |      |       |       |      |
|------------|-------------------------------------------------------------------------------------------|------|-------|-------|------|
| No.        | Sentence                                                                                  | Type | -það  | +það  | p    |
| 3A-17      | <i>Þeim leiddist það að ég væri alltaf að stoppa og taka myndir</i>                       | tht  | 4.66  | 5.56  | .040 |
| 3B-09      | 'They were sorry about the fact that I was always stopping and taking pictures.'          |      | (35)  | (126) |      |
| 6A-02      | <i>Mér leiddist það að fá enga athygli frá Silju</i>                                      | inf  | 5.62  | 3.10  | .002 |
| 6B-24      | 'I was sorry about not receiving any attention from Silja.'                               |      | (21)  | (21)  |      |
| 2A-10      | <i>Mér leiddist það hvað markhópur bókarinnar var í raun lítil</i>                        | que  | 4.93  | 3.42  | .002 |
| 2B-16      | 'I was sorry about how the target group of the book was actually small.'                  |      | (107) | (26)  |      |
| 7A-20      | <i>Mér líkar það að bókin sé myndskreytt</i>                                              | tht  | 5.38  | 5.09  | .528 |
| 7B-06      | 'I like the fact that the book is decorated with pictures.'                               |      | (26)  | (23)  |      |
| ACCUSATIVE |                                                                                           |      |       |       |      |
| No.        | Sentence                                                                                  | Type | -það  | +það  | p    |
| 4A-04      | <i>Ýmsir hafa gagnrýnt það að flóttafólki sé vísað úr landi</i>                           | tht  | 6.00  | 5.89  | .904 |
| 4B-22      | 'Various people have criticized the fact that refugees are expelled from the country.'    |      | (21)  | (36)  |      |
| 3A-14      | <i>Ég gagnrýni það hvernig þið öfluðuð ykkur þessara upplýsinga</i>                       | que  | 5.04  | 4.66  | .509 |
| 3B-12      | 'I criticize how you got this information.'                                               |      | (126) | (35)  |      |
| 3A-05      | <i>Við hörmum það að stofnunin vilji hætta við verkefnið</i>                              | tht  | 5.74  | 5.58  | .741 |
| 3B-21      | 'We regret that the institute wants to cancel the project.'                               |      | (35)  | (126) |      |
| 1A-23      | <i>Ýmsir hafa gagnrýnt það að börnum sé leyft að alast upp í fangelsinu</i>               | tht  | 4.86  | 5.16  | .522 |
| 1B-03      | 'Various people have criticized the fact that children are allowed to grow up in prison.' |      | (69)  | (126) |      |
| 6A-17      | <i>Ég harma það hvað gerðist í síðustu viku</i>                                           | tht  | 3.57  | 3.62  | .920 |
| 6B-09      | 'I regret what happened last week.'                                                       |      | (21)  | (21)  |      |
| 10A-14     | <i>María opinberaði það að hún hefði eytt fjórum milljónum í 750 gjafakort</i>            | tht  | 5.38  | 4.77  | .496 |
| 10B-12     | 'María revealed that she had spent four million on 750 gift cards.'                       |      | (34)  | (30)  |      |

Table 3.12: Questionnaires - Nominative and accusative clausal objects

As for the data from RMH in Table 3.13, they consistently show a low frequency of *það*. This means that the pattern we have observed with nominative clausal subjects in the first position, where NomCIs are less frequent in linguistic performance although both variants are acceptable, is once again displayed here:<sup>25</sup>

5A-02 and 5A-24, see Appendix B). Mood in embedded clauses just appears to differ on the basis of which clause type surfaces, but it does not seem to affect which clausal variant is more frequent. In support of this idea, I also quickly checked in RMH whether there was any difference in the frequency of *það* in *that*-clauses and embedded questions selected by a verb like *vita* 'know', which is well known for allowing both indicative and subjunctive mood in its clausal complement when *vita* itself surfaces in the past tense (cf. Þráinsson et al. (2005:465)):

- (i) Ég vissi að hann **var/væri** kominn  
 I knew that he was<sub>IND/SUBJ</sub> come  
 'I knew he had come.'

The frequency of *það* does not appear to be affected by mood (*that*-clauses in indicative past: 16098 vs. 323 examples (1.96%); in subjunctive past: 28233 vs. 622 (2.15%); embedded questions in indicative past: 5764 vs. 7 (0.12%); in subjunctive past: 2601 vs. 5 (0.19%)). With these facts at hand, we can assume that mood is outside the scope of this dissertation as it does not appear to affect clausal nominalization. Therefore, I will not discuss it any further.

<sup>25</sup>The examples presented here on accusative clausal objects are only in active voice. The passive ones have already been shown in our discussion on clausal subjects. The data on active and passive voice are shown together in the same table in Appendix A.

| NOMINATIVE AND ACCUSATIVE |               |                |                 |               |                 |
|---------------------------|---------------|----------------|-----------------|---------------|-----------------|
|                           | <i>líka</i>   | <i>leiðast</i> | <i>gagnrýna</i> | <i>harma</i>  | <i>opinbera</i> |
| <b>Valid</b>              | 246           | 577            | 4241            | 2492          | 462             |
| <b>-það</b>               | 212           | 553            | 3739            | 1897          | 364             |
| <b>+það</b>               | 34            | 24             | 502             | 595           | 98              |
| <b>%það</b>               | <b>13.82%</b> | <b>4.16%</b>   | <b>11.84%</b>   | <b>23.88%</b> | <b>21.21%</b>   |

Table 3.13: RMH - Nominative and accusative clausal objects

Let us focus for a moment on the verb *leiðast* in particular, before we move to the results on dative and genitive clausal objects. I tested this verb in the questionnaires more than once due to the fact that, in RMH, *það* presents a very low frequency with this particular verb (4.16%). I wondered whether such a low percentage could be a hint of ungrammaticality. The results from questionnaires in Table 3.12, however, indicate that *leiðast* allows nominalization, as shown in 3A-17 and 3B-09, where the pronoun appears to be more acceptable (the statistical difference appears to be significant, as  $p = .04$ ). But it is also true that, in the other two examples with *leiðast*, the clausal argument is considered more felicitous without a pronoun (and  $p = .002$  in both instances). The reader might wonder whether the difference among these examples could perhaps depend on the clause type, but the data from RMH show that the frequency of *það* before *that*-clauses and infinitives in the *leiðast* dataset is proportionally similar (7.41% vs. 4.06% respectively, see Appendix A), so that the difference between the pairs of examples with *leiðast* in Table 3.12 can hardly be explained by the clause type involved. My view, considering the data from the questionnaires in particular, is that *það* cannot be considered ungrammatical after *leiðast*, otherwise 3A-17 and 3B-09 should also show that a NomCl is degraded, which is not the case.

### 3.3.2.3 Dative and genitive

On the basis of our previous observations on clausal subjects, we can expect dative and genitive case in clausal direct objects to be correlated to a high frequency of *það* in the data from RMH as well as a significant difference between variants in favor of NomCIs in the results from the questionnaires. However, we should also expect there to be a contrast between certain verbs assigning dative. Recall our discussion on Thráinsson (1979), who claimed that there is a difference between verbs like *spá* ‘predict’ (also *fresta* ‘postpone’) and *fagna* ‘rejoice’, where only the former can be followed by an optional CN:

- (15) a. Jón spáði (**því**) að María væri farin  
 Jón predicted that<sub>DAT</sub> that María was gone  
 ‘Jón predicted that María would have left.’ (Thráinsson 1979:216)
- b. Ég fagna \*(**því**) að þú skulir vera kominn  
 I rejoice that<sub>DAT</sub> that you shall be come  
 ‘I am happy that you came.’ (Thráinsson 1979:230)

In order to verify the validity of this contrast, I tested both verbs in the judgment tasks and naturalistic data. As for the data from the questionnaires presented in Table 3.14, we can see that the non-NomCl variant is more degraded across the board. Apart from two pairs of examples in which no statistical significant difference has been found (see 3A-02/3B-24 and 5A-14/5B-12), probably due to a bias related to the acceptability of the clause type following the verb, we can see that NomClS are generally much better than non-NomClS in both dative and genitive case, as well as with verbs like *spá* and *fresta*, which were originally associated with optional nominalization based on Thráinsson's (1979) account:

| DATIVE   |                                                                                   |      |       |       |       |
|----------|-----------------------------------------------------------------------------------|------|-------|-------|-------|
| No.      | Sentence                                                                          | Type | -það  | +það  | p     |
| 2A-01    | <i>Sindri fagnar því að bændur hafi fengið lækkun á raforkuverði</i>              | tht  | 2.73  | 5.26  | <.001 |
| 2B-25    | 'Sindri is happy that farmers got a reduction of the price of electricity.'       |      | (26)  | (107) |       |
| 2A-13    | <i>Einar spáir því að liðið komist í úrslitaleikinn</i>                           | tht  | 4.15  | 6.18  | <.001 |
| 2B-13    | 'Einar predicts that the team will reach the final.'                              |      | (26)  | (107) |       |
| 3A-02    | <i>Ég þori ekki að spá því hvað verður um bækur í framtíðinni</i>                 | que  | 3.96  | 4.69  | .114  |
| 3B-24    | 'I do not dare predict what is going to happen to books in the future.'           |      | (126) | (35)  |       |
| 10A-17   | <i>Breska þingið frestaði því að fara í sumarfrí</i>                              | inf  | 3.63  | 6.00  | <.001 |
| 10B-09   | 'The British parliament postponed going on vacation.'                             |      | (30)  | (34)  |       |
| GENITIVE |                                                                                   |      |       |       |       |
| No.      | Sentence                                                                          | Type | -það  | +það  | p     |
| 1A-14    | <i>Ragnheiður krafðist þess að Árni bæði starfsmennina afsökunar</i>              | tht  | 3.76  | 6.30  | <.001 |
| 1B-12    | 'Ragnheiður demanded Árni to apologize to the staff members.'                     |      | (126) | (69)  |       |
| 6A-11    | <i>Dómarinn krafðist þess að fá að lesa skýrslurnar</i>                           | inf  | 3.48  | 6.62  | <.001 |
| 6B-15    | 'The judge demanded access to the reports.'                                       |      | (21)  | (21)  |       |
| 2A-07    | <i>Þóra var farin að sakna þess að búa á Íslandi</i>                              | inf  | 2.81  | 6.24  | <.001 |
| 2B-19    | 'Þóra had started to miss living in Iceland.'                                     |      | (26)  | (107) |       |
| 5A-14    | <i>Ég sakna þess að amma muni aldrei frammar bjóða mér aðra kökusneið</i>         | tht  | 2.78  | 3.50  | .284  |
| 5B-12    | 'I miss the fact that grandma will never offer me another slice of cake anymore.' |      | (23)  | (20)  |       |

Table 3.14: Questionnaires - Dative and genitive clausal objects

The data from RMH, on the other hand, shows that the frequency of *það* exceeds 90% with all the verbs involved, *spá* and *fresta* included, as we can see in Table 3.15. These data suggest that NomClS tend to be highly preferable, possibly mandatory when dative or genitive is assigned to the clausal complement. This contrasts with Thráinsson's idea that *það* is simply optional with verbs like *spá* and *fresta*, as the results from both questionnaires and corpora suggest that NomClS are preferable:<sup>26</sup>

| DATIVE AND GENITIVE |               |               |               |                 |               |
|---------------------|---------------|---------------|---------------|-----------------|---------------|
|                     | <i>spá</i>    | <i>fagna</i>  | <i>fresta</i> | <i>krefjast</i> | <i>sakna</i>  |
| <b>Valid</b>        | 11820         | 8926          | 1602          | 32593           | 5228          |
| <b>-það</b>         | 1045          | 157           | 154           | 278             | 124           |
| <b>+það</b>         | 11820         | 8769          | 1448          | 32315           | 5104          |
| <b>%það</b>         | <b>91.88%</b> | <b>98.24%</b> | <b>90.39%</b> | <b>99.15%</b>   | <b>97.63%</b> |

Table 3.15: RMH - Dative and genitive clausal objects

<sup>26</sup>The data from RMH presented on dative and genitive clausal direct objects are only the ones in active voice. The passive tokens have already been presented in our discussion on clausal subjects with lexical case.

It is true, however, that the *spá* dataset in RMH contains more than a thousand examples with a non-NomCl, unlike the other datasets (see Table 3.15). It is likely that *það* is highly preferable with *spá* rather than mandatory, which is also corroborated by the fact that the frequency of *það* with passivized clausal objects of *spá* was lower than 90% (see §3.3.1.2; the same happened with the *fresta* dataset).

Nevertheless, as a general trend, we can see that there is a distinction in clausal nominalization between structural case and lexical case (see §4.4 for a more detailed discussion on this distinction). In instances of structural case, NomCIs are less frequent in language performance although nominalization is still admissible. Lexical case tends, by contrast, to require or strongly prefer *það* to nominalize the clause.

### 3.3.3 Post-copular clauses

Let us now discuss post-copular clauses. As previously mentioned in §2.2.1, a CN in postcopular clauses agrees in gender and number with the subject:

- (16) Afleiðingin er [sú að við skuldum meira núna]  
 consequence.the<sub>NOM-F-SG</sub> is that<sub>NOM-F-SG</sub> that we owe more now  
 ‘The consequence is that we owe more money now.’ (Garofalo 2020:9)

Given that Icelandic features three genders and two numbers, I decided to test postcopular clauses and nominalization by selecting six nouns (with a fair number of valid examples in RMH) that play the role of the subject, as shown in Table 3.16, each with different gender and number features. I found it unnecessary to test all these nouns in the questionnaires, as I needed to simply verify whether masculine, feminine or plural (which we only find in post-copular clauses) have any impact on the acceptability of NomCIs and non-NomCIs. Hence, I tested one noun where gender and number features are default (*vandamál*), one masculine noun (*kostir*), one feminine (*afleiðing*) and one plural with a default gender (*rök*):<sup>27</sup>

| Gender | Singular                       | Plural                       |
|--------|--------------------------------|------------------------------|
| M      | <i>galli</i> ‘drawback’        | <i>kostir</i> ‘advantages’   |
| F      | <i>afleiðing</i> ‘consequence’ | <i>niðurstöður</i> ‘results’ |
| N      | <i>vandamál</i> ‘problem’      | <i>rök</i> ‘reasons’         |

Table 3.16: Classes of subjects - Postcopular clauses

The results from the questionnaires align with the data on structural case positions, as both variants are generally acceptable. This is shown in Table 3.17:

<sup>27</sup>The data from RMH on these constructions generally feature these nouns with a suffixed article. Therefore, I focused specifically on instances where these nouns are definite.

| No.    | Sentence                                                                                         | Type | -það  | +það  | p    |
|--------|--------------------------------------------------------------------------------------------------|------|-------|-------|------|
| 1A-02  | <i>Afleiðingin er sú að Bandaríkin eru í alvarlegri skuldakreppu</i>                             | tht  | 5.38  | 6.07  | .177 |
| 1B-24  | 'The consequence is that the USA is in a serious debt crisis.'                                   |      | (126) | (69)  |      |
| 2A-04  | <i>Kostirnir eru þeir að kerfið er ódýrt og einfalt í notkun</i>                                 | tht  | 6.65  | 6.38  | .435 |
| 2B-22  | 'The advantages are that the system is cheap and easy to use.'                                   |      | (107) | (26)  |      |
| 2A-19  | <i>Vandamálið er það að fólk vill ekki láta bólusetja sig</i>                                    | tht  | 5.88  | 5.44  | .218 |
| 2B-07  | 'The problem is that people do not want to be vaccinated.'                                       |      | (26)  | (107) |      |
| 5A-08  | <i>Vandamálið hjá þér er það að vera of kurteis við fólk</i>                                     | inf  | 4.48  | 3.90  | .242 |
| 5B-18  | 'Your problem is being too polite to people.'                                                    |      | (23)  | (20)  |      |
| 10A-11 | <i>Rökkin fyrir aukagjaldi eru þau að starfsmenn í sorphirðu séu undir svo miklu álagi</i>       | tht  | 4.83  | 5.62  | .238 |
| 10B-15 | 'The reasons for an extra charge are that workers for waste collection are under much pressure.' |      | (30)  | (34)  |      |

Table 3.17: Questionnaires - Postcopular clauses

Now, in relation to the data from RMH, based on our observations on nominative clausal subjects and nominative/accusative clausal direct objects, we should expect the non-NomCI variant to be the most frequent, due to the fact that postcopular clauses normally surface in nominative case (except when we deal with ECM).<sup>28</sup> The results from RMH differ to some extent from this prediction. In Table 3.18, we can see a clear contrast between the CN in neuter singular and the other instances of CNs. The frequency of *það* after nouns like *vandamál* 'problem' is very low (almost 6%), while it even reaches 70% and beyond with other nouns that are masculine, feminine and/or plural:

| GENERAL      |                      |                          |                         |                       |                            |                    |
|--------------|----------------------|--------------------------|-------------------------|-----------------------|----------------------------|--------------------|
|              | <i>galli (m.sg.)</i> | <i>afleiðing (f.sg.)</i> | <i>vandamál (n.sg.)</i> | <i>kostir (m.pl.)</i> | <i>niðurstöður (f.pl.)</i> | <i>rök (n.pl.)</i> |
| <b>Valid</b> | 325                  | 659                      | 721                     | 39                    | 54                         | 142                |
| <b>-það</b>  | 97                   | 123                      | 678                     | 19                    | 20                         | 32                 |
| <b>+það</b>  | 228                  | 536                      | 43                      | 20                    | 34                         | 110                |
| <b>%það</b>  | <b>70.15%</b>        | <b>81.34%</b>            | <b>5.96%</b>            | <b>51.28%</b>         | <b>62.96%</b>              | <b>77.46%</b>      |

Table 3.18: RMH - Postcopular clauses

In relation to the *vandamál* dataset, although the frequency of *það* is very low, the results from the questionnaires show that both variants are acceptable, which suggests that NomCIs are most probably less frequent in language performance than non-NomCIs. However, overall, naturalistic data suggest that whenever the gender or the number of a noun is non-default, NomCIs are preferable. However, when the subject is neuter singular, the post-copular clause still follows the same pattern we have observed in nominative subjects in the first position and nominative and accusative clausal objects.

As for the datasets *kostir* and *niðurstöður*, one would point out that the frequency of *það* (see Table 3.18) suggests that both variants are equipollent. This is definitely correct if we take these datasets on their own. But we cannot ignore the fact that there is a general trend clearly expressed

<sup>28</sup>I decided to leave out ECM clauses from my investigation because the data on accusative clausal postcopular clauses in RMH are very few and the great majority of them are in feminine singular, which means it is impossible to compare their frequency of *það* with neuter singular postcopular clauses or other instances of postcopular clauses in a different gender or number. However, just by quickly reviewing the existing data in RMH online, feminine singular NomCIs constitute about 66% of the relevant tokens (29 vs. 53), which is in line with the data on nominative postcopular clauses presented in Table 3.18.

across the board in postcopular clauses where the percentage of *það* is high (or at least much higher than in e.g. nominative and accusative clausal objects). Therefore, as a more general trend, it is possible that *það* is more frequent in nominal predicate position when the subject is not neuter singular (the average percentage, if we exclude the *vandamál* dataset, is 68.38%, which is very close to the threshold beyond which *það* is most likely preferable). Furthermore, these two datasets are also the ones where the number of examples gathered is the lowest (only 39 and 54 respectively). In other words, a lower frequency of *það* might be caused by a narrow sample in both datasets.<sup>29</sup>

Generally speaking, the results we have observed here suggest that gender and number also affect the distribution of NomCIs, as masculine and feminine CNs as well as the ones in the plural are more frequent than the non-NomCI variant. But the results also tell us something far more important. The postcopular clauses we have discussed are also in nominative case, which means that *það* is less likely to appear when all the gender, number and case features are default (i.e. neuter singular nominative). In other words, clausal nominalization tends to be triggered by non-default features belonging to these three grammatical categories, as summarized below:

(17) *Non-default feature effect on clausal nominalization*

A NomCI is at least preferable in a syntactic position if (fulfilling one of the following conditions is enough):

- a) it is assigned lexical case
- b) it is assigned non-default gender
- c) it is assigned non-default number

Note that there is a difference, however, between lexical case and non-default gender and number features, as the former can trigger obligatory nominalization while the latter cannot, as a NomCI is simply more frequent. This could depend on the fact that the distribution of postcopular clauses is limited to structural case positions (either nominative or accusative as in ECM constructions), which might cause both clausal variants to be acceptable or, alternatively, the non-NomCI to be specifically acceptable even if gender and number features are non-default.

### 3.3.4 Prepositional phrases

#### 3.3.4.1 Main properties

Now let us focus on the distribution of clausal nominalization in prepositional phrases. Before taking a look at the data from the questionnaires and corpora, I believe it is necessary to discuss

<sup>29</sup>A quick search among the data from RMH online shows that the frequency of *það* can be even higher than in *kostir* or *niðurstöður* with other non-neuter plural nouns. For instance, with the plural of *afleiðing* as subject, the percentage of *það* in postcopular clauses is 74.69% (82 vs. 242 examples). I did not take this dataset into account due to the fact that I preferred analyzing a totally different noun to represent a feminine plural subject, i.e. *niðurstöður*. However, these facts confirm the idea that a narrow sample could be the cause of a lower frequency of *það* in the *kostir* and *niðurstöður* datasets.

some important notions on prepositions, in particular some major distinctions between prepositions based on different factors. These factors are important as they also explain which types of prepositions have been tested.

A first factor to consider is that prepositions in Icelandic can be selected by other lexical items like verbs, adjectives or nouns, or can be found in an adjunct position. This difference is shown in the following examples:

- (18) a. *Petta bendir til þess að hann sé búinn að strauja tölvuna*  
 this points to that<sub>GEN</sub> that he is finished to format computer.the  
 'This indicates that he formatted the computer.'
- b. *Ég þóttist vera veikur til þess að skrópa*  
 I pretended be sick to that<sub>GEN</sub> to skip school  
 'I pretended to be sick to skip school.'

In the first example, *til* 'to' is selected by the verb *benda* 'point' in order to convey the meaning of indicating. In the second example, the preposition *til* introduces a purpose clause, which means that it surfaces in an adjunct position.

Secondly, Icelandic prepositions are also distinguished based on case assignment. It is commonly known that prepositions in Icelandic can either assign one case only (accusative, dative or genitive), as happens with e.g. *um* + ACC 'about', *frá* + DAT 'from', *vegna* + GEN 'because of' or *hjá* + DAT 'by', or two cases (accusative and dative), as generally occurs with prepositions like *í* 'in', *á* 'on', *yfir* 'over' or *fyrir* 'for, before'. At times, prepositions can also be combined with adverbs like *upp á* 'upon', *ofan í* 'into' and the case assigned is still the one that the preposition alone can generally assign, e.g. either accusative or dative in the case of prepositions like *á* or *í* (see also Þráinsson et al. (2005) for more information on Icelandic prepositions).

Thirdly, a further distinction that is made among prepositions is the one between functional and lexical prepositions (see Mardale (2011) for more discussion on this topic). In the linguistic literature, lexical prepositions display a full meaning, usually locative or temporal (see Mardale (2011)), while functional prepositions are abstract. In certain accounts like Rauh et al. (1994), it is also pointed out that lexical and functional prepositions also differ on the basis of their syntactic structure. While lexical prepositions can be followed or preceded by modifiers or adjuncts (and are also able to assign  $\theta$ -roles), functional prepositions are unable to do so. Consider, as an example of distinction between these two categories, prepositions like *meðfram* 'along', which displays a locative meaning and can also be preceded by a modifier like *alveg* 'entirely', in contrast with the preposition *á* 'on', which can be selected by many Icelandic nouns to introduce a complement alternative to a nominal genitive complement (the case assigned by *á* is dative). This preposition, in this case, cannot be preceded by a modifier:

- (19) a. *Ruslið lá (alveg) meðfram veginum*  
 garbage.the lay entirely along road.the  
 'The garbage lay (entirely) along the road.'

- b. Lýsingin (\*alveg) á myndinni var ekki nógu skýr  
 description.the entirely on picture.the was not enough clear  
 'The description of the picture was not clear enough.'

Lexical prepositions are also known to be often able to be replaced by other prepositions, while functional prepositions resist their replacement, as shown in the following contrast (see also Mardale (2011)):

- (20) a. Gísli gat ekki tjáð sig vegna/sökum reglna / út af reglum okkar  
 Gísli could not express himself because rules<sub>GEN</sub> because rules<sub>DAT</sub> our  
 um persónuvernd  
 about privacy  
 'Gísli could not speak because of our privacy regulations.'
- b. Það er misjafnt eftir/\*fyrir/\*á deildum hvenær endurtökuprófin eiga sér stað  
 it is dependent after/for/on faculties when resit exams.the take place  
 'It depends on the faculties when resit exams will take place.'

In the linguistic literature, a preposition is also considered functional when an item subcategorizes for it, even when it is preceded by adverbs, like the preposition *í* in *ofan í* (see cross-linguistically e.g. Gorrie (2008); Waters (2008)). Assuming this analysis of functional prepositions, if we want to map clausal nominalization in prepositional phrases, we can also assume as a rule of thumb that prepositions selected by lexical items like verbs or adjectives are functional, whereas lexical prepositions can often be found in adjuncts (and we can expect many of them to also convey a clear meaning). In addition to this distinction between lexical and functional prepositions, we also have to consider in our analysis those that assign only one case, the ones that assign two possible cases and the ones which are also preceded by adverbs.

### 3.3.4.2 Datasets

Since case appears to be pivotal in the distribution of NomCIs, I decided to study prepositions on the basis of the case(s) they assign. Here is the list of functional prepositions assigning only one case that I studied in questionnaires and corpora:

- only accusative: *um* 'about', in the predicates *vera meðvitaður um* 'be aware of' and *fjalla um* 'be about'
- only dative: *hjá* 'by', in the verb *komast hjá* 'avoid', and *af* 'from', in the verb *frétta af* 'get to know, learn'
- only genitive: *til* 'to', in the verbs *ætlast til* 'expect', *langa til* 'want' and *mælast til* 'require to do sth.'

The prepositions *um* and *til* have been tested more than once due to the fact that they are the only ones that are selected by verbs or adjectives when accusative and genitive case respectively

are assigned.<sup>30</sup>

The following list, on the other hand, contains the functional prepositions I tested that assign two possible cases, either accusative or dative. I made sure I tested both cases in corpora and judgment tasks for each preposition:

- *í* 'in': *styttast í* + ACC 'approach', *vera tilbúinn í* + ACC 'be ready to', *pæla í* + DAT 'reflect upon'
- *yfir* 'over': *komast yfir* + ACC 'get over', *vera svekkur yfir* + DAT 'be annoyed about', *gleðjast yfir* + DAT 'rejoice about'

With *í* and *yfir*, there are also two types of predicates under observation. On the one hand, there are stand-alone verbs, while on the other hand there are predicates composed of the copula and an adjective. Recall that, in our discussion of Thráinsson (1979) in §2.5.2, we saw that verbs like *gleðjast yfir* and *vera glaður yfir* 'be happy about sth.' feature an optional pronoun, despite the fact that they are emotive factives, which, according to Thráinsson, tend to be followed by a NomCl. Thus, the main reason why I wanted to review these two types of verbs specifically is to verify any possible difference between them in relation to the distribution of clausal nominalization.

Since functional prepositions can also be preceded by an adverb, I verified whether the presence of a modifier can impact the distribution of NomCls (genitive was not tested here as I am not aware of any verb which selects an adverb+genitive preposition with an embedded clause):

- only accusative: *komast upp um* 'come to light'
- only dative: *ganga út frá* 'assume'
- dative or accusative: *vanta upp á* + ACC 'lack', *finna upp á* + DAT 'find a way to do sthg.'

Lastly, I also considered three prepositions that are exclusively lexical, one for each case and with a fair amount of valid examples: *í kringum* 'around' for accusative,<sup>31</sup> *gagnvart* 'towards' for dative and *utan* 'apart from' for genitive.

### 3.3.4.3 Functional prepositions assigning one case

Let us focus first on single prepositions selected by verbs or adjectives that assign only one case. In relation to accusative case, as I have already mentioned, we will take a look at the preposition *um* 'about' in the predicates *vera meðvitaður um* 'be aware of' and *fjalla um* 'be about'. The data from the questionnaires in Table 3.19 indicate that both variants are acceptable:<sup>32</sup>

<sup>30</sup>I tested *til* three times due to the fact that the first two datasets I originally tested (*langa til* and *ætlast til*) presented two different patterns. I decided, therefore, to check a third verb in order to find out what the most common pattern is. We will see this in more detail in the next subsection.

<sup>31</sup>It is true, however, that *í kringum* appears to be a more complex example of a preposition due to the presence of multiple components at the surface. Our purpose here, however, is not to identify the structure of specific prepositions, but rather just the distribution of NomCls and non-NomCls.

<sup>32</sup>Tokens 1A-08 and 1B-18 (see Appendix B) are an alternative to 4A-01 and 4B-25 presented in the table. The problem is that the former pair of examples obtained negative judgments when I started testing them, so I decided

| No.    | Sentence                                                                             | Type | -það | +það | p    |
|--------|--------------------------------------------------------------------------------------|------|------|------|------|
| 4A-16  | <i>Hann er mjög meðvitaður um það hvaða áhrif notkun snjallsíma hefur á börn</i>     | que  | 6.19 | 5.58 | .483 |
| 4B-10  | 'He is very aware of the influence the use of smartphones has on children.'          |      | (21) | (36) |      |
| 4A-01  | <i>Við erum meðvituð um það að ákveðnir þjóðfélagshópar búi við lág laun</i>         | tht  | 4.67 | 5.24 | .589 |
| 4B-25  | 'We are aware that certain society groups live on a low salary.'                     |      | (36) | (21) |      |
| 5A-17  | <i>Margir eru meðvitaðir um það að borða hollt og reglulega yfir daginn</i>          | inf  | 5.30 | 5.00 | .968 |
| 5B-09  | 'Many are aware of eating healthy and regularly during the day.'                     |      | (20) | (23) |      |
| 10A-05 | <i>Lagið fjallar um það að njóta augnabliksins</i>                                   | inf  | 6.40 | 6.12 | .098 |
| 10B-21 | 'The song is about enjoying the moment.'                                             |      | (30) | (34) |      |
| 10A-20 | <i>Í ræðunni er fjallað um það að ekki megi dæma fólk eftir útliti</i>               | tht  | 6.24 | 6.30 | .928 |
| 10B-06 | 'In the speech, it is said that one may not judge people based on their appearance.' |      | (34) | (30) |      |

Table 3.19: Questionnaires - Accusative prepositional complements

On the basis of the data from structural case positions, we expect *það* to be less frequent according to the data from RMH. This prediction is confirmed in Table 3.20:

|              | <i>vera meðvitaður um</i> | <i>fjalla um</i> |
|--------------|---------------------------|------------------|
| <b>Valid</b> | 5401                      | 3145             |
| <b>-það</b>  | 4617                      | 2474             |
| <b>+það</b>  | 784                       | 671              |
| <b>%það</b>  | <b>14.52%</b>             | <b>21.34%</b>    |

Table 3.20: RMH - Accusative prepositional complements

These results align with the ones from accusative and nominative clausal objects as well as nominative subjects in the first position, and suggest, in particular, that accusative case in prepositional complements displays the same properties of accusative case in canonical argument positions when it comes to clausal nominalization: the pronoun is less common, although both variants are considered acceptable.

Now we can move to dative case and observe the prepositions *af* 'of' in the verb *frétta af* 'get to know' and *há* 'by' in *komast há* 'avoid'. Table 3.21 illustrates the relevant results from judgment tasks:

| No.    | Sentence                                                           | Type | -það | +það  | p     |
|--------|--------------------------------------------------------------------|------|------|-------|-------|
| 3A-23  | <i>Enginn kemst há því að kaupa sér mat</i>                        | inf  | 3.06 | 6.57  | <.001 |
| 3B-03  | 'No one can avoid buying food for themselves.'                     |      | (35) | (126) |       |
| 7A-08  | <i>Við getum ekki komist há því að verkefnið fari í klessu</i>     | tht  | 2.65 | 5.87  | <.001 |
| 7B-18  | 'We cannot avoid the fact that the project is going to be ruined.' |      | (26) | (23)  |       |
| 10A-02 | <i>Hann var búinn að frétta af því að María borðaði ekki fisk</i>  | tht  | 3.29 | 5.67  | <.001 |
| 10B-24 | 'He got to know that María did not eat fish.'                      |      | (34) | (30)  |       |

Table 3.21: Questionnaires - Dative prepositional complements

to replace them as soon as possible with the latter pair of tokens to test *that*-clauses once more. Later on, however, those examples obtained more positive judgments, especially 1B-18. Since we do not need two instances of the same type here, I decided to only present the results on 4A-01 and 4B-25. The same happened with tokens 2A-25 and 2B-01, which were an alternative to 4A-16 and 4B-10 and which obtained negative judgments for some time. As one can see in Appendix B, the *p*-value of the pairs that are not shown here is  $p > .05$ , so that there is no significant statistical difference between those variants either.

The data clearly indicates that *það* is mandatory, as the non-NomCl variant is highly degraded. The results from RMH corroborate pronominal obligatoriness, as shown in Table 3.22:

|              | <i>komast hjá</i> | <i>frétta af</i> |
|--------------|-------------------|------------------|
| <b>Valid</b> | 4941              | 1092             |
| <b>-það</b>  | 262               | 6                |
| <b>+það</b>  | 4679              | 1086             |
| <b>%það</b>  | <b>94.70%</b>     | <b>99.45%</b>    |

Table 3.22: RMH - Dative prepositional complements

In other words, dative case here displays properties that are similar to the ones of dative case in canonical argument positions, as the pronoun is required.

Let us now look at genitive case and, in particular, at the preposition *til*. Originally, I only observed the verbs *langa til* ‘want’<sup>33</sup> and *ætlast til* ‘expect’, but the data from RMH presented in Table 3.24 (see below) show that there was no clear pattern visible, as *það* is less common with *langa til* whereas it is more frequent with *ætlast til*. I tried, therefore, to observe a third dataset both with questionnaires and naturalistic data, namely *mælast til* ‘require to do sth.’. Let us observe the results from the questionnaires first:

| No.   | Sentence                                                                               | Type | -það  | +það  | p     |
|-------|----------------------------------------------------------------------------------------|------|-------|-------|-------|
| 1A-20 | <i>Mig langar til þess að þakka þeim sem sendu umsagnir</i>                            | inf  | 5.90  | 5.80  | .857  |
| 1B-06 | ‘I want to thank those who sent comments.’                                             |      | (126) | (69)  |       |
| 5A-05 | <i>Mig langar til þess að allir öðlist hamingju</i>                                    | tht  | 4.15  | 5.09  | .144  |
| 5B-11 | ‘I want everyone to achieve happiness.’                                                |      | (20)  | (23)  |       |
| 7A-02 | <i>Ég ætlast ekki til þess að Bjarna sé sagt upp</i>                                   | tht  | 5.15  | 6.04  | .200  |
| 7B-24 | ‘I don’t expect that Bjarni is going to be fired.’                                     |      | (26)  | (23)  |       |
| 9A-11 | <i>Það er ekki hægt að ætlast til þess að fá allar sínar óskir uppfylltar</i>          | inf  | 5.59  | 5.92  | .271  |
| 9B-15 | ‘It is not possible to expect all wishes to be fulfilled.’                             |      | (101) | (65)  |       |
| 9A-02 | <i>Sóttvarnalæknir mælist til þess að áhættuhópar njóti forgangs við bólusetningar</i> | tht  | 5.35  | 6.43  | .003  |
| 9B-24 | ‘The epidemiologist asks that groups at risk are prioritized for vaccinations.’        |      | (65)  | (101) |       |
| 9A-23 | <i>Það er mælst til þess að mæta á flugvöllinn þremur tímum fyrir brottför</i>         | inf  | 3.77  | 5.69  | <.001 |
| 9B-03 | ‘It is required to arrive at the airport three hours before the departure.’            |      | (101) | (65)  |       |

Table 3.23: Questionnaires - Genitive prepositional complements

Overall, the data show that, independently of the verb involved, both variants are acceptable if they are introduced by *til*. These results differ from canonical argument positions, where genitive case is correlated to mandatory *það*. I believe this difference is due to the fact that *til* is a highly

<sup>33</sup>Although the verb *langa* might be considered problematic due to the fact that its subject displays case variation (see Jónsson (2005b); Nowenstein (2014); Barðdal (2011); Jónsson and Eythórssón (2011) among others), I still chose to test *langa* due to the fact that it is a very common verb in Icelandic. What is more, I believe case variation here does not have any impact on the distribution of NomClIs and non-NomClIs. In support of this, I observed more closely the data from RMH and counted all the valid examples based on the case of the subject. With the exclusion of ambiguous tokens (3767), the data with an accusative subject show that *það* is less frequent (9809 vs. 1931; 16.45%), and the ones with a dative subject confirm a similar distribution (164 vs. 51; 23.72%). The same can also be observed with nominative subjects (36 vs. 5; 12.20%). The remaining examples are marginal and contain a genitive subject (7 tokens).

frequent preposition and can often precede adjuncts, in particular purpose clauses, where *þess* is generally optional:

- (21) Ég kom snemma í morgun til (**þess**) að læra í rólegheitum  
 I came early this morning to that<sub>GEN</sub> to study in tranquillity  
 'I came early this morning in order to study in tranquillity.'

Among the results from the questionnaires, however, we find an exception to the general trend. In fact, in infinitives with *mælast til*, the non-NomCI variant appears to be more degraded, which means that, at least in the case of *mælast til*, there is a difference in the distribution of clausal nominalization between *that*-clauses and infinitives. This is also supported by the fact that nominalized *that*-clauses are more acceptable than the non-NomCI ones (see 9A-02/9B-24).

Let us now observe the data from RMH, where, as already mentioned, *langa til* displays a different behavior compared to the other verbs:

|              | <i>langa til</i> | <i>ætlast til</i> | <i>mælast til</i> |
|--------------|------------------|-------------------|-------------------|
| <b>Valid</b> | 15770            | 7002              | 2199              |
| <b>-það</b>  | 13255            | 2643              | 270               |
| <b>+það</b>  | 2515             | 4359              | 1929              |
| <b>%það</b>  | <b>15.95%</b>    | <b>62.25%</b>     | <b>87.72%</b>     |

Table 3.24: RMH - Genitive prepositional complements

The question is how we can interpret this difference. First of all, I believe that the dominant trend in relation to genitive prepositional objects is to prefer *það*. By quickly looking for additional data in RMH, there are other verbs like *benda til* 'indicate' after which NomCIs are more frequent (20556 vs. 51297; 71.39%), as well as *verða til* 'lead to' (8631 vs. 59409; 87.31%) or *sjá til* 'see to' (3237 vs. 21276; 86.79%). Therefore, if *langa til* is an exception, my view is that it differs because it is the only verb in our study of genitive assigning prepositions that can optionally take a prepositional phrase as a complement in order to introduce an embedded clause. In fact, by default, *langa* can select a clause as complement without a preposition:

- (22) a. Mig langar (**\*það**) að fara í bíó  
 me<sub>ACC</sub> wants that to go to cinema  
 'I want to go to the cinema.'
- b. Okkur langar (**\*það**) að allir í heiminum verði bestu vinir  
 us wants that that everyone in world.the become best friends  
 'We want everyone in the world to become best friends.' (from *Morgunblaðið*)

In this construction, as we can observe, nominalization is disallowed, due to the fact that the clausal complement does not surface in a DP position:<sup>34</sup>

<sup>34</sup>Note that the accusative subject that precedes *langa* is not extracted from the embedded clause. Therefore, the ungrammaticality of *það* cannot be explained by its incompatibility with extraction. The impossibility of subcategorization for a DP, by contrast, can definitely account for the ungrammaticality of a CN in the examples with *langa*.

- (23) a. \*Mig langar **vatn**  
 me<sub>ACC</sub> wants water  
 Intended: 'I want water.'
- b. \*Okkur langar **breytingu**  
 us wants change  
 Intended: 'We want a change.'

*Ætlast* and *mælast*, by contrast, are unable to subcategorize for a bare complement clause, as illustrated in the following examples:

- (24) a. \*Læknar mælast að fólk drekki tvo lítra af vatni á dag  
 doctors ask that people drink two liters of water per day  
 'Doctors recommend people to drink two liters of water per day.'
- b. \*Ég ætlast ekki að hann fái tíu í öllum áföngum  
 I expect not that he gets ten in all subjects  
 'I don't expect him to get a ten in all subjects.'

It is possible, considering these facts, that NomCIs are less frequent with *langa til* due to the fact that its default complement (an embedded clause) is unable to be nominalized.<sup>35</sup> Verbs like *ætlast* and *mælast*, which must subcategorize for a prepositional complement, are much more likely to be correlated to clausal nominalization.

### 3.3.4.4 Functional prepositions assigning two cases

Let us now observe the group of prepositions selected by verbs or adjectives that can assign both accusative and dative case. The prepositions I studied, as already mentioned, are *í* 'in' and *yfir* 'over' (I have also investigated the preposition *á* 'on' in §3.3.4.5, where the preposition is also preceded by an adverb). Let us take a look again at the list of predicates involved in this analysis:

- *í* 'in': *stytlast í* + ACC 'approach', *vera tilbúinn í* + ACC 'be ready to', *pæla í* + DAT 'reflect upon'
- *yfir* 'over': *komast yfir* + ACC 'get over', *vera svekktur yfir* + DAT 'be annoyed about', *gleðjast yfir* + DAT 'rejoice about'

<sup>35</sup> *Langa* is actually not the only verb that displays this behavior. Recall the case of *þvinga* 'force' in Chapter 2, which cannot be followed by a NomCI as complement, as the clause itself is in a non-DP position. Verbs like *þvinga* have the option of selecting a prepositional phrase instead of a clausal complement. The preposition that is adopted is also *til*:

- (i) Hann þvingaði mig **til** að fara  
 he forced me to to go  
 'He forced me to go.'

A quick search in RMH shows that *þvinga* in constructions like the one above has a distribution of clausal nominalization similar to the one of *langa* (1936 vs. 303; 13.53%), which supports the hypothesis that the default complementation pattern of *langa* affects the distribution of clausal nominalization with *til*.

The data from the questionnaires are presented in Table 3.25 and are arranged on the basis of case assignment:<sup>36</sup>

| ACCUSATIVE |                                                                                |      |       |       |       |
|------------|--------------------------------------------------------------------------------|------|-------|-------|-------|
| No.        | Sentence                                                                       | Type | -það  | +það  | p     |
| 1A-17      | <i>Nú styttest í það að flautað verði til leiks á Laugardalsvelli</i>          | tht  | 6.54  | 5.50  | .004  |
| 1B-09      | 'Now the time for the game to start in Laugardagsvöllur is approaching.'       |      | (69)  | (126) |       |
| 8A-11      | <i>Hún er ekki tilbúin í það að eignast barn svona ung</i>                     | inf  | 5.01  | 4.46  | .080  |
| 8B-15      | 'She is not ready to have a child at such a young age.'                        |      | (132) | (76)  |       |
| 7A-05      | <i>María var ekki tilbúin í það að reka Sverri úr starfi</i>                   | inf  | 3.30  | 3.73  | .555  |
| 7B-21      | 'María was not ready to fire Sverrir.'                                         |      | (23)  | (26)  |       |
| 10A-23     | <i>Enginn getur komist yfir það að lesa allar bókmenntir heimsins</i>          | inf  | 6.30  | 5.71  | .096  |
| 10B-03     | 'No one can manage to read all the world's literature.'                        |      | (30)  | (34)  |       |
| DATIVE     |                                                                                |      |       |       |       |
| No.        | Sentence                                                                       | Type | -það  | +það  | p     |
| 8A-14      | <i>Þeir eru mjög svektir yfir því að tapa leiknum</i>                          | inf  | 3.91  | 5.27  | <.001 |
| 8B-12      | 'They are very upset about having lost the game.'                              |      | (76)  | (132) |       |
| 6A-23      | <i>Pau hafa aldrei þælt í því að þetta handrit geti verið fölsun</i>           | tht  | 3.57  | 3.14  | .515  |
| 6B-03      | 'They never thought that this manuscript could be false.'                      |      | (21)  | (21)  |       |
| 4A-10      | <i>Ég er að þæla í því að kaupa hlutabréf</i>                                  | inf  | 6.24  | 5.72  | .944  |
| 4B-16      | 'I am thinking of buying assets.'                                              |      | (21)  | (36)  |       |
| 5A-02      | <i>Ég þarf ekki að þæla í því hvað aðrir eru að hugsa</i>                      | que  | 5.04  | 5.95  | .062  |
| 5A-24      | 'I don't have to reflect upon what people are thinking.'                       |      | (23)  | (20)  |       |
| 8A-05      | <i>Við eigum að gleðjast yfir því að framkvæmdir séu loksins hafnar</i>        | tht  | 4.82  | 5.87  | <.001 |
| 8B-21      | 'We have to be happy about the fact that the operations have finally started.' |      | (132) | (76)  |       |

Table 3.25: Questionnaires - Double case prepositional complements

In relation to accusative case, we can observe that nominalization is acceptable in the majority of cases (7A-05 and 7B-21 received less positive judgments, but there is no significant difference between the two variants), which makes the results consistent with accusative case in clausal objects as well as with accusative case with the preposition *um*. As for dative case, the examples of *þæla í* and *gleðjast yfir* are for the most part quite acceptable (apart from 6A-23 and 6B-03 which received less positive judgments, but there is no significant difference between them), although the results on the latter suggest that *það* is more felicitous. However, as for *vera svektur*, the non-NomCl variant appears to be more degraded than the one in *gleðjast yfir*. It is possible that this discrepancy is due to the fact that *vera svektur yfir* is less common than *gleðjast yfir* (the data from RMH confirm this view; see below).

Let us now take a look at the data from corpora, which are presented in Table 3.26. I divided the datasets on the basis of case in order to verify whether there is a distinction between accusative and dative:

<sup>36</sup>Before testing 4A-10 and 4B-16, I had already tested a comparable example in 2A-22 and 2B-04 (see Appendix B), but at first it obtained highly negative judgments for both variants. That example is not going to be presented here, as 4A-10 and 4B-16 obtained more positive judgments from speakers overall.

| ACCUSATIVE   |                        |                             |                                |
|--------------|------------------------|-----------------------------|--------------------------------|
|              | <i>styttest í+acc.</i> | <i>vera tilbúinn í+acc.</i> | <i>komast yfir+acc.</i>        |
| <b>Valid</b> | 2964                   | 837                         | 196                            |
| <b>-það</b>  | 2773                   | 776                         | 147                            |
| <b>+það</b>  | 191                    | 61                          | 49                             |
| <b>%það</b>  | <b>6.44%</b>           | <b>7.29%</b>                | <b>25.00%</b>                  |
| DATIVE       |                        |                             |                                |
|              | <i>pæla í+dat.</i>     | <i>gleðjast yfir+dat.</i>   | <i>vera svekktur yfir+dat.</i> |
| <b>Valid</b> | 802                    | 1774                        | 780                            |
| <b>-það</b>  | 377                    | 253                         | 189                            |
| <b>+það</b>  | 425                    | 1521                        | 591                            |
| <b>%það</b>  | <b>52.99%</b>          | <b>85.74%</b>               | <b>75.77%</b>                  |

Table 3.26: RMH - Preposition *yfir* and *í* with accusative and dative case

Overall, the frequencies for accusative case suggest that NomCIs are once again less frequent, in line with our observations on *um* as well as structural case positions. As for dative case, the frequencies tend to be over 50% in a similar way as we have seen with post-copular clauses, in particular with *gleðjast yfir* and *vera svekktur yfir*. Apart from the possibility that *vera svekktur yfir* is less common than *gleðjast yfir*, I could not observe any significant difference between these two verbs, which indicates that whether the matrix verb is a stand-alone verb or not is irrelevant to the distribution of NomCIs and non-NomCIs.

As for the verb *pæla í*, the data from questionnaires and corpora suggest that *það* is optional, as the frequency of *það* almost reaches 53% while the examples from the questionnaires show that both variants are equally acceptable. This differs from what we have seen so far on dative case, as the frequencies of *það* have generally been higher. Hence, the question is whether the preposition *í* is an exception, in particular in comparison to *yfir* (possibly also *á*, as we are going to see that it behaves like *yfir*; see §3.3.4.5), or whether *pæla í* specifically is an exception for some reason. I believe the former possibility is not viable, as one can find in RMH other examples of predicates with *í* + DAT where the frequency of NomCIs is high, e.g. *vera fólgin í* 'be about' (88 vs. 497; 84.95%) and its stand-alone counterpart *felast í* 'be about' (5355 vs. 22414; 80.71%), as well as *lenda í* (641 vs. 4967; 88.56%). As for the latter option, my view is that the percentage displayed in the *pæla í* dataset only looks like an exception due to conflicting results on the basis of clause type. In fact, *that*-clauses preceded by *pæla í* display a high frequency of *það* (20 vs. 71; 78.02%), whereas NomCIs are slightly more frequent than non-NomCIs in embedded questions (157 vs. 229; 59.33%) and less frequent in infinitives (200 vs. 125; 38.46%).<sup>37</sup> In other words, the general percentage with *pæla í* is mainly caused by a contrast between infinitives and embedded questions. This raises an important question, i.e. whether the type of clause has some impact on the distribution of NomCIs and non-NomCIs. I will put this issue on hold for the moment as we still have to discuss the general data on other syntactic positions. However, I will return to this topic in

<sup>37</sup>These results are shown in Appendix A.

§3.3.9, after discussing all the general data from the questionnaires and corpora. This is because only a few datasets display a similar behavior as the one in *pæla í*, so I prefer to discuss them all together.

### 3.3.4.5 Functional prepositions preceded by an adverb

Now, we can consider the data on functional prepositions that are preceded by adverbs. As previously mentioned, we will take into account the following datasets:

- only accusative: *komast upp um* ‘come to light’
- only dative: *ganga út frá* ‘assume’
- dative or accusative: *vanta upp á* + ACC ‘lack’, *finna upp á* + DAT ‘find a way to do sth.’

On the basis of the data just discussed, we expect NomCIs to be less frequent with accusative case and more frequent with dative. The data from the questionnaires are presented in Table 3.27 and are largely consistent with our previous observations:

| No.   | Sentence                                                                      | Type | -það  | +það  | p     |
|-------|-------------------------------------------------------------------------------|------|-------|-------|-------|
| 9A-08 | <i>Það vantaði lítið upp á það að Chelsea næði stigi gegn Arsenal</i>         | tht  | 6.05  | 4.87  | .002  |
| 9B-18 | ‘Chelsea was so close to drawing against Arsenal.’                            |      | (65)  | (101) |       |
| 9A-14 | <i>Eftir slysið komst upp um það að bíllinn var ótryggður</i>                 | tht  | 2.45  | 2.84  | .384  |
| 9B-12 | ‘After the accident, it came to light that the car was not insured.’          |      | (65)  | (101) |       |
| 9A-17 | <i>Börn geta fundið upp á því að troða ótrúlegustu hlutum upp í nefið</i>     | inf  | 5.81  | 6.08  | .183  |
| 9B-09 | ‘Children can find a way to stick the most incredible stuff into their nose.’ |      | (101) | (65)  |       |
| 9A-20 | <i>Við getum ekki gengið út frá því að sagan í þessari grein sé sönn</i>      | tht  | 4.74  | 6.13  | <.001 |
| 9B-06 | ‘We cannot assume that the story in this article is true.’                    |      | (65)  | (101) |       |

Table 3.27: Questionnaires - Functional prepositions preceded by adverbs

The results on the preposition *á* (preceded by the adverb *upp*) with dative and accusative case confirm our predictions due to the fact that both variants are considered acceptable, although in the case of *vanta upp á* there is a significant statistical difference between the two variants. As for *komast upp um*, judgments from speakers on accusative NomCIs and non-NomCIs are very similar, although they are quite low, most probably due to the fact that *komast upp um* is not commonly used (see the data from RMH below). On the other hand, with *ganga út frá*, the NomCI variant was attributed more significantly positive judgments than the non-NomCI one, although both variants are considered acceptable.

Let us now add the data from corpora to our discussion. The results are illustrated in Table 3.28:

|              | <i>komast upp um+acc.</i> | <i>ganga út frá+dat.</i> | <i>vanta upp á+acc.</i> | <i>finna upp á+dat.</i> |
|--------------|---------------------------|--------------------------|-------------------------|-------------------------|
| <b>Valid</b> | 53                        | 3785                     | 1172                    | 118                     |
| <b>-það</b>  | 48                        | 250                      | 1140                    | 21                      |
| <b>+það</b>  | 5                         | 3535                     | 32                      | 97                      |
| <b>%það</b>  | <b>9.43%</b>              | <b>93.39%</b>            | <b>2.73%</b>            | <b>82.20%</b>           |

Table 3.28: RMH - Functional prepositions preceded by adverbs

The data from RMH on the preposition *á* corroborate our predictions, since the frequency of *það* is low with accusative case (see *vanta upp á*), while it is high with dative (see *finna upp á*). Moreover, the data on *ganga út frá* suggest that NomCIs are preferable. This result aligns with the data from the questionnaires, as the weighted average of judgments attributed to the NomCI variant is much higher than the one of the non-NomCI counterpart. As for *komast upp um*, the results suggest that NomCIs are less frequent, in line with our observations on accusative case so far.

At this point, considering the fact that the data on prepositional phrases up to this point make a distinction between accusative case on the one hand and dative and genitive case on the other hand (except the case of *langa til*, which we have already discussed) in a similar way as in canonical argument positions, we can revisit (17), by being more specific in condition (a):

(25) *Non-default feature effect on clausal nominalization (revisited)*

A NomCI is at least preferable in a syntactic position if (fulfilling one of the following conditions is enough):

- a) it is assigned dative or genitive case
- b) it is assigned non-default gender
- c) it is assigned non-default number

The analysis of clausal nominalization in (25) suggests that, for clausal nominalization to be preferred, it is enough that one non-default gender, number or case feature is assigned to the relevant clausal complement. Moreover, the fact that non-default case features tend to trigger clausal nominalization has an important implication, namely that the distinction between nominative/accusative and dative/genitive is extended beyond the boundaries of the structural vs. lexical case dichotomy (see further discussion on structural and lexical case in §4.2.3), as the same distinction can be observed with functional prepositions.

### 3.3.4.6 Lexical prepositions

Let us now take a look at lexical prepositions. The prepositions involved are *í kringum* ‘around’ for accusative case, *gagnvart* ‘towards’ for dative case and *utan* ‘apart from’ for genitive case. Let us take a look at the results from the questionnaires, which are presented in Table 3.29:

| No.   | Sentence                                                                                       | Type | -það  | +það  | p     |
|-------|------------------------------------------------------------------------------------------------|------|-------|-------|-------|
| 7A-14 | <i>Gögn þáttakenda eru ekki aðgengileg utan þess að aldur þeirra er birtur á vefsíðunni</i>    | tht  | 2.69  | 4.78  | .003  |
| 7B-12 | 'User data are not accessible apart from the fact that their age is published on the website.' |      | (26)  | (23)  |       |
| 8A-08 | <i>Það var mikið ferli í kringum það að byggja þetta hótél</i>                                 | inf  | 3.26  | 3.96  | .024  |
| 8B-18 | 'There was a huge process around the construction of this hotel.'                              |      | (76)  | (132) |       |
| 9A-05 | <i>Sumt fólk upplifir verkkviða gagnvart því að byrja á verkefnum</i>                          | inf  | 2.42  | 4.72  | <.001 |
| 9B-21 | 'Some people experience job anxiety at the moment of starting projects.'                       |      | (101) | (65)  |       |

Table 3.29: Questionnaires - Lexical prepositions

The results on *utan* and *gagnvart* suggest that the NomCl variant is much more acceptable than a non-NomCl, aligning with the idea that nominalization is triggered when dative or genitive is assigned to the relevant clause. However, the data on *í kringum* are mysterious, since the *p*-value indicates that there is a significant statistical difference between the two variants despite the fact that the two weighted averages are very close. For this reason, I investigated the judgments on *í kringum* more closely. These judgments, arranged on the basis of the value chosen by participants on a scale from 1 to 7, are illustrated in Table 3.30:

| Sentence          | 1  | 2  | 3  | 4  | 5  | 6  | 7  |
|-------------------|----|----|----|----|----|----|----|
| 8A-08 (NomCl)     | 26 | 18 | 14 | 10 | 26 | 17 | 21 |
| 8A-18 (non-NomCl) | 27 | 9  | 8  | 5  | 10 | 10 | 7  |

Table 3.30: Judgments on 8A-08 and 8B-18

We can see from the table that many more speakers judged the NomCl variant positively rather than the non-NomCl variant, that is 64 speakers vs. 27 (considering the judgments from 5 to 7 only). This significant difference suggests that *það* is most likely needed with *í kringum*. But let us also compare these results with the data from corpora. They are presented in Table 3.31:

|              | <i>í kringum + acc.</i> | <i>gagnvart + dat.</i> | <i>utan + gen.</i> |
|--------------|-------------------------|------------------------|--------------------|
| <b>Valid</b> | 215                     | 1144                   | 1230               |
| <b>-það</b>  | 25                      | 4                      | 513                |
| <b>+það</b>  | 190                     | 1140                   | 717                |
| <b>%það</b>  | <b>88.37%</b>           | <b>99.65%</b>          | <b>58.29%</b>      |

Table 3.31: RMH - Lexical prepositions

The data on dative case corroborate the idea that *það* is mandatory with prepositions like *gagnvart*. As for *utan*, the data from RMH apparently contrast with the data from the judgment tasks, as the non-NomCl variant is considered degraded by participants while the frequency of *það* only reaches 58.29% (which suggests that the pronoun might be optional). How can we explain this pattern in *utan*? This apparent issue is caused once again by a contrast between clause types (similarly to what we have observed with *pæla í*). *That*-clauses and infinitives after *utan* display a higher frequency of *það* (65.23% and 79.70% respectively), whereas embedded questions display a very low one (2.04%; see further details in Appendix A). The structure of embedded questions with *utan* is quite uncommon (and even degraded according to certain speakers I discussed this

construction with) and, consequently, I will discuss this exception in §3.3.9 after analyzing the syntactic positions that we still need to observe. If we focus exclusively on *that*-clauses and infinitives in the case of *utan* in our study of lexical prepositions, the results suggest that NomCIs are preferable (in particular with infinitives), in line with the results from questionnaires.

Regarding accusative case, where we would expect to find a low frequency of *það* based on the results for other prepositions like *um* or *yfir*, we find the opposite situation (88.37%). Considering the data from the questionnaires on *í kringum* presented above, the results overall show that *það* is preferable with this preposition. This contrasts with the results on accusative case we have discussed so far. The question is whether *í kringum* is just idiosyncratic or whether lexical prepositions that assign accusative case follow a pattern. The latter hypothesis is supported by other lexical prepositions like *umfram* ‘beyond’. A simple search in RMH shows that the frequency of *það* with this preposition is high (45 vs. 321; 87.70%). Lastly, as to why accusative-assigning lexical prepositions behave differently from other instances of accusative we have already discussed, my view is that the need for a NomCI is not triggered by case assignment in this case, but rather by DP selection, in the sense that lexical prepositions, unlike functional prepositions, tend to subcategorize for DPs independently of the case assigned.

### 3.3.4.7 Final observations on prepositional phrases

Now that we have taken a look at the data on various types of PPs, we see that there is a distinction between lexical and functional prepositions. The former tend to subcategorize for a DP, which causes NomCIs to be highly frequent independently of case. The latter tend to follow a similar pattern to the one of canonical argument positions. Accusative correlates with a lower frequency of NomCIs, while *það* is much more likely to emerge or is even required with dative- and genitive-assigning prepositions (except with verbs like *langa til*). This pattern also extends to double-case prepositions.

An important question in relation to functional prepositions is why it is that a dative non-NomCI is acceptable with double case prepositions (as with *í*, *yfir* and *á*), but much less with prepositions that assign only dative (like *af* and *hjá* and, possibly, also *frá*). Let us consider two possible options to explain this contrast. The first option is strictly connected to the nature of double case prepositions. Since prepositions like *í* and *yfir* are able to take an accusative complement, it would also entail that, by default, non-NomCIs are acceptable complements of such prepositions (this inference is also based on the data collected on accusative clausal direct objects, where non-NomCIs are more frequent). It is possible that the correlation between accusative case and acceptability of both variants causes non-NomCIs to be possible complements in dative case as well, which entails that *því* just appears preferable instead of mandatory. This first option, to be clear, is based on the assumption that prepositions like *af* and *hjá* represent the rule (also on the basis of our observations in canonical argument positions) while prepositions like *í* and *yfir* represent an exception.

However, it is also possible to see all this backwards, which is the premise of the second

hypothesis I present here. What if *í* and *yfir* represent the rule, i.e. functional prepositions tend by default to accept both variants? This would entail that prepositions like *af* or *hjá* have an exceptional property that distinguishes them from other prepositions by restricting non-NomCIs. It is possible that one property that makes them different is meaning. Prepositions like *hjá* or *af* in our observations have a clearer meaning compared to prepositions like *yfir* in the examples we analyzed. Therefore, we could argue that *hjá* and *af* can be classified as semi-functional (or semi-lexical, see e.g. Mardale (2011)) prepositions, in the sense that they are subcategorized for like functional prepositions but they still convey a clear meaning, which is more typical of lexical prepositions. By virtue of this additional property, these prepositions could be able to require nominalization like lexical prepositions.

In order to exclude the first option, we should be able to find double case prepositions that require *það* obligatorily in dative case, whereas we could exclude the second option if we find case prepositions assigning one case that have no clear meaning and have the same outcome of *hjá* and *af* with clausal nominalization. In order to verify the validity of my first hypothesis, I extended my research in RMH by quickly looking at different double case prepositions (selected by different lexical items like nouns, adjectives and verbs) assigning dative case. I could not find any preposition that would represent a clear exception, as, by average, the frequency of *það* does not corroborate pronominal obligatoriness. Here are some examples:

- *fyrir* ‘for’: *berjast fyrir* ‘fight for’ (523 vs. 5243; 90.92%), *áhugi fyrir* ‘interest for’ (330 vs. 655; 66.49%), *spenntur fyrir* ‘excited for’ (1215 vs. 4379; 78.28%)
- *eftir* ‘after’: *misjafn eftir* ‘dependent on’ (19 vs. 172; 90.05%), *taka eftir* ‘become aware’ (2410 vs. 5926; 71.08%), *sjá eftir* ‘regret’ (804 vs. 1751; 68.53%)
- *við* ‘at’: *bregðast við* ‘react to’ (121 vs. 493; 80.29%), *svar við* ‘answer to’ (80 vs. 1794; 95.73%), *vara við* ‘warn’ (1695 vs. 8668; 83.64%)
- *undir* ‘under’: *standa undir* ‘be able to bear’ (23 vs. 291; 92.67%), *eiga undir* ‘depend on’ (57 vs. 388; 87.19%)

On the other hand, regarding my second hypothesis, I found one single case preposition that can help us. I looked for examples with the preposition *að* ‘at, by’ on RMH. This preposition has a very abstract meaning and only assigns dative case. The results from RMH on different lexical items selecting *að* suggest that *það* is mandatory, as the frequencies are not lower than 99%:

- *komast að* ‘find out’ (191 vs. 20831; 99.09%), *miða að* ‘aim to’ (8 vs. 18662; 99.95%), *einbeittur að* ‘focused on’ (0 vs. 134; 100%), *hlæja að* ‘laugh about’ (1 vs. 213; 99.53%)

Since *að* requires *það* regardless of the fact that it does not convey any clear meaning, we can confidently discard the second hypothesis I proposed, namely that prepositions like *hjá* and *af* require *það* on the basis of a less abstract meaning than in double case prepositions like *yfir* and *í*. Empirical data rather support the possibility that double case prepositions allow non-NomCIs to

surface in positions where dative case is assigned by virtue of the fact that these prepositions are also able to assign accusative case, which allows non-NomCIs to emerge (further discussion on this issue is offered in §4.4).

### 3.3.5 Complements of nouns

Let us now consider the data on complements of nouns. Across languages, it is very common to find genitive complements of nouns, and Icelandic is not an exception. However, there is also a minority of nouns in Icelandic which can select dative complements. So, in relation to clausal complements of nouns, we have to focus on both cases. For genitive case, I selected two nouns that also have a fair amount of examples from RMH, namely *hagkvæmni* ‘practicality’ and *lögmæti* ‘legality’. As for nouns assigning dative case, considering that they are a minority, I decided to only test the noun *vernd* ‘defense, protection, prevention’.

Let us start by taking a look at the data from the questionnaires. They are presented in Table 3.32:

| No.    | Sentence                                                                                 | Type | -það | +það  | p     |
|--------|------------------------------------------------------------------------------------------|------|------|-------|-------|
| 1A-11  | <i>Verið er að kanna hagkvæmni þess að byggja sundlaug undir íþróttahúsinu</i>           | inf  | 1.64 | 5.83  | <.001 |
| 1B-15  | ‘The practicality of building a swimming pool under the sports centre is being checked.’ |      | (69) | (126) |       |
| 5A-20  | <i>Fyrirtækið kannaði hagkvæmni þess að vöruhúsið yrði byggt við ána</i>                 | tht  | 1.65 | 6.67  | <.001 |
| 5B-06  | ‘The company checked the feasibility of building the storage by the river.’              |      | (23) | (20)  |       |
| 7A-23  | <i>Ákvæðið er sett til verndar því að upplýsingar viðskiptavina verði misnotaðar</i>     | tht  | 1.87 | 4.85  | <.001 |
| 7B-03  | ‘The clause is set to prevent information on customers from being misused.’              |      | (23) | (26)  |       |
| 10A-08 | <i>Hún efast um lögmæti þess að loka opinberri stofnun með þessum hætti</i>              | inf  | 2.82 | 6.67  | <.001 |
| 10A-18 | ‘She doubts the legality of closing a public institution in this way.’                   |      | (34) | (30)  |       |

Table 3.32: Questionnaires - Complements of nouns

The results clearly indicate that NomCIs are required both with genitive and dative case. The data from RMH corroborate this analysis for genitive case, as shown in Table 3.33:

|              | <i>hagkvæmni</i> | <i>lögmæti</i> |
|--------------|------------------|----------------|
| <b>Valid</b> | 822              | 265            |
| <b>-það</b>  | 12               | 2              |
| <b>+það</b>  | 810              | 263            |
| <b>%það</b>  | <b>98.54%</b>    | <b>99.25%</b>  |

Table 3.33: RMH - Complements of nouns

In relation to dative case, unfortunately, only one valid example from RMH has been found, which contains a nominalized *that*-clause:

- (26) [...] líka ákvæðin sem eru einungis sett til verndar **því** að ekki sé gengið  
 also provisions.the which are only set to preventing that<sub>DAT</sub> that not is gone  
 fram hjá hæfari konum [...] past more capable women  
 ‘... also the provisions which are only set to prevent more capable women from being overlooked...’

Overall, the results suggest that the pronoun is mandatory both when genitive and dative case is assigned. However, regarding the trigger of clausal nominalization, complements of nouns are ambiguous as we can attribute obligatory *það* either to the fact that dative and genitive case are assigned, or to the fact that these nouns can only subcategorize for a DP. It is difficult to test either possibility, as we cannot use accusative case here as a diagnostic test (consider the contrast between functional and lexical prepositions, where only with the latter accusative NomCIs were preferable, which could entail that lexical prepositions in general tend to need a DP complement). Therefore, I cannot point out a specific rule behind clausal nominalization in complements of nouns, as both options are valid.

### 3.3.6 Complements of adjectives

Let us now observe the results from complements of adjectives. Here are the adjectives I picked specifically for our analysis:

- acc: *viðstaddur* ‘present’
- dat: *háður* ‘dependent’, *feginn* ‘satisfied, happy’
- gen: *minnugur* ‘mindful’

For dative case, I selected *háður* and *feginn* and not just one adjective as for the other cases, due to the contrast we observed in Chapter 2 between these two adjectives, based on which *feginn* is followed by an optional *það* while *háður* requires nominalization. The examples are re-illustrated here for simplicity:

- (27) a. Ég er feginn (**því**) að þú skulir vera kominn  
 I am satisfied that<sub>DAT</sub> that you shall be arrived  
 ‘I’m happy that you have come.’
- b. Sum hagkerfi heimsins eru háð **\*(því)** að fá orku úr  
 some economic systems world.the<sub>GEN</sub> are dependent that<sub>DAT</sub> to get energy from  
 jarðgasi  
 natural gas  
 ‘Some economic systems depend on getting energy from natural gas.’

Let us take a look at the data for each case. I will start with accusative and genitive first as they are easier to discuss, and then I will present the results on dative.

### 3.3.6.1 Accusative

Adjectives assigning accusative case are rare in Icelandic (see Jónsson (2005b:373)), so I did not expect to find significant results from the questionnaires or corpora. Let us first take a look at the data from the judgment tasks on the adjective *viðstaddur*:

| No.   | Sentence                                                                      | Type | -það | +það  | p     |
|-------|-------------------------------------------------------------------------------|------|------|-------|-------|
| 8A-20 | <i>Sara sagðist ekki vilja vera viðstödd það að hundurinn væri sprautaður</i> | tht  | 1.45 | 2.59  | <.001 |
| 8B-06 | 'Sara said she did not want to be there when the dog gets the injection.'     |      | (76) | (132) |       |

Table 3.34: Questionnaires - Accusative adjectival complements

Even if the example has not received very positive judgments in general, a NomCI is still significantly better than a non-NomCI ( $p < .001$ ). This is also confirmed if we take a closer look at the judgments in Table 3.35, where we can see that there are more speakers who tend to consider the NomCI acceptable than the non-NomCI one:

| Sentence          | 1  | 2  | 3  | 4 | 5  | 6 | 7  |
|-------------------|----|----|----|---|----|---|----|
| 8A-20 (NomCI)     | 62 | 19 | 16 | 6 | 15 | 3 | 11 |
| 8B-06 (non-NomCI) | 63 | 6  | 2  | 1 | 0  | 3 | 1  |

Table 3.35: Judgments on 8A-20 and 8B-06

Only 4 participants assigned the non-NomCI a value from 5 to 7, while 29 participants returned a positive judgment on the NomCI variant, which might suggest that *það* is required after the adjective. As for the data from RMH, only one valid example has been found, which contains a NomCI:

- (28) [...] hún hefði [...] verið viðstödd það að menn úr nefnd Katar hafi boðið  
 she had been present that<sub>ACC</sub> that men from committee Qatar have offer  
 fé á stjórnarmenn FIFA fyrir atkvæði þeirra í kosningunni  
 money to managers FIFA for vote their in election.the  
 '... she had been [...] present when people from the Qatari committee offered money to  
 the FIFA managers in exchange for their vote in the election.'

At least the results from the questionnaires tell us that there is the possibility that nominalization is required considering the fact that a NomCI is considered a significantly better variant than a non-NomCI.

### 3.3.6.2 Genitive

Let us now focus on the adjective *minnugur* 'mindful', which assigns genitive case. The data from the questionnaires are presented in Table 3.36:

| No.   | Sentence                                                      | Type | -það | +það | p     |
|-------|---------------------------------------------------------------|------|------|------|-------|
| 4A-19 | <i>KR-ingar voru minnugir þess að tapa 2-0 fyrir Völsurum</i> | inf  | 1.53 | 5.19 | <.001 |
| 4B-07 | 'KR players were mindful of having lost 2-0 against Valur.'   |      | (36) | (21) |       |

Table 3.36: Questionnaires - Genitive adjectival complements

The results indicate that the clausal complement of *minnugur* must be nominalized. As for the data from corpora, they are presented in Table 3.37:

| Adjectives | Valid | -það | +það | %það    |
|------------|-------|------|------|---------|
| minnugur   | 534   | 0    | 534  | 100.00% |

Table 3.37: RMH - Genitive adjectival complements

These results strongly confirm the data from the judgment tasks, as the frequency of *það* reaches 100%.

### 3.3.6.3 Dative

In the case of adjectives assigning dative, let us take a look at the results for *háður* and *feginn*. The data from the questionnaires are consistent with the original contrast I showed in Chapter 2, as both variants are acceptable with *feginn* and only NomCIs with *háður*:<sup>38</sup>

| No.   | Sentence                                                                              | Type | -það | +það  | p     |
|-------|---------------------------------------------------------------------------------------|------|------|-------|-------|
| 1A-05 | <i>Íbúarnir eru háðir því að bændur í nágrenninu sjái þeim fyrir mat</i>              | tht  | 1.77 | 6.31  | <.001 |
| 1B-21 | 'The inhabitants depend on farmers in the neighbourhood who take care of their food.' |      | (69) | (126) |       |
| 5A-23 | <i>María er feginn því að prófin séu loksins búin</i>                                 | tht  | 6.00 | 5.87  | .652  |
| 5B-03 | 'María is happy that the exams are finally finished.'                                 |      | (20) | (23)  |       |

Table 3.38: Questionnaires - Dative adjectival complements

In light of the empirical data observed so far across different syntactic positions, we should expect *það* to be preferable with both adjectives on the basis of the data from RMH (recall the results on dative case e.g. in prepositional phrases, where the frequency of *það* tends to be over 50% at least). However, the results shown in Table 3.39 tell us that *það* is not frequent with *feginn*, but it is very frequent with *háður*:

|              | <i>feginn</i> | <i>háður</i> |
|--------------|---------------|--------------|
| <b>Valid</b> | 5729          | 2689         |
| <b>-það</b>  | 4840          | 7            |
| <b>+það</b>  | 889           | 2682         |
| <b>%það</b>  | 15.52%        | 99.74%       |

Table 3.39: RMH - Dative adjectival complements

Now, the main question is how we can interpret the data presented here. First of all, I believe that *feginn* here is the exception rather than the rule, as other adjectives I quickly checked in RMH, in

<sup>38</sup>I also tested further examples with *háður* with a different clause type (see 3A-11/3B-15 and 6A-05/6B-21 in Appendix B), but the results were similar to the ones with a *that*-clause. Therefore, they are not presented here.

order to identify a possible pattern, show that *það* is quite frequent, e.g. *sammála* ‘agreeing’ (836 vs. 4979; 85.62%), *ósammála* ‘disagreeing’ (84 vs. 784; 90.32%), *andvígur* ‘opposed’ (155 vs. 3084; 95.21%), *mótfallinn* ‘opposed’ (11 vs. 1316; 99.17%). Therefore, it seems that, even in this case, adjectives tend to prefer or require clausal nominalization.

But why is *feginn* an exception to the rule? Compared to other adjectives like *háður* or *andvígur*, *feginn* does not require a complement in a default context, which might open the possibility for a CP to emerge:

- (29)
- a. **María er feginn**  
María is satisfied  
‘Maria is satisfied.’
  - b. \***Karl er háður**  
Karl is dependent  
Literally: ‘Karl is dependent.’
  - c. \***Deildarforseti er andvígur**  
faculty.chair is opposed  
Literally: ‘The Chair of the Faculty is opposed.’

Considering these facts, it is possible that clausal nominalization can be dropped if the relevant adjective has no strong requirement for a complement. The reader, however, might wonder whether this rule can be generalized to other syntactic positions, i.e. whether pronominal obligatoriness depends on the fact that the lexical item involved requires a complement. Let us consider, for instance, lexical prepositions. Prepositions like *gagnvart* or *utan* tend to need a NomCl. But these prepositions need to be followed by a complement as a general rule:

- (30)
- a. \***Við berum ábyrgð gagnvart**  
we bear responsibility towards  
Literally: ‘We are responsible towards.’
  - b. \***Þú mátt ekki taka próf utan**  
you may not take exams outside  
Literally: ‘You may not take exams apart from’

In other words, the interpretation proposed here for adjectives can be extended to lexical prepositions, as obligatory nominalization and requirement for a complement might go hand in hand. However, I would not extend this to all syntactic positions. For instance, nouns like *hagkvæmni* ‘practicality’ do not necessarily need to be followed by a genitive complement. However, nominalization in clausal complements of nouns is still mandatory:

- (31) **Staðurinn fyrir verksmiðjuna var valinn með hliðsjón af hagkvæmni**  
place.the for factory.the was chosen with respect to practicality  
‘The location of the factory was chosen by taking practicality into consideration.’ (from *nutimamalsordabok.is*)

Whether we interpret clausal nominalization in complements of nouns as the result of subcate-

gorization for a DP or as a consequence of dative/genitive case assignment to the clausal complement, obligatory nominalization cannot be considered, as a general rule, to be dependent on the need for a complement in light of examples like (31). Therefore, it is a possibility that, in certain syntactic positions, obligatory complements go hand in hand with obligatory or at least highly preferable clausal nominalization. But it is difficult to apply the same principle to all positions where clausal nominalization is mandatory. For our analysis of adjectives, it is necessary for us to know that it is plausible that *feginn* differs from other adjectives because it does not take a mandatory complement.

### 3.3.7 Clausal indirect objects

Let us move now to the analysis of clausal indirect objects. In Chapter 2, I compared dative clausal indirect objects (with the verb *veita* ‘provide’) and accusative clausal indirect objects (with the verb *svipta* ‘deprive’). I showed that nominalization is mandatory both with accusative and dative indirect objects. However, when the clausal argument is passivized, it must be nominalized when dative case is assigned while it is optional when structural case is involved:

- (32) a. Ég veitti [**(því)** að Jón var að gráta] enga athygli  
 I gave that<sub>DAT</sub> that Jón was to cry no attention  
 ‘I paid no attention to the fact that Jón was crying.’ (Garofalo 2020:18)
- b. [**(Því)** að Jón var að gráta] var engin athygli veitt  
 that<sub>DAT</sub> that Jón was to cry was no attention paid  
 ‘No attention was paid to the fact that Jón was crying.’ (Garofalo 2020:25)
- (33) a. Vopnahléið svipti [**(það)** að hermenn fórnðu lífi sínu] öllum hetjuljóma  
 armistice.the deprived that<sub>ACC</sub> that soldiers sacrificed life their all<sub>DAT</sub> heroism<sub>DAT</sub>  
 ‘The armistice deprived the soldiers’ sacrifice of their lives of all heroism.’  
 (Garofalo 2023:113)
- b. [**(Það)** að hermenn fórnðu lífi sínu] var svipt öllum hetjuljóma  
 that<sub>NOM</sub> that soldiers sacrificed life their was deprived all<sub>DAT</sub> heroism<sub>DAT</sub>  
 ‘The fact that soldiers sacrificed their lives was deprived of all heroism.’  
 (Garofalo 2023:113)

I obtained these judgments initially from a handful of speakers, so I decided to test their validity in the questionnaires (I did not look for examples in RMH as clausal indirect objects are rare). Some Icelandic students in linguistics advised me to test them by extraposing the clausal indirect object while leaving the pronoun *in situ*, as they predicted it would appear more acceptable to speakers. The results from the questionnaires are presented in Table 3.40:

| No.   | Sentence                                                                                             | Type | -það | +það | p     |
|-------|------------------------------------------------------------------------------------------------------|------|------|------|-------|
| 4A-13 | <i>Magnús veitti því enga athygli að Sara væri að gráta</i>                                          | tht  | 1.86 | 6.33 | <.001 |
| 4B-13 | 'Magnús did not pay attention to the fact that Sara was crying.'                                     |      | (36) | (21) |       |
| 4A-22 | <i>Heimsfaraldurinn svipti það öllu vægi að liðið hefði æft stíft fyrir næsta mót</i>                | tht  | 1.90 | 1.89 | .541  |
| 4B-04 | 'The pandemic deprived the team's hard training for the following tournament of all its importance.' |      | (21) | (36) |       |
| 6A-08 | <i>Vopnahléið svipti það öllum hetjuljóma að hermenn fórnðu lífi sínu</i>                            | tht  | 2.10 | 2.57 | .337  |
| 6B-18 | 'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'                      |      | (21) | (21) |       |

Table 3.40: Questionnaires - Clausal indirect objects

For dative case, we can clearly see that the pronoun is mandatory. However, the first example I tested on accusative indirect objects received very few positive judgments, i.e. 4A-22 and 4B-04, which led me to test another sentence, see 6A-08 and 6B-18. This second pair of examples did not obtain many positive judgments.<sup>39</sup> Just to be sure, I also tested a non-extrapolated variant of 6A-08 and 6B-18 in order to see whether there was any difference, but it received even worse judgments, consistently with the predictions made by students in linguistics (see example 7A-11 and 7B-15 in Appendix B). Furthermore, I also tested passivized examples like (32b), which also received highly negative judgments, in contrast with 4A-13 (see 6A-14 and 6B-12 in Appendix B).<sup>40</sup>

Using these methods, it was not possible to get a clear picture of clausal nominalization in the indirect object position. Consequently, I decided to apply a different methodology, only in the case of clausal indirect objects: I opted to interview orally a group of native speakers of Icelandic ( $n = 17$ ) with the purpose of gathering some more detailed feedback from them, as the results from the questionnaires could not be used to show what was specifically problematic in my examples (see Garofalo (2023)). In these oral interviews, I asked participants to judge my examples on dative and accusative clausal indirect objects, in active and passive voice as well as with and without extraposition (except passivized extrapolated accusative indirect objects, due to the fact that *það* would be homophonous with the expletive in first position).<sup>41</sup> All the examples were read out loud before they were evaluated. This time, however, instead of using a scale from 1 to 7 to let participants evaluate my examples, I listened to their comments on each pair of variants. Participants tended to either consider only one variant possible, or both, or would judge one variant better than the other (further information is offered in Appendix B), but, surprisingly, they

<sup>39</sup>The negative judgments given to the items with *svipta*, in my view, are not caused by some possible case variation. Although it is true that studies like Tómasdóttir (2021) show that there is case variation in Icelandic double object constructions that differ from the typical dative+accusative pattern, it is also true that Tómasdóttir showed that in RMH there are almost no examples of variation with *svipta*. As for judgment tasks, she also showed that a great majority of native speakers would tend to reject examples that do not match the accusative+dative pattern of *svipta*.

<sup>40</sup>I believe these results from the questionnaires have been problematic due to some difficulty for speakers in parsing the examples while reading them. As discussed below, oral interviews returned better results and that is most probably due to the fact that, by hearing the sentence read out loud, they could parse it more easily. Some speakers I interviewed actually had no difficulty at all in judging the examples tested immediately after they were read out loud.

<sup>41</sup>In passive voice, I asked speakers specifically about examples where the passivized indirect object was in the first position and not after the finite verb, as we have seen that Spec,T can only host NomCIs if there is no extraposition. In order to see whether nominalization could become optional, I rather needed to test Spec,C.

did not judge these examples as negatively as I could see in the data from the questionnaires. Rather, they were able to parse the examples properly when they were read out loud to them.

In order to map the various types of feedback I received from participants orally on clausal indirect objects, I encoded the comments into a narrow scale from 1 to 3 (1 = unnatural, 3 = natural). The scale is shown in Table 3.41. As the reader can see, each judgment affects both variants of one specific sentence on the basis of how natural/unnatural they were considered:

| Judgment                                              | NomCI | Non-NomCI |
|-------------------------------------------------------|-------|-----------|
| Both variants are natural and equally valid           | 3     | 3         |
| Both variants are possible, but a NomCI is better     | 3     | 2         |
| Both variants are possible, but a non-NomCI is better | 2     | 3         |
| Only a NomCI is natural                               | 3     | 1         |
| Only a non-NomCI is natural                           | 1     | 3         |
| Neither variant is natural, but a NomCI is better     | 2     | 1         |
| Neither variant is natural, but a non-NomCI is better | 1     | 2         |
| Neither variant is natural                            | 1     | 1         |

Table 3.41: Scale applied to judgments from interviews

Following this scale, the general results from interviews are presented in Table 3.42 (note that the data in the table only show whether NomCIs and non-NomCIs are acceptable independently of extraposition; a breakdown on the basis of extraposition as well as further details are offered in Appendix B). In this case, the minimum and maximum average values to be considered are 1.00 and 3.00 respectively, instead of 1.00 and 7.00 as we have seen in the questionnaires. The results show that a NomCI is much better than a non-NomCI, both when *það* appears in accusative and dative case. However, when the accusative indirect object is passivized and moved to the first position, both variants appear to be highly acceptable.<sup>42</sup>

| DATIVE     |                                                                                                                                                              |      |      |      |       |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|------|-------|
| Type       | Sentence                                                                                                                                                     | Type | -það | +það | p     |
| Active     | <i>Jón veitti því að Sara væri að gráta enga athygli</i>                                                                                                     | tht  | 1.12 | 2.88 | <.001 |
|            | <i>Jón veitti því enga athygli að Sara væri að gráta</i><br>'Jón did not pay attention to the fact that Sara was crying.'                                    |      | (17) | (17) |       |
| Passive    | <i>Því að Sara væri að gráta var engin athygli veitt</i>                                                                                                     | tht  | 1.47 | 2.88 | <.001 |
|            | <i>Því/það var engin athygli veitt að Sara væri að gráta</i><br>'No attention was paid to the fact that Sara was crying.'                                    |      | (17) | (17) |       |
| ACCUSATIVE |                                                                                                                                                              |      |      |      |       |
| Type       | Sentence                                                                                                                                                     | Type | -það | +það | p     |
| Active     | <i>Vopnahléið svipti það að hermenn fórnðu lífi sínu öllum hetjuljóma</i>                                                                                    | tht  | 1.82 | 2.59 | .010  |
|            | <i>Vopnahléið svipti það öllum hetjuljóma að hermenn fórnðu lífi sínu</i><br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.' |      | (17) | (17) |       |
| Passive    | <i>Það að hermenn fórnðu lífi sínu var svipt öllum hetjuljóma</i><br>'The fact that soldiers sacrificed their lives was deprived of all heroism.'            | tht  | 2.94 | 2.88 | .779  |
|            |                                                                                                                                                              |      | (17) | (17) |       |

Table 3.42: Oral interviews - Clausal indirect objects

<sup>42</sup>The results on the basis of extraposition, which are offered in Appendix B, show that the average of judgments on NomCIs is slightly better than the one on non-NomCIs, whether the clause is extraposed or not.

The main question at this point is how we explain: a) why *það* must surface with both accusative and dative case; and b) the contrast between *veita* and *svipta*, where only in the latter case are both variants acceptable when the clausal indirect object is passivized and moved to the first position. Based on our observations so far, accusative case is correlated to the acceptability of both variants in canonical argument positions (as well as a lower frequency of NomCIs, but we will put this aspect aside as we do not have any data from RMH on clausal indirect objects). The fact that this does not happen in active voice here indicates that there is a condition independent of case that causes *það* to be mandatory *in situ*. This is something we can also see with nominative subjects (see §3.3.1), as non-extraposed nominative NomCIs are mandatory after the finite verb. If we assume that the requirement for a subject to be a DP is determined by a feature in T that prevents non-DPs from emerging in subject position, i.e. a D-feature (see Chomsky (1995) and considerable subsequent work within the *Minimalist Program*; also Chapter 4), then the same feature may also be present in Appl (if we follow the *Applicative Head Approach*, see Pylkkänen (2000)).<sup>43</sup> If we assume that Icelandic is characterized by low applicatives<sup>44</sup> and not high applicatives (see Wood (2015a)), the D-feature in Appl would apply a restriction to non-DPs in Spec,Appl, while the direct object in the complement position of Appl is free from this restriction, as CPs can also emerge. Consider the case of the verb *segja* ‘say’:

- (34) Hann sagði mér (**það**) að hann vildi hætta  
 He told me<sub>DAT</sub> that<sub>ACC</sub> that he wanted quit  
 ‘He told me that he wanted to quit.’

Provided that this analysis is correct, if we focus on passivization, the fact that the clausal indirect object of *svipta* can drop nominalization after passivization and movement to the first position indicates that if the clausal argument can escape all the DP-only positions (Spec,Appl and Spec,T) and surface in matrix Spec,C, nominalization is no longer obligatory, and, since both variants are acceptable in instances of structural case (as we have seen in clausal subjects and objects), the pronoun turns out to be optional. The same, at this point, can be said of nominative subjects. If the clausal subject escapes Spec,T, which is a locus of obligatory nominalization on the basis of the data from non-extraposed nominative subjects, and moves to matrix Spec,C, then nominalization can be dropped. An important implication of this interpretation is that, if a clausal subject is unable to reach Spec,C, nominalization is mandatory as it can only surface in Spec,T. This is confirmed by the examples below with clausal subjects, which I already illustrated in §2.5.2 as a counterargument to Thráinsson’s (1979) Factivity Hypothesis:

<sup>43</sup>This possibility is also supported by cross-linguistic analyses that consider valid the presence of an EPP-feature in Appl which is checked by the argument that lands into Spec,Appl (see e.g. Georgala and Whitman (2009) for Modern Greek).

<sup>44</sup>Wood (2015a) also extended high-low applicatives to Icelandic, which are syntactically low but, from a semantic perspective, they apply an argument to an event as a high applicative does. This type of structure does not need to be discussed specifically here, as the analysis I offer is related to the syntactic side of applicatives and, since low and high-low applicatives at least share the same structure, I will focus on low applicatives to make things easier.

- (35) a. ... [C' að [DP \*(**það**) að Jón sé hættur] sé líklegt]  
           that that<sub>NOM</sub> that Jón is quit is likely  
           ‘... that it is likely that Jón quit.’
- b. ... [C' að [DP \*(**það**) að hann skuli vera farinn] sé hörmulegt]  
           that that<sub>NOM</sub> that he shall be gone is deplorable  
           ‘... that it is deplorable that he left.’

As expected, the same effect emerges in ECM constructions. This is confirmed in an example found in Thráinsson (1979:236), shown in (36), where the clausal subject of an ECM construction must be nominalized, as no CP projection over TP is available. Note that *það* is obligatory despite the fact that it surfaces in accusative case, which is generally correlated to pronominal optionality (as well as a lower frequency of *það*) on the basis of the data we have observed so far:

- (36) Jón telur [TP \*(**það**) að María skuli halda við Harald vera alveg  
       Jón considers that<sub>ACC</sub> that María shall have an affair to Haraldur be very  
       hörmulegt]  
       deplorable  
       ‘Jón believes the fact that María has an affair with Haraldur to be very deplorable.’

Considering the distribution of clausal subjects and clausal indirect objects we have just discussed and the possibility for non-NomCIs to surface in Spec,C, an important question these facts raise is how NomCIs and non-NomCIs are derived, in particular the latter since they are not allowed in Spec,T or Spec,Appl. I will discuss this aspect of clausal nominalization in Chapter 4.

### 3.3.8 Nominalization as a feature checking operation

Let us summarize the main findings we have discussed so far. The data from the questionnaires and RMH presented in this chapter have revealed some important facts about the distribution of NomCIs and non-NomCIs, which are summarized in the following table (OBL stands for obligatory, OPT for optional, DISP for dispreferred or less frequent, PREF for preferable or more frequent):

| Position                                              | Status of <i>það</i> | Remarks                               |
|-------------------------------------------------------|----------------------|---------------------------------------|
| Nominative subjects in first position                 | DISP                 | -                                     |
| Nominative subjects after the finite verb (non-extr.) | OBL                  | -                                     |
| Nominative subjects after the finite verb (extr.)     | DISP                 | -                                     |
| Lexical case subjects                                 | OBL                  | -                                     |
| Direct objects with structural case                   | DISP                 | -                                     |
| Direct objects with lexical case                      | PREF/OBL             | -                                     |
| Postcopular clauses neut.sing.                        | DISP                 | -                                     |
| Postcopular clauses not neut.sing.                    | PREF                 | -                                     |
| With functional Ps + only ACC                         | DISP                 | applies to Ps preceded by Adv         |
| With functional/lexical Ps + only DAT                 | OBL                  | PREF with Ps preceded by Adv          |
| With functional/lexical Ps + only GEN                 | PREF                 | except <i>langa til</i> 'want'        |
| With double-case functional Ps + ACC                  | DISP                 | applies to Ps preceded by Adv         |
| With double-case functional Ps + DAT                  | PREF                 | applies to Ps preceded by Adv         |
| With lexical Ps + ACC                                 | PREF                 | -                                     |
| Complements of nouns GEN/DAT                          | OBL                  | -                                     |
| Complements of adjectives ACC                         | PREF                 | -                                     |
| Complements of adjectives DAT/GEN                     | OBL                  | except <i>feginn</i> 'happy'          |
| Indirect objects ACC/DAT                              | OBL                  | OPT with passivization of ACC objects |

Table 3.43: Summary of the findings

A first very important finding is that, if a non-default case, gender or number is assigned to a clausal complement (e.g. dative, feminine or plural), a NomCl is either preferable or mandatory. Recall the following effect on clausal nominalization I presented after observing *það* with nominal predicates and functional prepositions:

(37) *Non-default feature effect on clausal nominalization*

A NomCl is at least preferable in a syntactic position if (fulfilling one of the following conditions is enough):

- a) it is assigned dative or genitive case
- b) it is assigned non-default gender
- c) it is assigned non-default number

Another crucial finding is that nominative clausal subjects in Spec,T and clausal indirect objects must be NomCls and, in particular, that pronominal obligatoriness in these positions is not regulated by case, gender or number, but rather by the fact that a DP is required. Now, let us consider these findings within a broad Minimalist approach to syntax (see Chomsky (1995) and considerable subsequent work; also Chapter 4 of this dissertation, where I will discuss in more detail some fundamental theoretical concepts with the Minimalist Program for my analysis of the derivation of NomCls and non-NomCls), and let us assume within this approach that a) case, gender and number are encoded as features that need to be checked during the derivation, b) subjects occupy Spec,T by checking the D-feature in T (which can only be checked by DPs; see §4.2.2), and c) indirect objects occupy Spec,Appl by checking a D-feature in Appl that can only be checked by DPs (as just proposed in §3.3.7). It appears that when a feature needs to be checked by the embedded clause, *það* usually emerges. In other words, it is plausible that feature checking is the

main trigger of clausal nominalization.

Let us also recall that, in §2.5.7, I had proposed the following mechanisms to map the distribution of NomCIs and non-NomCIs, which indicated that the distribution of NomCIs is subsumed within that of DPs and that case and mandatory DP selection trigger clausal nominalization:

(38) *NomCI selection mechanism*

A NomCI is always allowed if it is in a DP position

(39) *Non-NomCI filter*

A NomCI is required if (it is enough to fulfill one condition):

- a) a case feature must be checked by the relevant clause
- b) the relevant clause is in a position in which a DP is required

Let us combine these two mechanisms with (37) as well as the results related to nominative subjects in Spec,T and clausal indirect objects. Consider the following hypothesis, which I call *Feature Checking Hypothesis*:

(40) *Feature Checking Hypothesis (FCH)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and it is mandatory or at least preferable if:

- a. it checks at least one D-,  $\varphi$ - or case feature
- b. a DP complement is required

If conditions (a) and (b) are not met, CN is preferably absent

Based on FCH, *pað* checks features for the CP it precedes and can only emerge in DP positions. Conditions (a) and (b) account for the instances where *pað* is preferable or even mandatory, as D-features, case features (dative and genitive in particular),  $\varphi$ -features (specifically non-default gender and number) and requirement for a DP are correlated to a higher frequency of *pað* if not pronominal obligatoriness. Note that, specifically in relation to the requirement for a DP described in condition (b), I do not refer to clausal subjects in Spec,T or clausal indirect objects (as they are already accounted for by condition (a)), but rather to e.g. complements of adjectives and lexical prepositions.

If it is correct to assume that case features trigger clausal nominalization, this raises an important question in relation to nominative and accusative, namely whether they are not case features to check in syntax like dative or genitive (unlike e.g. Roussou (1991), who associated obligatory NomCI subjects in Modern Greek to nominative case assignment). In light of the empirical data, my view is that, since nominative and accusative are generally correlated to a lower frequency of *pað*, they do not correspond to case features that need to be checked (cf. Garofalo (2023); see §4.4 for further discussion on this matter). A similar view can be extended to neuter and singular if we consider gender and number ( $\varphi$ -features in condition (a)). In fact, the data on *vandamál* ‘problem’ have shown that a post-copular clause is preferably a non-NomCI only when the subject

is neuter singular. Therefore, we can infer that neuter and singular do not correspond to formal features that need to be checked (see Chapter 4 for further discussion).

### 3.3.9 Clausal nominalization based on clause type

Now that we have discussed FCH, let us move to the observation of some less general details on the empirical data. In the methodology on RMH, I stated that I looked for examples of *that*-clauses, infinitives and indirect questions. So far, we have mostly taken a look at the general results from all the data from the corpora, but, apart from few exceptions, we have not discussed whether there is any relevant systematic difference in the distribution of *það* based on clause type. As a general trend, the data from RMH (which can be consulted in Appendix A; a breakdown of the data based on clause type is illustrated for each dataset) reveal no systematic contrast across syntactic positions, which indicates that the distribution of *það* is not really affected by the type of clause we are dealing with. However, few datasets differ to some extent from the main trend. They are shown in Table 3.44:

| Clause type             | Nom. subj. 1st | <i>spá</i>    | <i>pæla í</i> | <i>utan</i>   |
|-------------------------|----------------|---------------|---------------|---------------|
| <i>tht</i> – <i>það</i> | 628            | 972           | 20            | 267           |
| <i>tht</i> + <i>það</i> | 991            | 11786         | 71            | 501           |
| <i>tht</i> % <i>það</i> | <b>61.21%</b>  | <b>92.38%</b> | <b>78.02%</b> | <b>65.23%</b> |
| <i>inf</i> – <i>það</i> | 6419           | 12            | 200           | 54            |
| <i>inf</i> + <i>það</i> | 1440           | 4             | 125           | 212           |
| <i>inf</i> % <i>það</i> | <b>18.32%</b>  | <b>25.00%</b> | <b>38.46%</b> | <b>79.70%</b> |
| <i>que</i> – <i>það</i> | 0              | 61            | 157           | 192           |
| <i>que</i> + <i>það</i> | 0              | 30            | 229           | 4             |
| <i>que</i> % <i>það</i> | –              | <b>32.97%</b> | <b>59.33%</b> | <b>2.04%</b>  |

Table 3.44: RMH - Exceptional instances based on clause type

One interesting trend that emerges in these datasets is that *það* is less frequent with infinitives than with *that*-clauses (except in the *utan* dataset). This contrast has also been observed to some extent in Garofalo (2015), where it has been proposed that infinitives display *það* less often compared to *that*-clauses due to the fact that the subject of the infinitive clause is bound to the matrix clause, contributing to a tighter integration between both clauses. Although the data presented above resonate with this hypothesis, it still does not hugely influence the main distribution of NomCIs and non-NomCIs, not to mention that the subset of infinitives in *spá* can hardly be considered significant due to the fact that only 16 valid examples were found (a similar reasoning can also be extended to embedded questions with *spá* since the number of examples is much lower than that of *that*-clauses, which are dominant in the dataset).

One last remark in relation to the *utan* dataset. Considering how different the frequency of *það* in embedded questions is compared to the other subsets, I checked those examples in more detail. It turns out that many of the tokens that do not involve clausal nominalization are embedded questions introduced by an interrogative pronoun *hvað* ‘what’, which has no argument

function and is not even in correlation with adverbs or quantifiers (as we would otherwise expect e.g. with *hvað margir* ‘how many’, *hvað lengi* ‘how long’). Here is an example:

- (41) Allar sýningar NFSu verða í Hólmaröst á Stokkseyri utan [hvað ein sýning  
all shows NFS will be in Hólmaröst in Stokkseyri apart from what one show  
verður sett upp í fangelsinu á Litla-Hrauni]  
becomes set up in prison in Litla-Hraun  
‘All shows of NFS will be in Hólmaröst in Stokkseyri apart from one show being performed  
in the prison in Litla-Hraun.’

Considering the special characteristics of embedded questions in this dataset, I asked some native speakers on the matter and they did not find this construction fully grammatical, independently of clausal nominalization. This means that the general results on *utan* we observed in our analysis of lexical prepositions were biased due to the presence of embedded questions that are considered problematic by speakers independently of *það*.

### 3.3.10 Data on age and gender groups

The data on age and gender groups are illustrated in Appendix B. The results do not mainly affect what we have already seen about the distribution of NomClIs and non-NomClIs. However, there are other types of results that emerged, which I will briefly discuss here. In relation to age groups, there are few cases where specific groups differ from the other two, so it is difficult to identify a pattern there. For instance, the youngest group (21–30) is more accepting of both variants with nominative subjects in Spec,T, but considers NomClIs significantly better than non-NomClIs with double case prepositions assigning dative. The oldest group (61–70) displays no significant statistical difference for our two variants with genitive subjects in Spec,T, but this is probably due to a bias caused by the low number of participants in this age group who judged the relevant sentences (12 vs. 7).

Regarding gender groups, on the other hand, it is worthy of attention that, in various syntactic positions where nominalization is mandatory based on the data we have seen, men are more accepting of both variants, while women require nominalization more often across many syntactic positions. This indicates that women are more conservative than men in relation to the distribution of clausal nominalization and, in particular, in relation to restrictions on non-NomClIs.

### 3.3.11 Data on sources

In relation to the data from RMH, I also verified whether there were any interesting differences between sources when it comes to clausal nominalization. As a general trend, sources tend to be compliant with the trends we have observed across syntactic positions. On the other hand, by looking at each dataset from RMH, there are very sporadic instances of sources that show a frequency of clausal nominalization that differs from the main trend observable in the same dataset. A list of these exceptions can be consulted in Appendix A but they are not discussed

here, as these sources are not quite recurrent. In other words, it was not possible to find any further information from source types on clausal nominalization.

### 3.4 Summary

The data from RMH and the questionnaires displayed a set of recurrent patterns across syntactic positions that cast a better light on the distribution of NomCIs and non-NomCIs. While instances of nominative and accusative case are correlated to a low frequency of NomCIs, although both variants are generally acceptable, dative and genitive (except *langa til*) generally either trigger a high frequency of *það* if not even pronominal obligatoriness. A similar distinction has also been observed with gender and number in post-copular clauses: although both variants are acceptable, NomCIs are less common when the subject of the matrix clause is neuter singular, whereas the frequency of *það* rises as soon as one gender or number feature is non-default.

Another pattern we have observed involves specific argument positions, namely nominative subjects and clausal indirect objects. The data from the questionnaires and corpora suggest that Spec,T and Spec,Appl are two positions where DPs are required, most likely due to the presence of a D-feature in both T and Appl. Moreover, with lexical prepositions as well as nouns and adjectives (except *feginn*), clausal nominalization tends to be needed by lexical items. The data overall indicate that feature checking plays an important role in the distribution of NomCIs and non-NomCIs and actually constitutes the main purpose of clausal nominalization, as summarized in FCH, which is re-illustrated below for simplicity:

(42) *Feature Checking Hypothesis* (FCH)

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and it is mandatory or at least preferable if:

- a. it checks at least one D-,  $\varphi$ - or case feature
- b. a DP complement is required

If conditions (a) and (b) are not met, CN is preferably absent

FCH, however, does not constitute the last step of our analysis. Although we have seen that *það* is preferable or even mandatory when formal features need to be checked or when a DP is required, or less frequent in other cases (provided that the position occupied by the clause is a DP position, as no instance was found in the empirical data where a NomCI is undoubtedly ungrammatical in a DP position), we also need to explore this system further from the point of view of grammar. A high frequency of *það* is, for instance, a possible sign that there are different competing grammars if we follow Kroch (1989) and also Yang (2000), in the sense that one configuration yielding a NomCI is preferable among speakers compared to one that allows non-NomCIs to surface in the same position. The opposite reasoning is applicable when *það* is less frequent. From this perspective, the next step we need to take is to identify the possible derivations that are in competition among speakers and which determine the distribution of NomCIs and non-NomCIs we have seen in this

chapter.

Furthermore, there is still another important issue we need to discuss and which is not yet captured by FCH, namely the fact that passivized accusative clausal objects tend to be more frequently non-NomCIs when the CP emerges as extraposed, despite the fact that a D-feature in T is supposed to be checked by the NomCI subject on the basis of our discussion on nominative clausal subjects. This shares similarities with the examples presented by Thráinsson (1979) with predicates like *vera hörmulegt* (see §2.5.2), where *það* was considered optional. This poses the question whether the examples discussed by Thráinsson and the ones with passivized accusative clausal objects presented in this chapter are derived in a similar way and, if yes, how. We will discuss all these issues in much more detail in Chapter 4.

Lastly, it is worth briefly mentioning some further details on the results that emerged in the current chapter. The data on clausal nominalization based on clause type only showed some sporadic instances where there was a difference in particular between infinitives and *that*-clauses. This means that there is no significant influence of the clause type on clausal nominalization. This does not support analyses like Garofalo (2015), where it was suggested that infinitives are more likely to drop clausal nominalization (see §2.5.6.1). Moreover, the data from RMH categorized by source did not reveal any striking difference in the frequency of *það*. The only additional piece of information that was interesting to find is the comparison between gender groups in the questionnaires. The data revealed in particular that women tend to be more conservative than men in that the former tend to more accepting of NomCIs, especially where *það* is normally required, while the latter are more accepting of both variants. The results on age groups, by contrast, only displayed sporadic instances where a certain group would differ from the others. Hence, no pattern that diverges from the main trend could be observed.

# 4 Theoretical implications

## 4.1 Introduction

In this chapter, we will take a step forward into the analysis of clausal nominalization and discuss from a theoretical perspective the results observed in Chapter 3. So far, we have come to the conclusion that the purpose of *pað* as CN is to check features assigned by other items, whether lexical or functional, as summarized in FCH (re-illustrated below). The relevant features checked by *pað* are case, gender and number features as well as D-features:

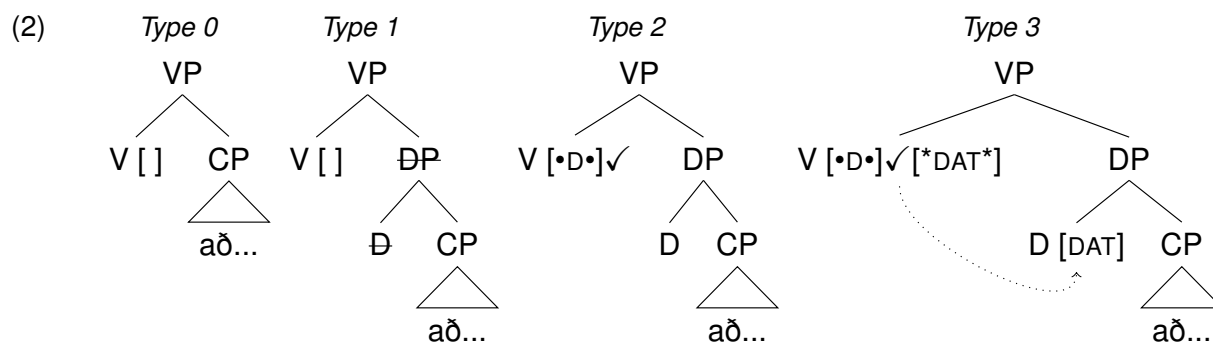
(1) *Feature Checking Hypothesis (FCH)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and it is mandatory or at least preferable if:

- a. it checks at least one D-,  $\varphi$ - or case feature
- b. a DP complement is required

If conditions (a) and (b) are not met, CN is preferably absent

I will argue in this chapter that, from a derivational perspective, the distribution of NomCIs and non-NomCIs is mainly the result of a competition between four different configurations, which are presented below (the tree diagrams display complements of V as an example, but the same mechanism applies to other syntactic positions as well):



The first configuration, *Type 0*, is for non-DP positions (where only CPs can be merged; see the case of *þvinga* ‘force’ and *hjálpa* ‘help’ in Chapter 2), while the other three account for positions where DPs can surface. *Type 3* yields a NomCI due to feature checking as described in condition (a) of FCH (in the tree diagram, a case feature is assigned to the CN). The *Type 2* configuration yields a NomCI as well, but only by virtue of the fact that the item subcategorizing for a DP contains a [•D•] feature (I will follow here the notation proposed by Heck and Müller (2007)), which can only

be checked by a merging category of type D.<sup>1</sup> Regarding *Type 1*, I will argue in this chapter that non-NomCIs that are merged in DP positions are originally merged as DPs. However, when the highest copy of the clausal argument does not have to check any formal feature locally, the DP shell undergoes structural removal, yielding a CP as a result. I will argue for the possible existence of structural removal in Icelandic in §4.3.5.

Furthermore, in §4.3.9, I will also propose that post-copular clauses introduced by *sá* and *sú*, as well as their accusative forms, are exceptionally characterized by a slightly different variant of *Type 3*, in which the CN is in the specifier position of a null D, in line with proposals like Brugé (2002) on demonstratives. As for the other post-copular NomCIs, I still maintain that the CN occupies the D-head of the projection.

In light of my proposal, I will argue in relation to nominative clausal subjects that, while the non-extraposed ones in Spec,T are only accounted for by *Type 3* due to D-feature checking, the extraposition constructions we observed in Chapter 3 (see §3.3.1) are instead the result of three competing configurations, *Type 0*, *Type 1* and *Type 3*, two of which yield a non-NomCI. In *Type 1*, after the clausal DP has been attracted to Spec,T due to EPP, it is moved further by right-adjunction, leaving a DP trace in Spec,T which checks the D-feature in T (cf. Stowell (1981) regarding DP traces left by clausal arguments). Since the highest copy of the NomCI has not directly checked the D-feature locally, the DP shell is removed. As for *Type 0*, the apparent extraposed clause is merged as CP (as already mentioned in §2.5.2), and the expletive occupies Spec,T in order to satisfy EPP. The presence of *Type 0* as a competitor in these constructions is also supported by some observations on stylistic fronting (see §4.3.8).

On the other hand, in relation to case, gender and number, I argue that nominative and accusative case do not correspond to features to check in narrow syntax, whereas dative and genitive do (I made a similar claim in Garofalo (2023)). The data from Chapter 3 suggest it is plausible that nominative and accusative are realized at PF. However, since nominative and accusative differ in Icelandic from a morphological perspective, they also must differ in how they are derived. Such a hypothesis is supported by configurational analyses of structural case, where nominative is the morphological manifestation of unmarked case and accusative is the product of a relationship between DPs respectively, as in *Dependent Case Theory* (see e.g. Marantz (2000) or a preliminary discussion in Yip et al. (1987)). In addition, I believe that the idea of a case that does not correspond to a feature to check in narrow syntax can also be extended beyond canonical argument positions, if we consider, for instance, the contrast between accusative (with which *það* is not quite frequent) and dative (with which *það* is preferable) in prepositional objects. I will propose in this regard that accusative could be a default case in Icelandic prepositional objects, which entails that prepositions that can assign both accusative and dative actually assign dative or no

---

<sup>1</sup>Note that *Type 2* accounts for condition (b) specified in FCH and, therefore, it also extends the idea that the purpose of *það* is to check features to those positions where a DP is required independently of case, gender, number etc., as the item selecting the clausal complement contains a [ $\bullet$ D $\bullet$ ] feature that needs to be checked. This feature differs from the D-feature in T or Appl, as the D-feature in these positions is also an EPP-feature. Further discussion is offered in §4.3.

case feature in narrow syntax. The results from post-copular clauses also indicate that neuter and singular do not correspond to features to check, as they are default gender and number (cf. Garofalo (2023)). This explains why *það* is only dispreferred in clausal post-copular clauses when all of its features are unmarked, i.e. neuter singular nominative.

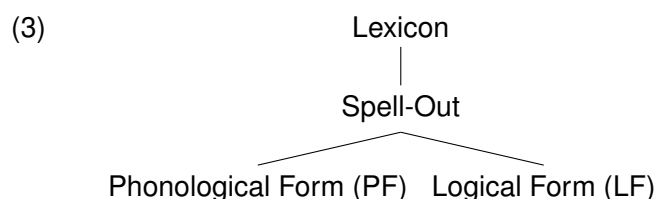
As for the comparison between Icelandic and other languages, I analyzed clausal nominalization in Swedish, German, Persian and Russian (the criteria for this choice are described in §4.5.1). The results of this comparative analysis show that lexical case only triggers nominalization when a language has a morphological case system. By contrast, structural case is not a trigger regardless of whether the language displays a morphological case system or not. In absence of a case system, languages can default to pronominal optionality (as in Swedish) or to obligatory DP selection *in situ* (as in Persian). However, all the languages involved require nominative subjects in Spec,T to be preceded by a CN, similarly to Icelandic, indicating that D-feature checking is responsible for obligatory nominalization in other languages as well. This is also supported by the fact that, if sentential subjects can escape Spec,T, non-NomCl subjects are allowed to emerge.

This chapter is structured as follows. In the first section, I will present the core assumptions necessary to discuss the theoretical issues of this chapter. In the second section, I will delve into the theoretical issues related to the derivation of NomClS and non-NomClS across syntactic positions and argue for the four configurations mentioned in this introduction. In the third, I will focus on the issues related to case, gender and number and what clausal nominalization can tell us about these grammatical categories. In the fourth, I will discuss the comparison between clausal nominalization in Icelandic and the same phenomenon in other languages. In the fifth, the main findings of this chapter will be summarized.

## 4.2 Core assumptions

### 4.2.1 Basic concepts of Minimalism

The theoretical issues discussed in this chapter will be analyzed within a broad Minimalist approach. I assume, first of all, following the standard view of the *Minimalist Program* (see Chomsky (1993; 1995) and much subsequent work), that the derivation of a sentence is the result of different steps and levels of representation. The standard model generally proposed is sketched below:



This model proposes that the building process of a sentence in a language starts with the selection of a set of lexical items from the lexicon (the so-called *lexical array*) which are progressively

merged in the syntactic derivation. All the building blocks are combined bottom-up in order to build the syntactic structure and they are also moved to a different position from the one where they are originally merged if there is any need to, as syntactic movement needs to be motivated, e.g. because a certain grammatical feature needs to be valued (cf. the discussion on the Agree operation below). When all the building and movement operations are completed, the derived structure is *spelled out* and sent to two different levels of representation: Phonological Form (PF), where the output of the syntactic derivation is rendered morpho-phonologically, and Logical Form (LF), where semantic interpretations of the derived structure are applied.

The Minimalist Program aims to reduce the number of syntactic operations to the bare minimum. Three important operations that are relevant in our analysis are Merge, Move and Agree. Merge (with this term I specifically refer to External Merge) combines two different items in the syntactic structure to create a new syntactic object. Consider the following example, where an item of category V (a past participle) and an NP are merged together:

- (4) a. bjargað hundi  
 rescued dog<sub>DAT</sub>  
 'Rescued a dog'
- b.
- 
- ```

graph TD
  VP[VP] --- V[V]
  VP --- NP[NP]
  V --- bjargað[bjargað]
  NP --- hundi[hundi]
  
```

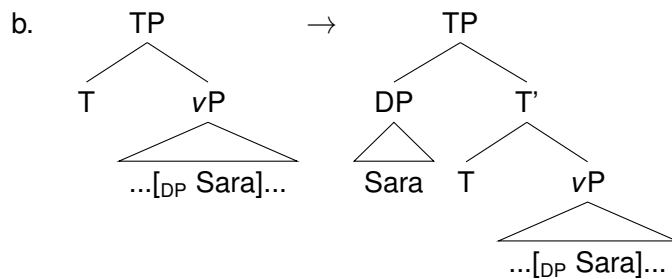
The Merge operation is recursive, which means that the object just created can be one of the two items of a new Merge operation which will produce a more complex item, as shown in the following example, where the VP shown above is embedded within another VP as the complement of the verb *hafa* 'have':

- (5) a. hafa bjargað hundi  
 have rescued dog<sub>DAT</sub>  
 'Have rescued a dog'
- b.
- 
- ```

graph TD
  VP1[VP] --- V1[V]
  VP1 --- VP2[VP]
  V1 --- hafa[hafa]
  VP2 --- V2[V]
  VP2 --- NP[NP]
  V2 --- bjargað[bjargað]
  NP --- hundi[hundi]
  
```

Internal Merge, which is often referred to as Move, occurs when an embedded item is remerged with a higher item in the structure, as shown in the following example of movement of a DP to Spec,T from a lower position (note that the lower copy of the item is generally covert):

- (6) a. Sara sefur  
Sara sleeps  
'Sara is sleeping.'



As for Agree, this operation has been formalized in different ways (consider e.g. Chomsky (1995; Chomsky (2000)); Zeijlstra (2012); Wurmbrand (2012) among others), but, in general terms, it takes place when a head carrying a certain feature checks and marks for deletion a matching feature of another item. Syntactic items carry various types of grammatical features when they enter the derivation, such as semantic, formal features, EPP-features etc. (see Chomsky (1995; 1999)). Some of these features borne by syntactic items are already valued. Consider the example in (4): the noun *hundur* 'dog' is masculine, therefore it bears a gender feature which is already valued from the start, i.e. gender: MASC. Some other features, by contrast, are not valued at the beginning of the derivation but are valued before *Spell-Out*. The case of a noun like *hundur*, for instance, is not pre-determined at the beginning of the derivation. Therefore, the case feature it bears is unvalued. It is usually proposed that a) the item which carries an unvalued feature, the *probe*, checks and marks for deletion the matching valued feature of a different item, the *goal*, and b) the probe c-commands the goal, e.g. when T attracts via Move a DP in a lower position to Spec,T to have its D-feature checked (see Chomsky (1995) among others; also §4.2.2).<sup>2</sup> I assume the following definition of c-command (Chomsky 1995:85):

- (7) *Constituent-command* (c-command)  
X c-commands Y if:  
(i) every Z that dominates X dominates Y and  
(ii) X and Y are disconnected<sup>3</sup>

However, I do not exclude the possibility for the direction of Agree to be reversed as well, in the sense that the goal c-commands the probe (the so-called Upward Agree or Reverse Agree), as proposed by Zeijlstra (2012) and Wurmbrand (2012). I assume this possibility, in particular, in the case of post-copular clauses when the CN checks the gender and number features of the subject.

Let us consider once again the example in (4). The noun *hundur* has a case feature which needs to be valued, i.e. case:  $\emptyset$ . The verb *bjarga* 'rescue', on the other hand, assigns dative case, which means that it bears a valued case feature, i.e. case: DAT. The noun checks the

<sup>2</sup>In §4.4.3, however, I will point out, in relation to case assignment, that the data on clausal nominalization suggest that the item carrying a valued lexical case feature is the trigger of case assignment rather than the probe DP.

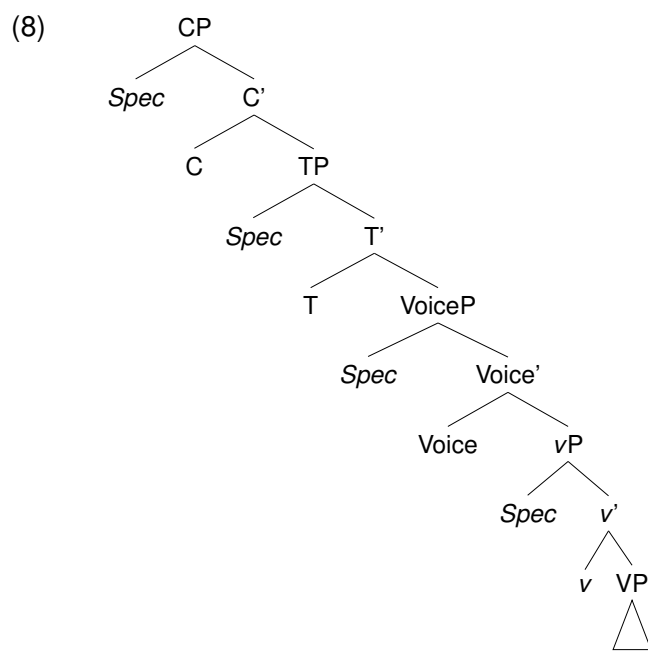
<sup>3</sup>This condition means that X cannot dominate Y nor vice-versa.

case feature of the verb and is assigned the DAT case feature. Subsequently, these features are marked for deletion. For the derivation of a sentence to converge, all the features carried by syntactic items must be valued and marked for deletion. If other features remain unvalued or unchecked at *Spell-Out*, the derivation crashes.

I will assume that Merge is an operation that is triggered by features in a similar way as Agree, in line with analyses like Heck and Müller (2007), which make a distinction between features that trigger an Agree operation and other features, called bullet features, which cause an item to be merged with a category that specifically corresponds to that feature (in particular, I will consider this approach in instances of subcategorization for a NomCl in positions where a NomCl is dispreferred, e.g. accusative direct object position, or where a DP is required). Consider once again the example in (4). Let us assume that *bjarga* carries a bullet feature of type [ $\bullet$ N $\bullet$ ]. This bullet feature triggers Merge of a category that can check the [ $\bullet$ N $\bullet$ ]-feature, that is an NP. An item of a different category, for example a PP, would be incompatible with the feature borne by V and the derivation would therefore crash. I will adopt in my analysis the notation used by Heck and Müller (2007) to distinguish between features triggering Agree, marked with a star [ $\ast$ X $\ast$ ], and the bullet features triggering Merge. More discussion in this regard is offered in §4.3.7.

## 4.2.2 Main clausal structure

Let us now move to the theoretical assumptions I make in relation to the main structure of a clause in Icelandic. These assumptions are important for the discussion of the derivation of clausal subjects and, by extension, of NomCls and non-NomCls in general (see §4.3). The structure I assume is shown in the following tree diagram:



The main point to make here is about the *Tense* head (T). T is the syntactic representation of ver-

bal tense,<sup>4</sup> and has the power of attracting the subject to Spec,T. It is usually assumed that, in the syntactic derivation, the item that is selected to be a subject is merged at a lower position than T. Throughout the linguistic literature, different positions have been proposed as a merge position for nominative subjects, for instance Spec,V (see e.g. Sportiche (1988); McNally (1992); Burton and Grimshaw (1992)), but also Spec,Voice (see Kratzer (1996)) and Spec,v (see Chomsky (1995); note that certain analyses also consider the existence of both VoiceP and vP one over the other, cf. e.g. Sybesma (2021) for Chinese; see the tree diagram in (8)). The hypothesis of the existence of additional functional projections embedding VP originally stems from Marantz's (1981) intuition that external arguments do not have as much impact on the meaning of the main verb as internal arguments do. This indicates that external arguments are merged in a position external to VP. I will assume Spec,v to be the merge position for nominative subjects (following Chomsky's (1995) analysis) although nothing in my analysis rules out Spec,Voice as an alternative.

I also assume that T carries a D-feature, previously defined as an EPP-feature in the literature, that attracts the closest DP subject to Spec,T.<sup>5</sup> EPP refers to the so-called *Extended Projection Principle* (Chomsky (1981) within Government and Binding theory; also Chomsky (1995) and considerable subsequent work within the Minimalist Program), which prescribes that every sentence must have a subject. Due to the presence of a D-feature in T, a DP merged in Spec,v, which is the closest DP c-commanded by T, is promoted to Spec,T to check this feature.

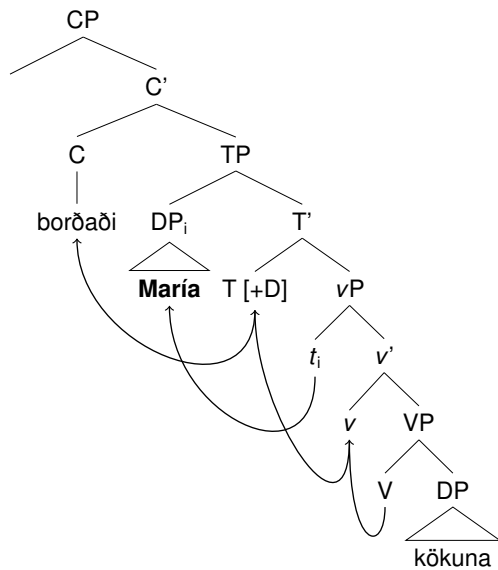
For a better understanding of this mechanism, let us consider some examples of a derivation before we move to next subsection. In particular, let us see how DP subjects in first position and after the finite verb are derived, so that we already have a clear idea about these two different configurations before we discuss in more detail the distribution of NomCl and non-NomCl subjects (see §4.3.2). When a subject emerges after the finite verb, we can assume the derivation to look as follows. I illustrate below an example of subject-verb inversion:

- (9) Borðaði **María** kökuna?  
 ate María cake.the  
 'Did María eat the cake?'

<sup>4</sup>T is also a cover term for other features related to subject and verbal agreement such as person and number (see Sigurðsson (2012)), which can be represented as further projections in the syntactic structure, in case one wants to focus in more detail on them. Moreover, TP has also been considered part of an extended IP structure (see e.g. Pollock (1989)), which contains agreement projections such as AgrSP for subject agreement. Subject-verb agreement is not strictly relevant here, so I will not specifically assume a split-IP structure.

<sup>5</sup>This assumption is valid for those instances where there is no subject gap. By contrast, in instances where we have a subject gap, Icelandic allows stylistic fronting (see e.g. Holmberg (2000); Ott (2009)). This property of Icelandic does not need to be discussed at this juncture as we are dealing the main clausal structure of Icelandic and not specific exceptional cases. However, we will discuss stylistic fronting in §4.3.8.

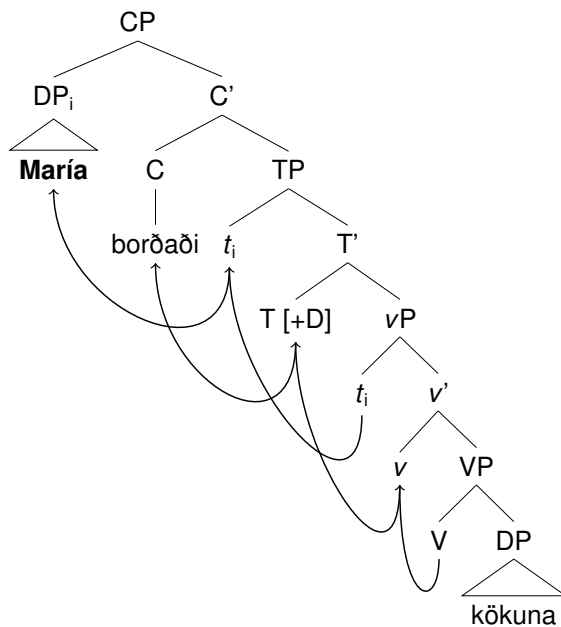
(10)



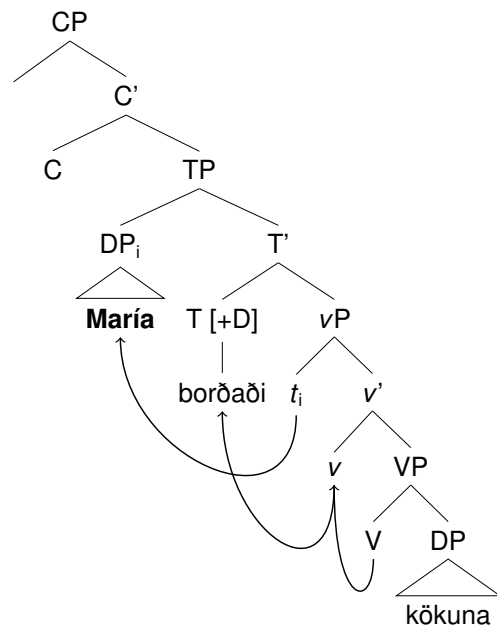
In this construction, the DP subject is moved to Spec,T from its merge position in order to check the D-feature in T, while the finite verb moves from V up to T and then C before the subject. By contrast, when the subject surfaces before the finite verb, that position is ambiguous, as it can be Spec,T once again (while the matrix verb is just moved up to T) or Spec,C if the subject is moved further (while the verb is moved to C; this second configuration would be a string-vacuous movement). I assume, therefore, that two derivations are possible in this case:

- (11) **María** borðaði kökuna  
 María ate cake.the  
 'María ate the cake.'

(12) *Movement of the subject to Spec,C*



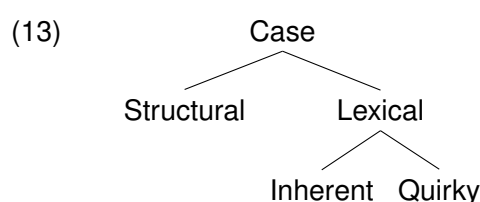
*Movement of the subject to Spec,T*



In the case of sentential subjects, I assume Spec,T and Spec,C are the two possible landing positions when the sentential subject itself is clause-initial.

### 4.2.3 Case

Let us now move to the assumptions I make in my analysis of case. I agree with the common distinction made in the linguistic literature between structural and lexical case (cf. Chomsky (1981) and considerable subsequent work throughout the last decades, e.g. Sigurðsson (1989); Woolford (2006); Yip et al. (1987); Marantz (2000) among many others), which is supported by various linguistic phenomena in Icelandic, as we will discuss below.<sup>6</sup> Moreover, I also assume a further distinction between subcategories of lexical case (cf. e.g. Yip et al. (1987); Jónsson (2003)), which are illustrated in (13). Labels may vary in linguistic accounts (as in Woolford (2006)) and may be more or less inclusive of certain instances of case assignment, but the distinctions made are to a great extent similar:<sup>7</sup>



*Inherent case* refers to those manifestations of case that are not determined by the structure, but are rather predictable on the basis of semantics (see e.g. Yip et al. (1987); Jónsson (2013; 2005b) among many others). On the other hand, those instances of lexical case that are unpredictable have been classified as *quirky case*. As for structural case, the linguistic literature traditionally proposes that it does not rely on a semantic role (see e.g. Butt (2006)) and is assigned by functional heads, like T or C (see e.g. Pesetsky and Torrego (2001) on nominative case), unlike lexical case, where it is, for instance, lexical heads like V that assign case to their complements.

I assume that case assignment is an Agree operation that takes place in the syntactic side of the derivation, i.e. before *Spell-Out*: a DP checks the valued case feature carried by a c-commanded item and the feature is marked for deletion (a similar process takes place for gender and number agreement if we consider post-copular NomCIs; see §4.4.6). As previously mentioned, the derivation crashes if any feature remains unvalued at *Spell-Out*, i.e. not checked and marked for deletion. I assume that this is the mechanism behind lexical case assignment, although in §4.4 I will argue for a different mechanism taking place with structural case.

Icelandic supports the existence of a distinction between structural and lexical case: nominative and accusative are usually identified as structural cases for subjects and objects respectively, whereas dative and genitive are considered lexical cases. This can be observed, for instance, with case preservation in passives, which normally only occurs with lexical case, as we can see in the following examples:<sup>8</sup>

<sup>6</sup>There are, however, other accounts on Icelandic which entirely reject the structural-lexical dichotomy of case and rather consider case as entirely lexical (see Barðdal (2011)). We will see that the data on Icelandic clausal nominalization contradict such a claim, as they clearly display a difference that reflects the structural-lexical case dichotomy.

<sup>7</sup>Although I agree with this distinction, it is a fact that the data from Chapter 3 on the distribution of *það* do not show any clear difference between inherent and quirky case (consider the data on dative and genitive).

<sup>8</sup>It is true, however, that passivization also displays some variation in Icelandic. On the one hand, it has been

- (14) a. Strákarnir brutu **gluggann**  
guys.the broke window.the<sub>ACC</sub>  
'The guys broke the window.'
- b. \***Gluggann** var brotið  
window.the<sub>ACC</sub> was broken  
Intended: 'The window was broken.'
- c. **Glugginn** var brotinn  
window.the<sub>NOM</sub> was broken  
'The window was broken.'
- (15) a. Nefndin hafnaði **tillögunni**  
committee.the rejected proposal.the<sub>DAT</sub>  
'The committee rejected the proposal.'
- b. **Tillögunni** var hafnað  
proposal.the<sub>DAT</sub> was rejected  
'The proposal was rejected.'
- c. \***Tillagan** var höfnuð  
proposal.the<sub>NOM</sub> was rejected  
Intended: 'The proposal was rejected.'
- (16) a. Jón saknaði **Sigríðar**  
Jón missed Sigríður<sub>GEN</sub>  
'Jón missed Sigríður.'
- b. **Sigríðar** var saknað  
Sigríður<sub>GEN</sub> was missed  
'Sigríður was missed.'
- c. \***Sigríður** var söknuð  
Sigríður<sub>NOM</sub> was missed  
Intended: 'Sigríður was missed.'

Dative and genitive objects, in fact, normally cannot take up nominative morphology when they are passivized (traditionally, a way to explain case preservation in passives is that lexical case must be licensed in connection with  $\theta$ -marking, cf. Chomsky (1986); Woolford (2006)). With structural case, by contrast (as in (14)), accusative corresponds to nominative after passivization, as is commonly known in nominative-accusative languages. To be clear, it is not movement itself that causes the DP in accusative case to become nominative. This is supported by examples like

---

observed that there are speakers who let passivized objects display case absorption (as discussed in particular in Benediktsdóttir (2023)), especially in double object constructions. On the other hand, in the so-called new passive, which is emerging in the new generations (cf. Sigurðsson (2011)), accusative direct objects remain low and do not turn into nominative subjects, as would normally happen with lexical case rather than structural case:

- (i) a. %Það var barið **mig**  
it was beaten me<sub>ACC</sub>  
'I was beaten.'
- b. %Það var sagt **mér** ...  
it was said me<sub>DAT</sub>  
'I was told ...'

the following (the passive example is from Jónsson (2005b:390; 2005a:455)):

- (17) a. Tæknimaðurinn keypti **tölvu** fyrir starfsfólkið  
 technician.the bought computer<sub>ACC</sub> for workers.the  
 'The technician bought a computer for the workers.'
- b. Það var keypt **tölva** fyrir starfsfólkið  
 it was bought computer<sub>NOM</sub> for workers.the  
 'A computer was bought for the workers.'

The direct object of *kaupa* 'buy' in the example above can remain in its merge position after passivization. This is due to the fact that indefinite nominals in Icelandic (i.e. NPs) can remain low in a sentence (see e.g. Thráinsson 2007:361ff.), unlike definite nominals (DPs), which are attracted to Spec,T (this also applies to NomCIs; cf. §4.3, in particular §4.3.8).

A further argument in favor of a distinction between structural and lexical case that has been pointed out in the literature on Icelandic is the fact that, in ECM constructions, lexical case must be preserved, whereas nominative case becomes accusative (the examples are from Thráinsson (2007:182–183)):

- (18) a. **Þeir** hafa sofið  
 they<sub>NOM</sub> have slept  
 'They have slept.'
- b. Við teljum [**þá** hafa sofið]  
 we consider them<sub>ACC</sub> have slept  
 'We believe them to have slept.'
- (19) a. **Þeim** hefur leiðst  
 them<sub>DAT</sub> has bored  
 'They have been bored.'
- b. Við teljum [**þeim**/\***þá** hafa leiðst]  
 we consider them<sub>DAT/ACC</sub> have bored  
 'We believe them to have been bored.'

As for the difference between inherent case and quirky case, although no specific test can justify such a distinction, we can still observe that there are semantic patterns that point to the existence of inherent case in Icelandic, for example the fact that dative case is usually assigned to subjects that play the role of experiencers (see also §2.4.4; cf. Thráinsson (2007:200); Blake (2001:68ff.); Þráinsson et al. (2005:322); Jónsson (2003)):

- (20) a. **Mér** leiðist þessi fundur  
 me<sub>DAT</sub> is bored this<sub>NOM</sub> meeting<sub>NOM</sub>  
 'I'm bored of this meeting.'
- b. **Honum** tókst að ná prófinu  
 him<sub>DAT</sub> managed to pass exam.the  
 'He managed to pass the exam.'

- c. **Þeim** er kalt  
 them<sub>DAT</sub> is cold  
 'They are feeling cold.'

Genitive case with subjects and objects, on the other hand, has often been considered to be a quirky case as it is unpredictable (cf. Práinsson et al. (2005:322); Zaenen et al. (1985)), not to mention that it is also quite rare in both subject and object position (as we have seen to some extent in Chapter 3 with subjects; see also Jónsson and Eythórsson (2011)).

Before moving to the next subsection, it is important to mention a final point about indirect objects, which can only emerge with accusative or dative case marking (see §2.3.5) and with which the distinction between lexical and structural case emerges. Although my view is that Appl carries a D-feature as clausal indirect objects must be NomCIs in Spec,Appl (see Chapter 3), I assume that dative is a lexical case assigned by Appl to indirect objects, whereas accusative is structural in Spec,Appl (cf. e.g. Wood (2015b)). This is supported by data on passivization, as dative indirect objects promoted to subjects preserve their case (while the accusative direct object becomes nominative, see (21b)), whereas accusative indirect objects become nominative:

- (21) a. Ég gaf **honum** bókina  
 I gave him<sub>DAT</sub> book.the<sub>ACC</sub>  
 'I gave him the book.'
- b. **Honum**/\***Hann** var gefin bókin  
 him<sub>DAT</sub>/he<sub>NOM</sub> was given book.the<sub>NOM</sub>  
 'He was given the book.'
- (22) a. Lögreglan svipti **hana** ökuskírteininu  
 police.the confiscated her<sub>ACC</sub> driving-license.the<sub>DAT</sub>  
 'The police confiscated her driving license.'
- b. **Hún** var svipt ökuskírteininu  
 she<sub>NOM</sub> was confiscated driving-license.the<sub>DAT</sub>  
 'Her driving license was confiscated.'

This concludes our discussion on the core assumptions upon which the analysis presented in this chapter is built.

## 4.3 Deriving NomCIs and non-NomCIs

### 4.3.1 Introduction

In this section, I will discuss the derivation of NomCIs and non-NomCIs in Icelandic in light of the empirical data presented in Chapters 2 and 3, by especially focusing on clausal subjects, which have been the most mysterious. With the support of the data from subjects and by extending the analysis to other syntactic positions, I will mainly argue in favor of the possibility that all NomCIs and non-NomCIs that are merged in DP positions are originally base-generated as DPs, while the other non-NomCIs that are merged in non-DP positions are base-generated as CPs. Moreover, I

will propose that non-NomCIs originally merged as DPs undergo structural removal if they escape those positions where features need to be checked by *það* locally. I will also present the four configurations that account for the whole distribution of NomCIs and non-NomCIs and, in light of this proposal, refine the hypothesis described in FCH by considering condition (b) as another condition where a feature needs to be checked, namely a [ $\bullet$ D $\bullet$ ] feature that causes Merge of DPs exclusively (see Heck and Müller (2007)).

### 4.3.2 The problem of nominative subjects

We have observed in our analysis so far various instances of clausal nominative subjects that display a different distribution of NomCIs and non-NomCIs. They are re-illustrated below for simplicity:

- (23) a. Er \*(**það**) að læra á hljóðfæri eins og að læra að hjóla?  
is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
'Is learning to play an instrument like learning to ride a bike?' (Garofalo 2023:106)
- b. (**það**) að læra á hljóðfæri er eins og að læra að hjóla  
that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
'Learning to play an instrument is like learning to ride a bike.' (Garofalo 2023:106)
- c. Á árunum fyrir hrun var (**það**) gagnrýnt að bankarnir réðu til sín  
On years.the before crisis was that<sub>NOM</sub> criticized that banks.the hired to themselves  
besta fólkið úr eftirlitinu  
best people.the from supervision.the  
'In the years before the crisis, banks were criticized for hiring the best people from financial supervision.'
- d. Er (**það**) hörmulegt að Jón skuli hafa barið Maríu?  
is that<sub>NOM</sub> deplorable that Jón shall have beaten María  
'Is it deplorable that Jón has beaten María?'  
(adapted from Thráinsson (1979:192, 194))

The first example in (23) is the only one where *það* is mandatory (as the whole clausal subject emerges after the finite verb). In fact, the data from Chapter 3 showed that *það* is acceptable but less frequent in the (b-c) examples, which display respectively an instance where the clausal subject emerges entirely in the first position and one where the clausal object is passivized and the CP appears extraposed. Copula+adjective constructions like the ones presented by Thráinsson (1979) (see the (d) example) align with these results as *það* is optional (and, possibly, less frequent as well if we consider the trend shown in Chapter 3 in structural case positions).<sup>9</sup>

<sup>9</sup>The reader might argue that the (c-d) examples in (23) are of a different nature. As said in §2.5.2, adjectives like *hörmulegt* 'deplorable' do not take a DP as complement and, assuming that the clause is merged as a complement of the adjective (cf. e.g. Roussou (1991)), then one would conclude it must be merged as a CP. However, in the case of passivized constructions, the position where the clause is originally merged should be a DP position if we assume that active and passive voice are derived from one single structure (see Chomsky (1965) and considerable subsequent work). This would suggest that, even if passivized and copula+adjective constructions look very similar at the surface, we should analyze them differently. My view, however, is that these constructions are derived in a similar way, as we will see in §4.3.7 and §4.3.8.

The main question is how we can explain the contrast shown in (23). Given that the position occupied by the subject before the finite verb in a matrix clause can either be Spec,T or Spec,C whereas the position after the finite verb can only be Spec,T, let us return to the variegated distribution we observed in (23). How can we explain from a derivational perspective pronominal obligatoriness in Spec,T when the clausal subject is not extraposed, while both clausal variants are possible in the other examples? On the basis of the analysis in Chapter 2 on the structure of NomCIs and non-NomCIs, we have come to the conclusion that the former emerge as DPs and the latter as CPs. This, however, is the result that we only see at the surface. The question is whether NomCIs and non-NomCIs undergo any structural operations during the derivation that cause the contrast in (23) to be the way it is. We have three possibilities that we need to explore:

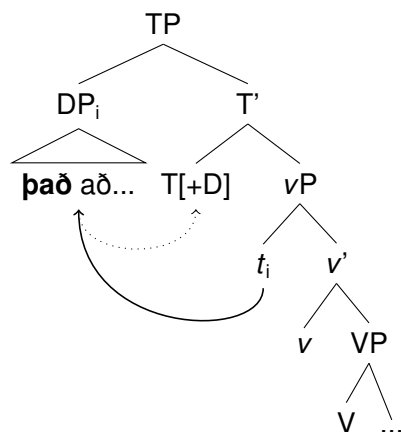
- both clausal types, DPs and CPs, are merged as such and do not change their structure during the derivation (this is our Null hypothesis)
- both clausal types are merged as CPs and nominalized when needed
- both clausal types are merged as DPs and the DP shell is peeled out under certain conditions

Let us start discussing these three options, one by one, by observing clausal subjects in particular.

### 4.3.3 Option 1: DPs and CPs merged as such

Our Null hypothesis consists in the argument that NomCIs and non-NomCIs are merged as DPs and CPs respectively and do not undergo any type of structural transformation. In the case of NomCI subjects, the derivation is pretty simple to reconstruct. Initially, the NomCI is merged as a DP in Spec,v, or even a different position (consider, for instance, passivized clausal direct objects, which are merged as complements of V). Let us now consider the derivation of NomCIs for the cases illustrated in (23), one by one:

(24) *To Spec,T without extraposition:*

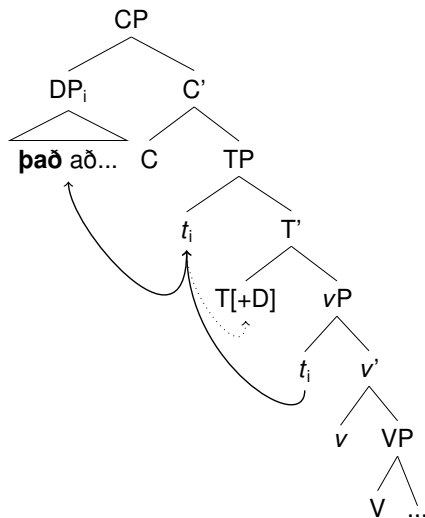


This first scenario is the one of clausal subjects that surface after the finite verb. The NomCI is

promoted from its original position to Spec,T, where it can check the D-feature in T. The clausal subject does not move any further.

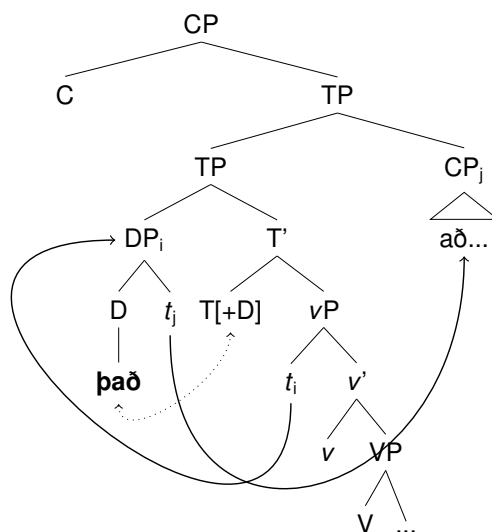
The next case is the one where the NomCl is moved further to Spec,C, and is illustrated in the following tree diagram:

(25) *To Spec,C without extraposition:*



This time, after the NomCl has checked the D-feature in T, it is moved further to Spec,C. The remaining scenario, at this point, is the one of extraposition after the clausal subject has landed in Spec,T, which accounts for the NomCl variant of the (c-d) examples in (23). Assuming that *pað* and the CP are merged as one constituent, one possible way to analyze extraposition constructions with NomCl subjects is to posit that after the NomCl is promoted to Spec,T and *pað* checks the D-feature in T, the CP component of the NomCl is right-adjoined. The tree diagram is sketched below:

(26) *To Spec,T with extraposition:*



I assume here that the CP is right adjoined to TP in line with the *Complement principle* proposed in Rochemont and Culicover (1997) (cf. also Culicover and Rochemont (1990)):

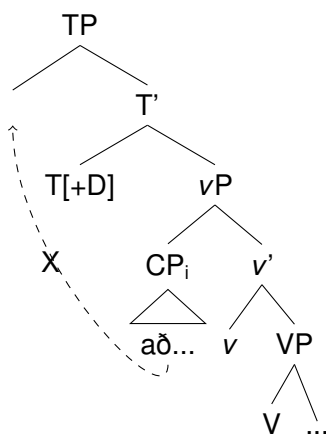
(27) *Complement principle*

An extraposed phrase must be adjoined to the minimal maximal projection that contains its antecedent (Rochemont and Culicover (1997))

Since the CN is in subject position, the extraposed CP is right-adjoined inside the maximum projection that embeds the pronoun, i.e. TP. However, I do not share Culicover and Rochemont's idea that the extraposed phrase is actually already merged as adjoined, due to the fact that their analysis is built upon data from English, where clausal extraposition is mandatory. Since extraposition is not strictly mandatory in Icelandic and we know for a fact that *það* and the CP form one constituent (see Chapter 2), I would rather propose that the CP is moved from inside the NomCl to the right-adjoined position.<sup>10</sup>

As for non-NomClS, the matter is more complex and must be discussed more thoroughly. This time, since a non-NomCl is not a DP, it cannot be attracted to Spec,T in the first place (this is a common cross-linguistic issue, see e.g. Roussou (1991); Alrenga (2005); Takahashi (2010); Hartman (2012) among others), as sketched in the following tree diagram:

(28) *To Spec,T without extraposition:*



This configuration would explain why only NomClS are possible in Spec,T when extraposition does not take place. On a side note, this would also explain why clausal subjects embedded in a subordinate clause can only be NomClS. Consider the example illustrated in (29). The position of the sentential subject in the example cannot be occupied by non-NomClS, which indicates that the position in question is Spec,T. A non-NomCl has no other positions available to emerge as the subordinating conjunction occupies C and, therefore, a) the finite verb cannot move up to C as that position is already taken, and b) the subject cannot move up to Spec,C preceding the conjunction:<sup>11</sup>

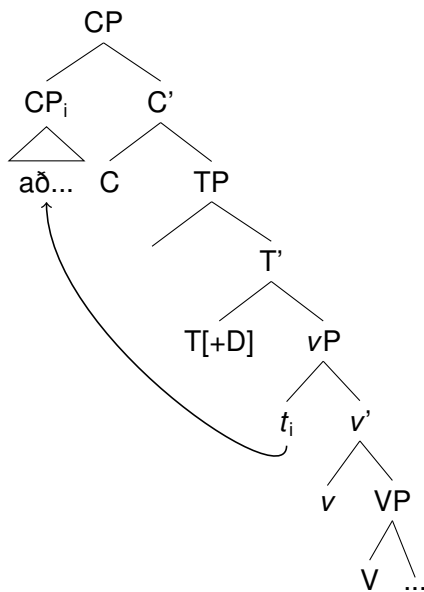
<sup>10</sup>Consequently, if a NomCl subject is moved further to Spec,C and the CP is extraposed, then I also assume that the CP in this case is right-adjoined to matrix CP. This configuration, however, is not strictly relevant here as it is not necessary to explain the contrast in (23).

<sup>11</sup>In §4.3.8, however, we will see that it is possible that other positions are available over Spec,T in subordinate

- (29) Hann sagði [<sub>CP</sub> að [\***það**] að fara á skíði] væri ekki skemmtilegt]  
 he said that that<sub>NOM</sub> to go skiing was not fun  
 'He said that going skiing was not fun.'

As for the next scenario, where non-NomCIs are possible in first position, we are unable to claim, following our Null hypothesis, that the clausal subject crosses Spec,T and then moves further to Spec,C, as movement to Spec,T is not motivated in the first place. An alternative way to derive a construction where the clausal subject surfaces in the first position would be a movement of the non-NomCI straight to Spec,C from its merge position. This movement is in line with Koster (1978), who considered clausal subjects in clause-initial position as syntactic topics:

- (30) *To Spec,C without extraposition:*



This, however, poses a problem for the D-feature, which still needs to be checked. We might try to solve this problem by exploring the possibility that a null expletive checks the D-feature in T, since expletive *það* is generally unable to emerge in Spec,T in Icelandic (see also Thráinsson (2007:312)):

- (31) **Það** rignir á morgun / Rignir (\***það**) á morgun?  
 it rains tomorrow / rains it tomorrow  
 'It is raining tomorrow. / Is it raining tomorrow?'

But this option is problematic as well, at least from a cross-linguistic perspective. As is commonly known, expletives in Mainland Scandinavian languages, unlike expletive *það*, must be phonetically realized after the finite verb. Since the distribution of clausal nominalization with clausal subjects in Icelandic and, for example, Swedish, is similar (see the examples below), we would expect to see an overt expletive after the verb in Swedish if the CP is a topic moved to Spec,C. But this is not the case (the examples are from Josefsson (2006:footnote 12); the last example is my adaptation

---

clauses. This does not affect the validity of the example in (29), as pronominal obligatoriness indicates that the clausal subject is in Spec,T.

with an expletive in Spec,T):<sup>12</sup>

- (32) a. Därför är **(det)** att professorn avgår just nu ganska olyckligt  
therefore is that<sub>NOM</sub> that professor.the resigns just now quite unfortunate
- b. **(Det)** att professorn avgår just nu är därför ganska olyckligt  
that<sub>NOM</sub> that professor.the resigns just now is therefore quite unfortunate  
'Therefore it is unfortunate that the professor is resigning right now.'
- (33) [Att professorn avgår just nu] är **(det)** därför ganska olyckligt  
that professor.the resigns just now is it therefore quite unfortunate  
'Therefore it is unfortunate that the professor is resigning right now.' SWEDISH

If the expletive is not a valid option, then the question remains whether a CP can still somehow cross Spec,T before landing in Spec,C. Even if this possibility existed, the question is how the D-feature would be checked. We might try and follow Shahar (2008) (see our discussion in §2.5.6.2) and posit that the CP subject moves to Spec,T and then onwards, leaving an underspecified DP copy of itself in Spec,T. But it is unclear how this transformation of a CP trace into a DP trace in Spec,T can happen. It is true, though, that other analyses like Koster (1978), Takahashi (2010) and Alrenga (2005) suggest that the trace left by a clausal subject displays DP properties; also Stowell (1981), who proposed, in relation to the *Case Resistance Principle*, that a CP can let its trace check case after  $\bar{A}$ -movement. But if we assume that non-NomCIs are merged as CPs, it is unclear why their traces display DP properties and can check case while the CP itself is unable to emerge in Spec,T. Hence, the derivation of non-NomCI subjects in Spec,C remains mysterious within the Null hypothesis.

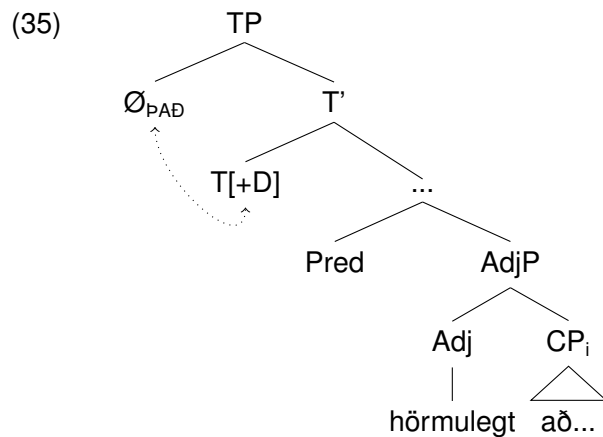
As for the scenario with extraposition, on the other hand, how can we explain the derivation of non-NomCIs if they are merged as CPs? Let us take a look once again at the examples in (23) where the CP appears extraposed, but without *það*:

- (34) a. Er hörmulegt að Jón skuli hafa barið Maríu?  
is deplorable that Jón shall have beaten María  
'Is it deplorable that Jón has beaten María?'  
(adapted from Thráinsson (1979:192, 194))
- b. Á árunum fyrir hrun var gagnrýnt að bankarnir réðu til sín besta  
On years.the before crisis was criticized that banks.the hired to themselves best  
fólkið úr eftirlitinu  
people.the from supervision.the  
'In the years before the crisis, banks were criticized for hiring the best people from financial supervision.'

Let us focus on the first example for the moment. As we said in Chapter 2 (§2.5.2) in relation to copula+adjective constructions, it is possible that the CP is merged as the complement of adjectives like *hörmulegt* 'deplorable' (see also accounts like Roussou (1991)). Within our Null

<sup>12</sup>Note that, exactly as in Icelandic, the pronoun used to introduce clauses in Swedish also plays the role of the expletive (see §4.5.2 for further discussion).

hypothesis, the presence of the CP at the end of sentence can be the result of a simple merge of the CP inside AdjP (note that adjectives like *hörmulegt* do not subcategorize for DPs, therefore we can be confident that the CP is merged as such assuming it is a complement of Adj). In this configuration, since the D-feature in T must be checked, one can suppose a null expletive (symbolized with  $\emptyset_{PAD}$ ) occupies Spec,T. This yields a construction where the CP only appears extraposed at the surface, but it is actually not moved from its position:<sup>13</sup>



This interpretation is supported by analyses like Vangsnes (2002), who proposed that the expletive is not merged in Spec,C but rather in subject position,<sup>14</sup> and can be possibly applied to passivized constructions like (34b) as well. In fact, let us consider the following. The position occupied by the clausal complement of a verb like *gagnrýna* ‘criticize’ is one where DPs can appear, unlike what happens with *hörmulegt*. Nevertheless, also in passivized constructions it is unclear how a non-NomCl can be promoted to Spec,T in the first place if we assume it is merged as a CP from the beginning of the derivation. Therefore, we should assume for the sake of D-feature checking that a covert expletive occupies Spec,T while the CP merged as complement of the matrix verb remains behind.

Such an interpretation, though, is also problematic to some extent. If a covert expletive checks the D-feature in T, it is still puzzling that a non-NomCl can still play the role of the subject, as in the following diagnostic example (adapted from Thráinsson (1979:192, 194)):

- (36) a. **Hvað** er hörmulegt?  
 what is deplorable  
 ‘What is deplorable?’
- b. **Að Jón skuli hafa barið Maríu**  
 that Jón shall have beaten María  
 ‘That Jón has beaten María.’

Moreover, the idea of *það* merged in Spec,T contrasts with other studies like Bowers (2002)

<sup>13</sup>I assume here the analysis of copular constructions presented in accounts like Mikkelsen (2005) and Myler (2018) among many others, where *vP* embeds a *PredP* small clause, in which the specifier position is occupied by the merging subject, and the complement position for the nominal predicate (see also §4.3.9).

<sup>14</sup>Vangsnes identified the subject position as Spec,AgrS following a Split-IP structure.

and Christensen (1991), which rather argue that expletive *það* in Icelandic is merged at a higher position than Spec,T, namely Spec,C, e.g. on the basis of *Transitive Expletive Constructions*. In TECs, while *það* precedes the verb, the subject follows it. This indicates that, while the subject surfaces in Spec,T, the expletive must be higher:

- (37) **Það** hafa **einhverjir nemendur** kvartað um hann  
 it have many students complained about him  
 ‘Many students have complained about him.’

In other words, our Null hypothesis becomes problematic when we consider the derivation of non-NomCl subjects. They remain mysterious as they are not able to emerge in Spec,T although they still appear to be able to play the role of the subject. This means that we need to drop our Null hypothesis and venture the possibility that either NomCls or non-NomCls modify their structure during the derivation.

#### 4.3.4 Option 2: All-CP merge hypothesis

The second option we have to derive NomCls and non-NomCls is the all-CP merge hypothesis. This proposal is not a novelty in the linguistic literature on clausal nominalization, as Hartman (2012), for instance, claimed that clausal arguments are originally merged as CPs, but they can be embedded into a DP projection as a last resort option whenever a DP is required. He observed, in fact, that in languages like Russian, for example, clausal objects are not preceded by a determiner (see also Knyazev (2016) in this regard), while they are when they are passivized and moved to subject position. This happens because only DPs can move to Spec,T (the examples are from Hartman (2012:61–62)):

- (38) a. \*(**To**) čto Daša ušla izvestno vsem  
 that<sub>NOM</sub> that Dasha left known everyone<sub>DAT</sub>  
 ‘That Dasha left is known to everyone.’  
 b. Ja znaju (??**to**) čto Daša ušla  
 I know that<sub>ACC</sub> that Dasha left  
 ‘I know that Dasha left.’ RUSSIAN

The fact that the object clause can only be preceded by the determiner when it is moved to subject position suggests that the only way for a CP to be moved to Spec,T is by being embedded into a DP projection (this has similarities with the process described in Takahashi (2010); see Chapter 2).

Icelandic does not follow the same pattern displayed in Russian, first of all because clausal direct objects can be generally followed by *það*. Consider the following examples already illustrated in Chapter 2:

- (39) a. Ég harma (**það**) að ég sé ekki búinn að senda umsóknina  
 I regret that<sub>ACC</sub> that I am not finished to send application.the  
 ‘I regret the fact that I did not send my application.’

- b. Karl samþykkir (**það**) að einhver annar stjórnir fundinum  
Karl accepts that<sub>ACC</sub> that someone else leads meeting.the  
'Karl accepts the fact that someone else will lead the meeting.'
- c. Umræðan endurspeglar (**það**) að löggin eru ekki almenn lög  
discussion.the reflects that<sub>ACC</sub> that laws.the are not general laws  
'The discussion reflects the fact that the laws are not general laws.'

However, the main question here is whether Hartman's approach can be extended to Icelandic as well. If we consider the examples in (23) in light of an all-CP merge analysis, all the examples of NomCIs are to be interpreted in the following way. In order to reach Spec,T, the merged CP subject is embedded into a DP projection so that it can be promoted to Spec,T and check the D-feature in T. At that point, the clausal subject can a) remain in Spec,T, b) let the CP component be extraposed via right-adjunction to TP, or c) advance entirely to Spec,C. The relevant tree diagrams are basically the same as in our Null hypothesis, with the only difference that the NomCI is originally merged as CP and not as a DP, and the DP shell is added to the embedded CP as a last resort operation.

As for non-NomCI subjects, however, it is still unclear how they can be derived, because, similarly to what we have already discussed in our Null hypothesis, CPs cannot be attracted by the D-feature in T, otherwise we would at least expect them to emerge in Spec,T. In other words, following an all-CP merge approach, we are still unable to explain the derivation of non-NomCIs, as there is no motivation for a CP to be attracted to Spec,T in the first place (which explains why CPs need to be embedded into a DP projection within this approach). But there is also a further problem for this type of model. Consider the case of passivized accusative clausal indirect objects. Recall that accusative *það* is mandatory *in situ* with verbs like *svipta* 'deprive', but if the clausal indirect object is moved to the first position of the sentence after passivization, *það* becomes optional or possibly even dispreferred (assuming that the general results from RMH on nominative case can be extended to passivized accusative clausal indirect objects):

- (40) a. Vopnahléið svipti \*(**það**) að hermenn fórnúðu lífi sínu öllum hetjuljóma  
armistice.the deprived that<sub>ACC</sub> that soldiers sacrificed life their all<sub>DAT</sub> heroism<sub>DAT</sub>  
'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'  
(Garofalo 2023:113)
- b. (**Það**) að hermenn fórnúðu lífi sínu var svipt öllum hetjuljóma  
that<sub>NOM</sub> that soldiers sacrificed life their was deprived all<sub>DAT</sub> heroism<sub>DAT</sub>  
'The fact that soldiers sacrificed their lives was deprived of all heroism.'  
(Garofalo 2023:113)

The fact that *það* is optional in Spec,C indicates that a non-NomCI, i.e. a CP, is able to move from Spec,Appl, cross Spec,T and move further to Spec,C. But how can that even be possible in an all-CP merge analysis if clausal indirect objects must always be DPs in Spec,Appl in the first place in light of the data from Chapter 3? The fact that *það* is mandatory in Spec,Appl would mean that a DP shell is added as a last resort option *in situ* and, consequently, we should expect *það* to be

mandatory in Spec,C after passivization, but this appears to not be the case. In sum, an all-CP merge hypothesis is as problematic for Icelandic as the Null hypothesis described in §4.3.3, due to the fact that it is inadequate to explain the derivation of non-NomCIs.

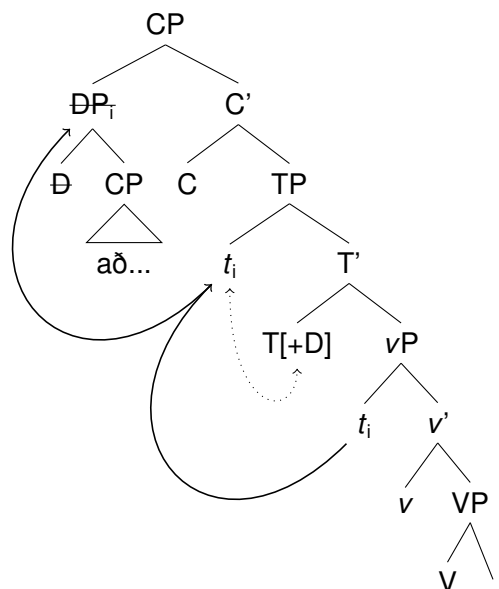
### 4.3.5 Option 3: All-DP merge hypothesis

The third option we are left with at this point is the all-DP merge hypothesis, which basically consists of two main points: a) clausal subjects (and, by extension, clausal complements) are merged as DPs in DP positions, regardless of whether they surface in the end as NomCIs or non-NomCIs, and, consequently, non-DP positions can only host CPs that are merged as such; and b) non-NomCIs undergo a removal operation, where the DP shell is peeled out. To be clear, with DP positions here I am referring to those positions where a DP complement of a lexical/functional item or a DP argument can be merged.

This hypothesis has an advantage over the other two options we have just discussed. If clausal subjects are merged by default as DPs, they can all be attracted to Spec,T to check the D-feature in T, which means that non-NomCIs are also able to cross that position before being moved elsewhere during the derivation. This is a very important point that would solve the issues we have encountered with non-NomCIs in our Null hypothesis and the all-CP merge analysis. An important consequence of this possibility is that structural removal, if it occurs, happens *after* the clausal subject is moved from Spec,T.

In order to understand this better, let us take a closer look at how non-NomCIs are derived, as the derivation of NomCIs is exactly the same as the one presented in our discussion on the Null hypothesis. Consider the following derivation of non-NomCIs that surface in Spec,C:

(41) *To Spec,C without extraposition:*



Within this approach, the non-NomCI is merged as a DP and is attracted to Spec,T by the D-feature in T. This process would now be common to both NomCIs and non-NomCIs so far. Now,

suppose the DP subject moves further to Spec,C before D-feature checking takes place. At this point, it would be the lower trace of the DP subject left at Spec,T that would check the D-feature in T. Since the overt copy of the clausal subject has not checked the D-feature directly but rather the trace left behind in Spec,T, the DP shell in matrix Spec,C is not necessary anymore and can be removed.<sup>15</sup>

There is, however, an important question in relation to the derivation of clausal subjects in (23). If it is true that both NomCIs and non-NomCIs are originally merged as DPs, how do we explain the derivation of clausal subjects in Spec,T with extraposition? Recall the following examples:

- (42) a. Er (**það**) hörmulegt að Jón skuli hafa barið Maríu?  
 is that<sub>NOM</sub> deplorable that Jón shall have beaten María  
 ‘Is it deplorable that Jón has beaten María?’  
 (adapted from Thráinsson (1979:192, 194))
- b. Á árunum fyrir hrun var (**það**) gagnrýnt að bankarnir réðu til sín  
 On years.the before crisis was that<sub>NOM</sub> criticized that banks.the hired to themselves  
 besta fólkið úr eftirlitinu  
 best people.the from supervision.the  
 ‘In the years before the crisis, banks were criticized for hiring the best people from financial supervision.’

Let us look more closely at the example with *vera hörmulegt* ‘be deplorable’. My view is that the optionality of *það* in examples like (42a) is accounted for by three distinct derivations, two of which yield a non-NomCI, while the third one yields a NomCI. The first configuration for non-NomCIs is based on the common assumption that the *that*-clause is merged as a complement of AdjP as already mentioned previously. In this case, the *that*-clause is not merged as a DP, but rather as a CP, due to the fact that the position where it is merged with *hörmulegt* is a non-DP position (see the example below, where *hörmulegt* is compared to other adjectives like *feginn* ‘happy, satisfied’, which can take a DP complement and a NomCI as well; see §3.3.6). In other words, this instance is outside the scope of a generalized DP merge, but it still remains a valid derivation of the non-NomCI variant in (42a). As for D-feature checking, we can suppose it is checked by a covert expletive in this specific configuration:<sup>16</sup>

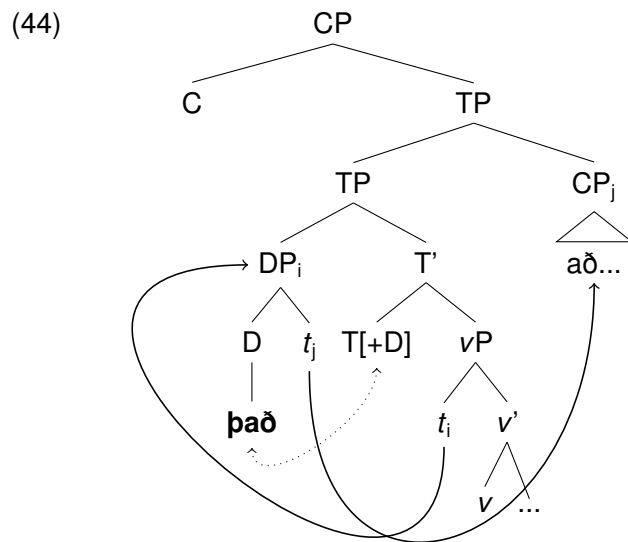
- (43) a. \*Aðstæðurnar voru hörmulegar **þetta/þessu/þessa**  
 circumstances were deplorable this<sub>ACC/DAT/GEN</sub>  
 Literally: ‘The circumstances were deplorable this.’

<sup>15</sup>The reader might wonder whether another option is plausible, namely whether *það* is simply removed from non-NomCIs leaving the DP projection intact (similarly to Knyazev (2016)). I have already discussed this to some extent in Chapter 2, but it is perhaps worth mentioning once again that I do not consider this option viable for Icelandic, especially in light of the fact that the data on extraction from clausal direct objects clearly shows that a DP projection is non-existent in non-NomCIs, at least at the end of the derivation considering the discussion in this chapter. Moreover, since these same clausal objects can be passivized and then display the same distribution of nominative subjects, I would therefore argue that a non-NomCI is more likely to be a product of structural removal rather than the product of a silent DP.

<sup>16</sup>This interpretation would align with (35). However, I do not make any specific claims here about the merging position of the expletive (recall the discussion in §4.3.3 in this regard).

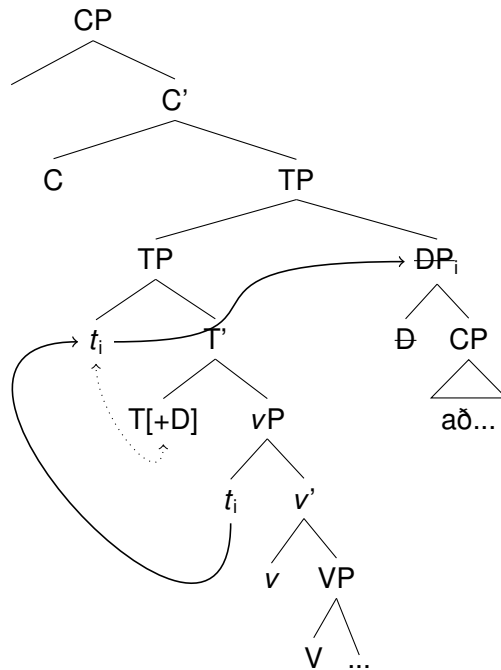
- b. Hann er feginn **þessu**  
 he is happy this<sub>DAT</sub>  
 'He is happy about it.'

The second configuration to discuss is the one that yields a NomCl. Since the *that*-clause cannot be merged inside AdjP as a DP with *hörmulegt*, suppose it is merged in a different position, for example Spec,*v* or Spec,Pred (see Myler (2018) for a discussion on the presence of PredP in copular constructions). In this case, the NomCl would follow the canonical derivation with promotion to Spec,T, D-feature checking and extraposition of the CP, as illustrated in the following tree diagram:



Within an all-DP merge analysis, the existence of a derivation as in (44) where a DP is attracted to Spec,T also suggests that another configuration is possible, which is the third one I am proposing for structures like (42a) and which yields a non-NomCl. Assuming for a moment that nominative clausal subjects are able to remove their DP shell if they escape Spec,T, suppose that a clausal subject is moved from Spec,T to the right-adjointed position before the D-feature is checked. In that case, the trace checks the D-feature in T, while the overt copy of the clausal subject in the right-adjointed position has no feature to check locally. At that point, the DP-shell is removed:

(45)



At the surface, the word order with this configuration will become the same as when a CP is embedded into AdjP, i.e. even if the first and the third configuration start in a different way, the outcome at the surface is the same as when a non-NomCl is produced.

As for passivized clausal objects, as in (42b), the merge position of the clausal object as a complement of V is a DP position (unlike copula+adjective constructions if we assume active and passive voice are derived from the same structure). Therefore, assuming an all-DP merge hypothesis, I believe NomCIs and non-NomCIs in structures like (42b) are explained at least with the following competing configurations based on what we have discussed so far: a) the clause is merged as a DP complement of V and is promoted to Spec,T, and then the CP component is right-adjoined (similarly to (44)); and b) the clause is merged as a DP complement of V and is promoted to Spec,T and, before D-feature checking, the whole DP is moved further to the right-adjoined position, so that the trace left in Spec,T checks the D-feature and the DP shell is removed in the overt copy of the clausal subject due to the fact that it does not have to check any feature locally (similarly to (45)).

However, it is important to inform the reader at this juncture that I will propose in §4.3.8 that an additional third configuration is involved with passivized constructions, namely a derivation where the CP is merged as such. I will discuss this possibility after observing some data on stylistic fronting.

### 4.3.6 Structural removal

If the all-DP merge analysis appears to explain in a better fashion the distribution of NomCIs and non-NomCIs than the Null hypothesis and the all-CP merge analysis, we need to gather further evidence on the validity of this approach. In particular, since the all-DP merge analysis implies

structural removal, we need to gather evidence that structural removal exists in Icelandic syntax in the first place. A first step we can take is to briefly discuss what the cross-linguistic literature offers about structural removal, so that we can have more tools at hand to analyze Icelandic. Mainly, cross-linguistic accounts claim that structural removal occurs in syntax due to two possible triggers: the first is the existence of structure-destroying features, while the second is movement from an embedded clause. Müller (2017; 2018), for instance, claimed that certain features, if checked, can cause the removal of structural material, while other features, on the other hand, trigger Merge of a specific category (see our discussion in §4.2.1; also Heck and Müller (2007)):

(46) *Structural Removal* (Müller (2017; 2018))

As well as structure-building features [ $\bullet X \bullet$ ] that trigger Merge of a category X, there are structure-destroying features [ $-X-$ ] that trigger Removal of an existing projection X in the structure

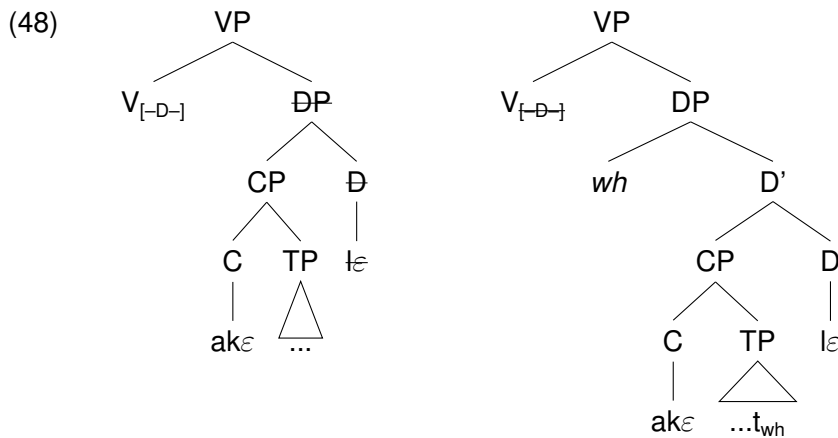
Clausal nominalization in Kwa languages like Gã (spoken in Ghana; see Korsah and Murphy (2019)) confirm this hypothesis. The data show that the determiner of a clausal argument in Gã is regularly removed unless the DP shell contains other items in its specifier position, e.g. traces of items that are extracted from the clausal complement (Korsah and Murphy 2019:255):

(47) a. Kwei yose [<sub>CP</sub> ak $\epsilon$  Y $\epsilon$ mo he y $\epsilon$  $\epsilon$  (\*I $\epsilon$ )]  
 Kwei realize that Y $\epsilon$ mo buy yam D  
 'Kwei realized that Yemo bought yam.'

b. M $\epsilon$ ni<sub>i</sub> ni Kwei yose [<sub>DP</sub> ak $\epsilon$  Y $\epsilon$ mo he    <sub>j</sub> \*(I $\epsilon$ )]?  
 what FOC Kwei realize that Y $\epsilon$ mo buy D  
 'What did Kwei realize that Yemo bought?'

Gã

As we can see from these examples, the determiner  $\epsilon$  is ungrammatical if no extraction applies, but it becomes mandatory as soon as a *wh*-pronoun is moved from the clausal argument itself, which is a diametrically opposite situation compared to the one we found in Icelandic, where *það* is impossible in contexts of extraction (note that Kwa languages like Gã are not island-sensitive like Icelandic, see Korsah and Murphy (2019); cf. Saah and Goodluck (1995)). Here is a sketch of the two different structures in Gã (the tree diagram on the left is the one where no extraction has occurred, while the one on the right involves successive-cyclic movement of an interrogative pronoun):



In the first tree diagram, the [-D-] feature in V destroys the DP shell, leaving the CP behind as a complement of V. In the second one, the DP contains a *wh*-pronoun that is moved successively crossing Spec,D (originally embedded into TP, see the  $t_{wh}$ -trace). This prevents the [-D-] feature from destroying the DP shell.

As for movement as a possible trigger of structural removal, an interesting account is Pesetsky's *Exfoliation Hypothesis* (2016 and subsequent work). Pesetsky's hypothesis abandons the traditional lexicalist idea that clause types like infinitives or *that*-clauses are freely selected by lexical items (see Kiparsky and Kiparsky (1971); Bresnan (1972)) and proposes that the distinction between clausal types is derivationally derived: infinitives are reduced *that*-clauses where the embedded subject is moved beyond the clause boundary, for instance via raising-to-object, raising-to-subject, or  $\bar{A}$ -movement, as illustrated in the following examples (in particular (a-c)). By contrast, if movement beyond the clause boundary is not required (as in the (d-e) examples), the infinitive is ungrammatical (Pesetsky 2021:21; 2016:3):<sup>17</sup>

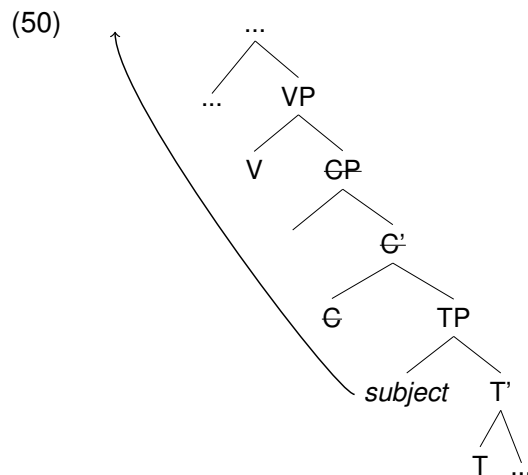
- (49)
- a. Sue considers [**even more important than linguistics**]<sub>i</sub> [ \_\_\_<sub>i</sub> to be the fate of the planet]
  - b. [**Even more important than linguistics**]<sub>i</sub> seems [ \_\_\_<sub>i</sub> to be the fate of the planet]
  - c. [**Even more important than linguistics**]<sub>i</sub> I believe [ \_\_\_<sub>i</sub> to be the fate of the planet]
  - d. \*Mary was assured [**even more important than linguistics**] to be the fate of the planet
  - e. \*It is likely [**even more important than linguistics**] to be the fate of the planet

The following tree diagram illustrates the process of exfoliation (cf. Pesetsky (2016:6)). If we focus, for instance, on raising-to-subject constructions, we can generally assume that e.g. T

<sup>17</sup>The same results occur with fronted predicates as well as fronted locatives (see Pesetsky (2021) for more details). As for control infinitives, like the one below, Pesetsky also suggested that, if control is interpreted as movement as proposed in the *Movement Theory of Control* (see e.g. Hornstein (1999); Hornstein and Polinsky (2010)), the Exfoliation hypothesis would be able to account for all categories of infinitives, as the subject of the embedded clause, even in this case, would be moved beyond the clause boundary:

- (i) Mary persuaded **Sue**<sub>i</sub> [ \_\_\_<sub>i</sub> to talk to me]  
(Pesetsky 2021:113)

in the matrix clause attracts the subject of the embedded clause by virtue of an EPP-feature. However, the embedded subject targeted by matrix T cannot move if it is not at the edge of the embedded CP, due to the so-called *Phase Impenetrability Condition* (see e.g. Chomsky (2008)). Therefore, in order to raise the subject beyond the clause boundary and avoid violating PIC, the CP projection is removed. In other words, exfoliation is a last resort operation to not violate PIC:



Let us now return to Icelandic and the possibility of structural removal. In a similar way to Pesetsky's (2016; 2021) observations, Icelandic also shows that movement from the embedded clause causes structural removal. In support of this, I will show some examples of movement from a clausal object with verbs with which nominalization is mandatory. Consider a verb like *krefjast* 'demand', which requires *það* on the basis of the empirical data presented in Chapter 3. In the following example, the CN *þess* is ungrammatical despite the presence of *krefjast*. This is due to the fact that the pronoun *eiðhvað* 'something' has moved from its previous position inside the embedded clause beyond the clausal boundary (in a similar way as with the examples from extraction). This indicates that the DP shell is not present:

- (51) Þetta er eiðhvað<sub>i</sub> sem ég krefst (\*þess) [að \_\_\_<sub>j</sub> verði gert]  
 this is something that I demand that<sub>GEN</sub> that becomes done  
 'This is something I demand to be done...'

More interestingly, if a DP is extracted from the embedded clause and moved to the first position, nominalization is impossible, but the DP itself takes the case assigned by the matrix verb, as shown in the following examples with an interrogative pronoun:<sup>18</sup>

- (52) a. **Hverju**<sub>i</sub>/**%Hvað**<sub>i</sub> fagnaðir þú (\*því) að hafa fengið \_\_\_<sub>j</sub> ?  
 what<sub>DAT/ACC</sub> rejoiced you that<sub>DAT</sub> to have received  
 'What were you happy of having received?'

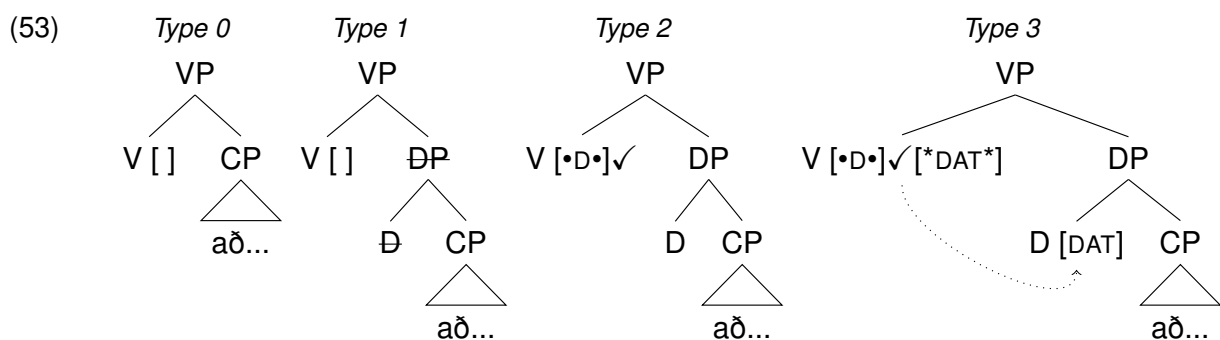
<sup>18</sup>These examples are also going to be discussed in relation to case assignment. For the moment, they confirm that nominalization is excluded and that it is the interrogative pronoun moving via successive-cyclic movement that takes up the case assigned by the matrix verb. Although the majority of the native speakers I asked would only assign lexical case to the interrogative pronoun after movement, there have also been some cases of speakers who would rather keep the accusative form of the pronoun *hvað* 'what' (which should actually be the 'rule' considering the fact that *hvað* is merged as complement of *fá* 'get', which selects an accusative complement).

- b. **Hvers<sub>i</sub>/%Hvað<sub>i</sub>** saknar þú (**\*þess**) að sjá ekki lengur \_\_\_\_ ?  
 what<sub>GEN/ACC</sub> miss you that<sub>GEN</sub> to see not anymore  
 'What do you miss not having seen anymore?'

These facts indicate that the DP projection has been removed in the examples above, and shows that one of the possible triggers of structural removal in Icelandic is movement. However, the problem is whether structural removal can also be triggered when movement does not take place, as non-NomCIs are not necessarily characterized by extraction. Unfortunately, there is no independent evidence in Icelandic that clearly indicates the occurrence of structural removal in absence of movement, as such an operation would only be observable in the domain of clausal complementation. But it is at least a fact that structural removal is an existing operation in Icelandic syntax, and this corroborates the possibility that clausal complements can undergo DP shell removal. Moreover, we also have to keep in mind that the other hypotheses we have discussed, namely the Null hypothesis and all-CP merge analysis, are less adequate than the all-DP merge analysis to interpret clausal nominalization in Icelandic, and since this interpretation is built upon structural removal, we cannot discount it entirely due to the fact that independent evidence cannot be provided.

### 4.3.7 Four possible configurations

Let us now take a step forward in our analysis, assuming the validity of the all-DP merge hypothesis. Considering the fact that this model does not just affect nominative subjects but complements as well (consider, for instance, passivized clausal direct objects), let us now map the distribution of NomCIs and non-NomCIs across syntactic positions on the basis of an all-DP approach as well as by taking into account non-DP positions that can only host non-NomCIs merged as CPs. The whole distribution can be mapped with the following four configurations:



The tree diagrams show an example of a clause that is merged as complement of a verb, but the same mechanism described in the four configurations can be extended to all the other syntactic positions (also those where a clause lands into a specifier position). The first configuration, which I call *Type 0*, covers those positions where clauses can only surface as CPs due to the fact that DPs are not allowed in that same position. This is what we have seen e.g. with the verbs *þvinga* 'force' and *hjálpa* 'help':

- (54) a. \**Það er enginn að þvinga fólk birtingu þessara mynda hérna*  
 there is nobody to force people publication<sub>ACC</sub> these<sub>GEN</sub> photos<sub>GEN</sub> here  
 Literally: 'There is no one forcing people to the publication of these photos here.'
- b. *Það er enginn að þvinga fólk (\*það) að hafa heimasíður barnanna sinna hérna*  
 there is nobody to force people that<sub>ACC</sub> to have homepages children.the their  
 here  
 'There is no one forcing people to have their children's homepages here.'
- (55) a. \**María hjálpaði honum flutningana*  
 María helped him<sub>DAT</sub> moving out<sub>ACC</sub>  
 Intended: 'María helped him moving out.' (Garofalo 2023:107)
- b. *María hjálpaði honum (\*það) að flytja*  
 María helped him that<sub>ACC</sub> to move out  
 'María helped him moving out.' (Garofalo 2023:107)

The other three configurations, on the other hand, cover positions where DPs can surface. In these three configurations, the outcome depends on whether *það* checks features or not. Suppose the D-head has a set of unvalued features:

- (56)  $D \rightarrow [D: \emptyset] [\text{Case}: \emptyset] [\text{Gender}: \emptyset] [\text{Number}: \emptyset]$

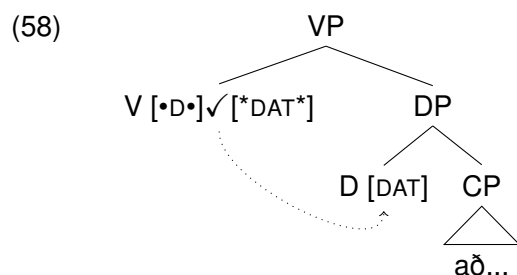
Since *það* is mandatory in positions where D-features must be checked and it is also inflected based on gender, number and case depending on the syntactic position involved (see e.g. the case of postcopular clauses, where *það* agrees with the features of the subject), we can assume the D-head is characterized by all these features that are unvalued at the beginning of the derivation. Keeping this in mind, let us now take a look at *Type 1* in (53). This time, the clause, which is originally merged as a DP, drops its DP-shell as V cannot value any of the features in the D-head. In *Type 2*, by contrast, the DP shell can remain, which begs the question how the clausal complement can remain a NomCl. If we consider positions where a D-feature does not have to be checked, for example direct object position, then the presence of a NomCl must be due to a subcategorization for a DP. In this regard, let us take into account the notation proposed by Heck and Müller (2007) (see §4.2.1), based on which Merge is triggered by structure-building features (bullet features). Let us suppose that complement selection is determined by the presence of a bullet feature borne by the lexical item selecting a NomCl, and let us consider the following example:

- (57)
- 

The bullet feature in the tree diagrams contains a D value, which indicates that the category merging with the preposition P has to be a DP for the bullet feature to be checked. In the tree diagram

on the left, the derivation converges as the complement of P is a DP, while the derivation crashes in the right tree diagram because a PP is incompatible with the bullet [ $\bullet$ D $\bullet$ ]-feature. Assuming that this is a correct analysis of Merge, then *Type 2* would be configured as shown in (53).<sup>19</sup>

The last type in (53), *Type 3*, is the configuration where *það* checks at least one case, gender, number or D-feature. Except the last, which is correlated to EPP, the other ones, called star features if we follow Heck and Müller (2007), trigger Agree. Checking these features prevents the DP shell from being removed, as shown in the following example where dative case is assigned by V to D (consider, for instance, a matrix verb like *fagna* ‘rejoice’):<sup>20</sup>



Provided that this analysis is correct, let us consider for a moment FCH once again, which is re-illustrated below for the ease of the reader:

(59) *Feature Checking Hypothesis (FCH)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and it is mandatory or at least preferable if:

- a. it checks at least one D-,  $\varphi$ - or case feature
- b. a DP complement is required

If conditions (a) and (b) are not met, CN is preferably absent

The four configurations we have seen here indicate that we can refine FCH by considering clausal nominalization as a feature-checking mechanism as a whole (in particular by modifying condition (b) of FCH), due to the fact that the configurations that yield a NomCl, i.e. *Type 2* and *Type 3*, are always characterized by the presence of a feature to check. Therefore, here is a refined (and final) version of my hypothesis, which explains the role and distribution of *það*, which I call *Clausal*

<sup>19</sup>The reader might wonder, considering the configuration in *Type 2*, whether it is possible to affirm that *Type 1* is also triggered by a feature, in particular a structure-destroying feature (as proposed by Müller (2017; 2018)). Although this could make my hypothesis even more generalized for Icelandic in the sense that the purpose of *það*, i.e. checking features, would also be extended to features that destroy the DP shell itself, I believe this might not be the case. If we assume for argument's sake that the structural removal operation is triggered by a feature, then we can only infer that in all positions where the clausal complement emerges as a non-NomCl there is a local structure-destroying feature, e.g. in matrix C (affecting Spec,C) as well as adjoined positions. But this inference is problematic, primarily because it is possible for DPs to be moved to positions like Spec,C, and no DP-shell is removed in those instances. Hence, my view is that it is a much simpler option from a derivational perspective to propose that the lack of a relevant [ $\bullet$ D $\bullet$ ]-feature itself (as well as the lack of any feature *það* can check, for that matter) causes structural removal in clausal complements embedded into a DP projection, rather than the presence of a structure-destroying feature.

<sup>20</sup>I assume here that a [ $\bullet$ D $\bullet$ ] feature could also be present among the features of the item subcategorizing for a clausal complement in order to trigger Merge.

*Nominalization Mechanism:*

(60) *Clausal Nominalization Mechanism (CNM)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and can only surface if (it is enough to satisfy one condition):

- a. CN checks at least one D-,  $\varphi$ - or case feature, or
- b. the clause merges with an item bearing a [ $\bullet$ D $\bullet$ ] feature

In CNM, condition (a) would account for a *Type 3* configuration, while condition (b) would account for a *Type 2*. CNM also implies that the positions where no feature is checked or no DP is possible yield a non-NomCl. This is accounted for by *Type 1* and *Type 0* respectively.

Now, if we consider the syntactic positions we observed in Chapter 3, these four configurations we have discussed are distributed as in Table 4.1:<sup>21</sup>

| Position                                             | <i>Type 0</i> | <i>Type 1</i> | <i>Type 2</i> | <i>Type 3</i> |
|------------------------------------------------------|---------------|---------------|---------------|---------------|
| Non-DP position (see <i>hjálpa/bvinga</i> )          | x✓            |               |               |               |
| Nominative subjects in Spec,C                        |               | x✓            |               | x             |
| Nominative subjects in Spec,T (non-extr.)            |               |               |               | x✓            |
| Passivized nom. subjects in Spec,T (extr.)           |               | x✓            |               | x             |
| Copula+adjective constr. in Spec,T (extr.)           | x✓            | x✓            |               | x             |
| Lexical case subjects                                |               |               |               | x✓            |
| Direct objects with structural case                  |               | x✓            | x             |               |
| Direct objects with lexical case                     |               | (x)           |               | x✓            |
| Indirect objects                                     |               |               |               | x✓            |
| Postcopular clauses in neut.sing.                    |               | x✓            | x             |               |
| Postcopular clauses not in neut.sing.                |               | x             |               | x✓            |
| With functional Ps + accusative case                 |               | x✓            | x             |               |
| With functional Ps + dative or genitive              |               | x             |               | x✓            |
| With lexical Ps                                      |               | (x)           |               | x✓            |
| Complements of nouns                                 |               |               |               | x✓            |
| Complements of adjectives (excluding <i>feginn</i> ) |               |               |               | x✓            |

Table 4.1: Competing configurations across syntactic positions

*Type 3* is the most common configuration as there are often features that *það* has to check, such as lexical case, non-default gender or number or a D-feature. Furthermore, we can notice two recurrent patterns of competition between configurations. The first pair of competitors is *Type 1* and *Type 2*, where the former is more preferable as *það* does not have to check any case, gender or number feature. These two configurations tend to go hand in hand across the board, i.e., if *Type 1* is an option, then also *Type 2* emerges, except with complements of functional prepositions in dative or genitive, where *Type 2* is not possible due to the fact that, if no case feature were checked

<sup>21</sup>Note that the most frequent configuration in a certain position is marked with x✓. Moreover, in direct objects with lexical case and also complements of lexical prepositions, I set the applicability of *Type 1* between brackets due to the fact that the data on *spá* 'predict', *fresta* 'postpone' and also *í kringum* 'around' show that *það* is highly frequent but not strictly mandatory, which might leave some room for *Type 1* to emerge in the grammar of speakers.

with dative or genitive assigning prepositions, we should expect *það* to surface in a default form, i.e. as *það* instead of *því* or *þess*. However, this does not happen. Therefore, we can safely assume that *Type 2* is disallowed in those positions.

The second pair of competing configurations is *Type 1* vs. *Type 3*. In this case, the latter configuration is preferable, except in nominative subjects in Spec,C and extraposed passivized clausal subjects in Spec,T. I believe these exceptions are explained by the fact that Spec,C and adjoined positions can host different types of constituents and not just DPs and CPs, which entails that these positions are not quite characterized by case, gender or number marking. By contrast, in positions where *Type 3* is more preferable like complements of dative- or genitive-assigning functional prepositions, only DPs or CPs can be licensed, which increases the possibilities for case-marking to surface, making *Type 3* more frequent.

As for copula+adjective constructions with extraposition, the table shows three configurations competing (as already discussed in the previous paragraph): *Type 0*, which yields a non-NomCl if the CP is originally merged as such inside AdjP; *Type 1*, which yields a non-NomCl after movement to Spec,T of the clausal subject without *það* checking the D-feature, with consequent right-adjunction and structural removal; *Type 3*, which yields a NomCl. I could not pinpoint which configuration between *Type 0* and *Type 1* is the most frequent, since both yield a non-NomCl, which is more preferable according to the data from RMH. Therefore, I marked them both as most frequent. Nevertheless, interestingly, the fact that two configurations out of three yield a non-NomCl can most probably account for why the non-NomCl variant is more frequent than a NomCl in extraposed copula+adjective constructions.

#### 4.3.8 Stylistic fronting

In our discussion on clausal subjects, we have seen that CPs are unable to emerge in Spec,T due to the fact that the D-feature in T only attracts DPs, as they are the only categories that can check the D-feature itself. However, this raises an important question about another phenomenon that occurs in Icelandic, namely *stylistic fronting* (see Holmberg (2000); Thráinsson (2007); Ott (2009; 2018); Hrafnbjargarson (2004) among many others), which is not characterized by D-feature checking. The main question in this regard is how CP subjects are derived in these kind of instances.

Stylistic fronting is an optional movement that has little repercussion on the meaning of the sentence (possibly a focusing effect; see Hrafnbjargarson (2004)). The main condition for this movement to occur is that there must be a subject gap, i.e. the DP subject is not overt (this also indicates that stylistic fronting is in complementary distribution with expletive *það* as well; see Maling (1990); Ott (2009); Holmberg (2000) among others). Various items are able to be fronted, namely past participles, adjectives or verb particles, but also entire constituents like PPs (see Hrafnbjargarson 2004:98):

- (61) a. Sá sem er **fyrstur** að skora mark fær sérstök verðlaun  
 he that is first to score goal gets special prize  
 'The first one to score a goal will get a special prize.'
- b. Sá sem **fyrstur**<sub>i</sub> er \_\_\_\_<sub>j</sub> að skora mark fær sérstök verðlaun  
 he that first is to score goal gets special prize  
 (Jónsson 1991:1) ADJECTIVE
- (62) a. Hver heldur þú að hafi **stolið** hjólinu?  
 who think you that has stolen bike.the  
 'Who do you think has stolen the bike?'
- b. Hver heldur þú að **stolið**<sub>i</sub> hafi \_\_\_\_<sub>j</sub> hjólinu?  
 who think you that stolen has bike.the  
 (Jónsson 1991:30) PAST PARTICIPLE
- (63) a. Svo leit hann á þá sem höfðu komið **inn**  
 then looked he at those that had come in  
 'Then he looked at those that had come in.'
- b. Svo leit hann á þá sem **inn**<sub>i</sub> höfðu komið \_\_\_\_<sub>j</sub>  
 then looked he at those that in had come  
 (Hrafnbjargarson 2003:88) VERB PARTICLE
- (64) a. Þeir sem hafa verið [<sub>PP</sub> í **Oslo**] segja að ...  
 those who have been in Oslo say that ...  
 'Those who have been in Oslo say that...'
- b. Þeir sem [<sub>PP</sub> í **Oslo**]<sub>i</sub> hafa verið \_\_\_\_<sub>j</sub> segja að ...  
 those who in Oslo have been say that ...  
 (Holmberg 2000:449) PP

The accounts in the literature vary regarding the landing position of the fronted element. Different accounts identify Spec,T as a landing position (see e.g. Holmberg (2000); Ott (2009; 2018) among others), but there are analyses that argue for alternatives, e.g. a functional projection over subject position (as in Bošković (2004)), a FocusP projection (see Hrafnbjargarson (2004)), or even adjunction to Infl/T (see Jónsson (1991)).<sup>22</sup> I assume here the analysis offered in Ott (2018), in which it is pointed out that all instances of stylistic fronting are XP-movements rather than head-movements. On the basis of this assumption, those accounts that propose that the fronted element occupies a head position (e.g. Jónsson (1991), Bošković (2004), Platzack (1987) among others) are problematic, since entire constituents are unable to occupy a head position (as in (64)). Consequently, it is more likely that a specifier position rather than a head is occupied by the fronted element, which suggests that Spec,T might still be taken into account or another specifier position. In this regard, some accounts, in particular Rögnvaldsson and Thráinsson (1990), propose that stylistic fronting might be a form of topicalization, but, as Ott (2018) pointed out, topicalization and stylistic fronting display evident asymmetries, e.g. only the latter is possible in an embedded clause if another item has been extracted. Therefore, they cannot be considered

<sup>22</sup>Further discussion on this topic is offered in Thráinsson (2007:385ff.) and also Ott (2018).

the same type of movement (further arguments are presented in Ott (2018)):

- (65) a. \*Hversu lengi heldur þú [<sub>CP</sub> að [<sub>PP</sub> í **Oslo**]<sub>i</sub> hafi hann búið t<sub>i</sub>]?  
 how long think you that in Oslo has he lived  
 Intended: ‘How long do you think that people have lived in Oslo?’ TOPICALIZATION
- b. Hversu lengi heldur þú [<sub>CP</sub> að [<sub>PP</sub> í **Oslo**]<sub>i</sub> hafi verið búið t<sub>i</sub>]?  
 how long think you that in Oslo has been lived  
 ‘How long do you think that people have lived in Oslo?’ STYL. FRONTING  
 (Ott 2018:5–6)

Considering the fact that Ott (2018) also identified Spec,T as the landing position of fronted elements in Icelandic, let us assume for the moment that such an analysis is correct and let us investigate the possible derivation of clausal subjects with the co-occurrence of stylistic fronting. Consider the following preliminary examples:

- (66) a. Ég veit að **það** hefur verið sagt [að samningurinn sé ólöglegur]  
 I know that it/that? has been said that contract.the is illegal  
 ‘I know it has been said that the contract is illegal.’
- b. Samt hefur (**það**) verið sagt [að samningurinn sé ólöglegur]  
 anyway has that<sub>NOM</sub> been said that contract.the is illegal  
 ‘It has been said anyway that the contract is illegal.’
- c. Ég veit að \*(**það**) rignir oft hér  
 I know that it rains often here  
 ‘I know it often rains here.’

(66a) shows us that *það* is positioned in Spec,T in the subordinate clause and is ambiguous, as it can be interpreted as an expletive or as a CN. This is based on the fact, on the one hand, that the expletive can emerge in that same position with weather verbs, as shown in (66c) (the expletive is mandatory with weather verbs). On the other hand, in line with our observations on clausal subjects, *það* can emerge in root environments, which indicates that a CN can occupy Spec,T (see (66b); recall that an expletive does not emerge in Icelandic after the finite verb).

Since *það* can be an expletive in (66a), this means that there is also a possible subject gap that can be filled with a fronted element. The past participle *sagt* ‘said’ in (66a) can in fact be moved via stylistic fronting:

- (67) Ég veit að **sagt** hefur verið \_\_\_\_ [að samningurinn sé ólöglegur]  
 I know that said has been that contract.the is illegal  
 ‘I know it has been said that the contract is illegal.’

The main question here is how we derive the structure illustrated in (67) – also in comparison with (66a) – in light of our analysis on NomCIs and non-NomCIs so far. As we have already discussed in §4.3.5, passivized constructions with extraposition could be the result of *Type 1* competing against *Type 3* configuration. However, now that stylistic fronting comes into play, I propose it is possible that the structure in (66a) can be produced by *Type 0* and *Type 3*, and that (67) is

a consequence of the fact that *Type 0* is allowed. *Type 1*, on the other hand, is excluded in structures like (66a) due to the fact that the item in Spec,T must be overt inside an embedded clause. However, it remains a valid configuration for non-NomCIs in root environments, as in (66b).

Let us discuss this proposal in more detail. The NomCI variant in (66a) is easily accounted for by *Type 3* with attraction of the clausal subject to Spec,T, D-feature checking and consequent right-adjunction of the CP component. But things become more complex as soon as we consider *það* as expletive, as this would entail a) that the clausal subject is not attracted at all to Spec,T, i.e. it is just a CP rather than a DP (as with copula+adjective constructions), and b) that the position in which the clausal subject is cannot host DPs. This becomes an unexpected problem for CNM, as it apparently calls into question the claim that NomCIs are only allowed in positions where DPs can emerge. We know for a fact that the object position of a transitive verb like *segja* ‘say’ can host DPs:

- (68) a. *Ég segi (það) að samningurinn sé ólöglegur*  
 I say that<sub>ACC</sub> that contract.the is illegal  
 ‘I say that the contract is illegal.’  
 b. *Ég segi söguna*  
 I say story.the<sub>ACC</sub>  
 ‘I tell the story.’

How can we solve this contradiction? My view is that *Type 0* emerges due to a reanalysis of the past participle as an adjective, which would yield a similar structure to copula+adjective constructions, i.e. a past participle like *sagt* can take a DP if it is interpreted as V but not if it is interpreted as Adj.<sup>23</sup> In the latter case, the CP is the only category that can be merged as complement, and is therefore unable to be promoted to subject position. Hence, expletive *það* is added to check the D-feature in T. The data from (67) in which stylistic fronting takes place corroborate this possibility. The fact that *sagt* is moved indicates that no DP has merged in the syntactic structure. This leaves *Type 0* as the only configuration valid with stylistic fronting. But this raises an important question about the D-feature: if it is the past participle that is promoted to Spec,T, what happens to the D-feature, especially considering the fact that such a feature generally prevents any non-DP from occupying Spec,T? There are two possible options we can discuss to answer this question. The first possibility is that stylistic fronting deletes the D-feature. But, assuming for a moment that that is the case, the fact that a low CP cannot be fronted and delete the D-feature as well undermines the validity of this interpretation:

- (69) \**Ég veit að [að samningurinn sé ólöglegur] hefur verið sagt \_\_\_\_*  
 I know that that contract.the is illegal has been said  
 Intended: ‘I know it has been said that the contract is illegal.’

<sup>23</sup>Note that it is commonly known in Icelandic that past participles share many similarities with adjectives as a) they can be inflected like adjectives, and b) can also be found as modifiers of nouns (see e.g. Thráinsson (2007:232ff.)).

The second option to consider is that the position of stylistic fronting is not Spec,T. One piece of evidence in support of this hypothesis is that, in non-root environments, it is possible to front constituents in absence of a subject gap (Práinsson 2005:284):

- (70) a. Ég held [að **forsetanum** hafi María gefið \_\_\_ þessa bók í fyrra]  
 I think that president.the<sub>DAT</sub> has María given this book last year  
 'I think María has given this book to the President.'
- b. Ég held [að **þessa bók** hafi María gefið forsetanum \_\_\_ í fyrra]  
 I think that this book has María given president.the<sub>DAT</sub> last year
- c. Ég held [að **í fyrra** hafi María gefið forsetanum þessa bók \_\_\_ ]  
 I think that last year has María given president.the<sub>DAT</sub> this book

These examples suggest that, since no subject gap is involved and, consequently, the D-feature has been checked, there might be a further position available for fronted elements that is not Spec,T. In the case of stylistic fronting, if we propose that the landing position of the fronted element is above Spec,T, then it is also likely that the D-feature is checked by a covert expletive. The CP, on the other hand, is merged as such and is not promoted to higher positions.

In sum, instances of stylistic fronting can be explained by CNM if we posit that they are the product of a *Type 0* configuration where the embedded clause is merged as a CP, while a covert expletive occupies Spec,T. This, however, also entails that copula+adjective constructions as well as passivized constructions can be derived in a similar way, i.e. with the configurations *Type 0*, *Type 1* and *Type 3*. Let us therefore revisit Table 4.1, as illustrated below:

| Position                                             | <i>Type 0</i> | <i>Type 1</i> | <i>Type 2</i> | <i>Type 3</i> |
|------------------------------------------------------|---------------|---------------|---------------|---------------|
| Non-DP position (see <i>hjálpa/þvinga</i> )          | x✓            |               |               |               |
| Nominative subjects in Spec,C                        |               | x✓            |               | x             |
| Nominative subjects in Spec,T (non-extr.)            |               |               |               | x✓            |
| Passivized nom. subjects in Spec,T (extr.)           | x✓            | x✓            |               | x             |
| Copula+adjective constr. in Spec,T (extr.)           | x✓            | x✓            |               | x             |
| Lexical case subjects                                |               |               |               | x✓            |
| Direct objects with structural case                  |               | x✓            | x             |               |
| Direct objects with lexical case                     |               | (x)           |               | x✓            |
| Indirect objects                                     |               |               |               | x✓            |
| Postcopular clauses in neut.sing.                    |               | x✓            | x             |               |
| Postcopular clauses not in neut.sing.                |               | x             |               | x✓            |
| With functional Ps + accusative case                 |               | x✓            | x             |               |
| With functional Ps + dative or genitive              |               | x             |               | x✓            |
| With lexical Ps                                      |               | (x)           |               | x✓            |
| Complements of nouns                                 |               |               |               | x✓            |
| Complements of adjectives (excluding <i>feGINN</i> ) |               |               |               | x✓            |

Table 4.2: Competing configurations across syntactic positions - revisited

*Type 0* is now added as a possible configuration for passivized constructions on the basis of the data from stylistic fronting.

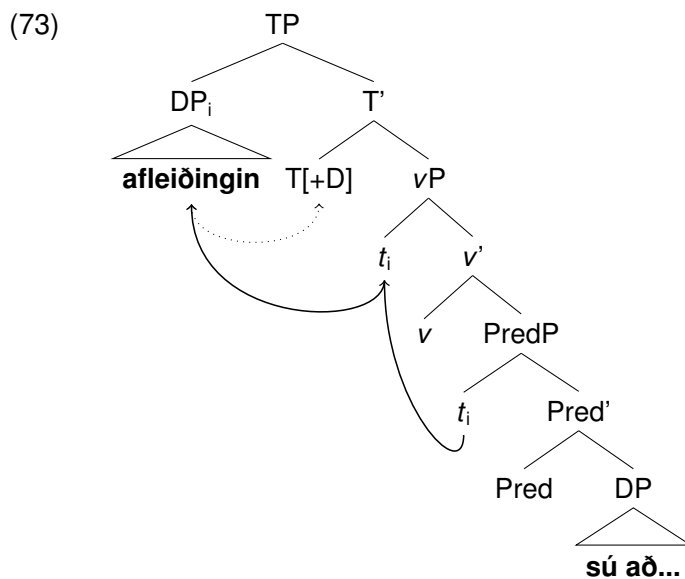
### 4.3.9 Postcopular clauses: mandatory extraposition?

In §2.2.4, I presented an important issue related to postcopular clauses, namely that movement of the whole NomCl to first position is ungrammatical when the determiner emerges (see (71a)), while the sentence is felicitous when the clause appears to be extraposed (see (71b)). When the CN is absent, however, the clause is able to occupy the first position of the sentence (see (71c)); the examples are adapted from Garofalo (2020:9):

- (71) a. \***Sú** að við skuldum meira núna er afleiðingin  
 that<sub>F-SG</sub> that we owe more now is consequence.the<sub>F-SG</sub>  
 b. **Sú** er afleiðingin að við skuldum meira núna  
 that<sub>F-SG</sub> is consequence.the<sub>F-SG</sub> that we owe more now  
 c. **Að** við skuldum meira núna er afleiðingin  
 that we owe more now is consequence.the<sub>F-SG</sub>  
 'The consequence is that we owe more money now.'

The main question here is how we can account for the puzzling contrast between these examples, in particular why (71a) is ungrammatical while (71b) is possible. One important basic point to keep in mind at the beginning of this discussion is that the gender and number of the CN is determined by the gender and number features of the subject, which asymmetrically c-commands the post-copular clause.<sup>24</sup> This is sketched in the following tree diagram, which accounts for the default word order, shown in (72):

- (72) **Afleiðingin** er **sú** að við skuldum meira núna  
 consequence.the<sub>NOM-F-SG</sub> is that<sub>NOM-F-SG</sub> that we owe more now  
 'The consequence is that we owe more money now.' (Garofalo 2020:9)



I assume a PredP-analysis of copular constructions (see e.g. Mikkelsen (2005); Myler (2018) among others), in which vP embeds PredP. The structure in (72) is classified in the literature as

<sup>24</sup>Further discussion on gender and number agreement with post-copular clauses is offered in §4.4.6.

*specificational* (see e.g. Higgins (2015); also Mikkelsen (2005)), in the sense that the post-copular element expresses what the subject is, rather than predicating a property of the subject (as in the so-called *predicational* clauses, cf. Mikkelsen (2005); Bondaruk (2012)). In specificational constructions, both items are merged as DPs: in our example, *afleiðingin* is merged in Spec,Pred and is attracted to Spec,v and then Spec,T to check the D-feature in T. The post-copular item is merged as complement of Pred instead.

I assume that *sú* checks the  $\varphi$ -features of the subject, possibly via reverse Agree (see Wurmbrand (2012); Zeijlstra (2012)), as it is *afleiðingin* that c-commands the CN. If we take into account (71a), then the fact that such an example is ungrammatical can be explained by the fact that the post-copular clause is promoted to a higher position than the subject itself, violating the c-command condition necessary for Agree to take place. This is also supported by the fact that the post-copular NomCl as a whole is only allowed in first position when the CN emerges in its default form, i.e. *það*, as illustrated in the following example:

- (74) **Það** að við skuldum meira núna er afleiðingin  
 that<sub>N-SG</sub> that we owe more now is consequence.the<sub>F-SG</sub>  
 ‘The consequence is that we owe more money now.’ (adapted from Garofalo (2020:9))

However, if c-command is the reason why (71a) is ungrammatical, then (71b) apparently contradicts this hypothesis, as the CN still emerges in a position higher than *afleiðingin*. How can we explain this contradiction? My proposal in this regard is the following: NomCls introduced by *sá* and *sú* and, by extension, their accusative counterparts, have a more complex structure than the other NomCls. This complex structure generates the contrast between (71a) and (71b).

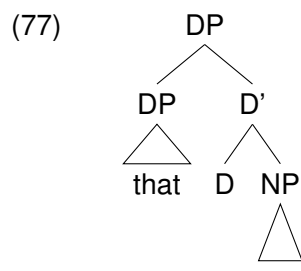
A major piece of evidence in support of this claim is that CNs like *þeir*, *þær*, or *þau* are unable to reproduce the contrast we observed with *sú*, as the structure in (71b) is degraded. *Sá*, by contrast, behaves exactly like *sú*:

- (75) a. **\*Þeir** að kerfið er ódýrt og einfalt í notkun eru kostirnir  
 those<sub>M-PL</sub> that system.the is cheap and easy to use are advantages.the  
 Intended: ‘The fact that the system is cheap and easy-to-use are the advantages.’  
 b. **?\*Þeir** eru kostirnir að kerfið er ódýrt og einfalt í notkun  
 those<sub>M-PL</sub> are advantages.the that system.the is cheap and easy to use  
 Intended: ‘These are the advantages, that the system is cheap and easy-to-use.’
- (76) a. **\*Sá** að bærinn er líka hæðóttur í meira lagi er gallinn  
 that<sub>M-SG</sub> that town.the is also hilly quite a lot is drawback.the  
 Intended: ‘The fact that the town is also very much hilly is the drawback.’  
 b. **Sá** er gallinn að bærinn er líka hæðóttur í meira lagi  
 that<sub>M-SG</sub> is drawback.the that town.the is also hilly quite a lot  
 ‘The drawback is that the town is also very much hilly.’ (adapted from *fararheill.is*)

What is interesting to note here is that those CNs that are homophonous with personal pronouns are unable to yield an extraposition construction after a movement to the first position, while those CNs that only exist as demonstratives in Icelandic behave differently (neuter singular *það* is not

part of this contrast, as it is the default pronoun introducing NomCIs, so we expect it to behave exactly as with nominative subjects, which is proven by (74)). This suggests that the structure of these two types of NomCIs could differ.

In light of these facts, let us consider the following. According to analyses like Brugé (2002) among others, demonstrative pronouns are located in the specifier position of a DP rather than the head, which remains unrealized, as sketched below:

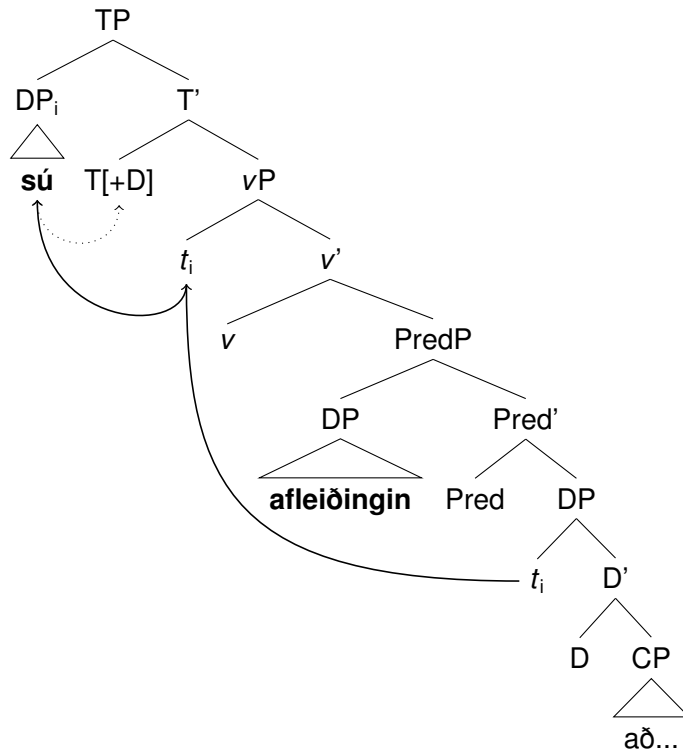


Now, suppose for a moment that *sá* and *sú* are also located in Spec,DP following Brugé's approach, while CNs like *þeir* or *þau*, due to the fact that they are homophonous with personal pronouns (see Chapter 2), are interpreted as D-heads instead. Examples like (75a) as well as (71a) are explained by the fact that the NomCI is higher than the definite noun, so that c-command is unable to trigger Agree. As for the extraposition constructions, the ungrammaticality of (75b) can be explained by the fact that the D-head cannot be promoted to a specifier position leaving the rest of the constituent behind. However, if the CN is in Spec,D, like *sá* or *sú*, then promotion can be performed, as the CN is already in a specifier position. To be clear, my view is that the landing position of *sá* and *sú* after promotion is at least Spec,T, which is proved by the following examples with a yes/no-question, where the same contrast between (71a) and (71b) emerges:

- (78) a. Er **sú** afleiðingin að við skuldum meira núna?  
 is that<sub>F-SG</sub> consequence.the<sub>F-SG</sub> that we owe more now  
 'Is the consequence that we owe more money now?'  
 b. \*Er [**sú** að við skuldum meira núna] afleiðingin?  
 is that<sub>F-SG</sub> that we owe more now consequence.the<sub>F-SG</sub>  
 'Is the consequence that we owe more money now?'

The movement of *sá* or *sú* as CN is sketched in the following tree diagram:

(79)



Let us take a closer look at this proposal of derivation. First of all, the CN is promoted to Spec,T although *afleiðingin* is closer to that position (in violation of the so-called *Minimal Link Condition*, see Chomsky (1993)). Nevertheless, the fact that a lower DP can access Spec,T is not a novelty in the cross-linguistic literature on specificational constructions, as it has been observed that a DP predicate can be raised to subject position instead of the canonical subject (cf. e.g. Moro (1997)). This movement is possible due to the fact that the items revolving around the copula in specificational constructions are both DPs. But one might argue that this movement of the CN to Spec,T still constitutes a problem for agreement, as *sú* emerges in a higher position than the definite noun and one would therefore expect the derivation to crash on the basis of our discussion on c-command. Although it is true that *afleiðingin* cannot c-command the pronoun positioned in Spec,T, it is also true that stand-alone pronouns in Spec,T, as is the case of *sú* if we consider it a specifier of a DP, do not necessarily have to agree with a lower DP in specificational constructions. Consider the following examples, where the neuter pronoun *þetta* ‘this’ can emerge in Spec,T, but not if it is lower than the definite noun *kennarinn* ‘the teacher’. This indicates that, if agreement takes place between the pronoun in Spec,T and the lower DP, it is not necessarily triggered by c-command:

- (80) a. Er **þetta/þessi** kennarinn?  
is this<sub>N-SG/M/F-SG</sub> teacher.the<sub>M-SG</sub>  
‘Is this the teacher?’
- b. Kennarinn er **þessi/\*þetta**  
teacher.the<sub>M-SG</sub> is this<sub>M-SG/N-SG</sub>  
‘The teacher is this one.’

Provided that this analysis is correct, then we can assume that a variant of *Type 3* exists for *sá* and *sú* where the CN is not a head but rather a specifier. The existence of this variant causes *sá* and *sú* to be able to be separated from the rest of the NomCl and be moved to Spec,T, as in (71b), yielding an *only apparent* extraposition construction in which no item is actually right-adjoined. This movement, by contrast, is not possible if the CN is the D-head of the NomCl.

## 4.4 Case, gender and number

### 4.4.1 Introduction

Now that we have discussed the derivation of NomCls and non-NomCls from a structural perspective and proposed a) that they are both merged as DPs (except in non-DP positions, where clauses are merged as CPs), and b) that non-NomCls undergo structural removal during the derivation due to the fact that *það* has no formal features to check, let us focus on the issue of case, gender and number. The main purpose of this subsection is to analyze these three grammatical categories from a derivational perspective in light of our findings so far on clausal nominalization in Icelandic. In other words, we are going to discuss what Icelandic can tell us about the theoretical discussion on case and  $\varphi$ -features. The discussion will primarily focus on case assignment, considering the fact that it is generally involved in all DP positions where NomCls and non-NomCls can surface. Subsequently, we will also discuss gender and number in postcopular clauses.

### 4.4.2 Structural vs. lexical case

A first important implication that derives from the data on clausal nominalization in canonical argument positions is that there is a clear distinction between structural and lexical case (see §4.2.3). We have seen in Chapter 3 that whenever an embedded clause is supposed to receive nominative or accusative case, NomCls are less frequent than non-NomCls, whereas the opposite happens when dative or genitive is assigned to the clausal complement. Although the distribution of NomCls and non-NomCls confirms the traditional model of case marking, it also adds other interesting elements to the discussion on case. Firstly, the data in Chapter 3 suggest that there is not much difference between inherent and quirky case when it comes to clausal nominalization, as NomCls are highly frequent in both cases. Secondly, semantic roles assigned to DPs are normally of various nature, like *Agent*, *Patient*, *Experiencer*, *Instrument*, *Theme*, *Recipient*, *Causer* etc. (see e.g. Práinsson et al. (2005:321); Thráinsson (2007:200); Blake (2001:68ff.)). However, the set of possible semantic roles that a clausal complement can be assigned is more restricted, possibly *Causer* or *Theme* in the case of subjects, and *Theme* in the case of objects. Despite this thorough restriction on semantic roles, the structural-lexical case dichotomy still stands by solely looking at embedded clauses.

### 4.4.3 The trigger of case

An important question in relation to case is what item triggers case assignment. One possibility is that the DP itself is the trigger as it needs to receive (i.e. check and mark for deletion; see §4.2.1) a case feature borne by an assigner (consider accounts like e.g. Knyazev (2016); see also §2.5.5). An alternative possibility is that the item bearing a case feature triggers the mechanism as it needs to assign its case feature to another (c-commanded) item that needs to have its case valued, i.e. a DP. Consider Roussou (1991), for instance, who argued that clausal nominalization takes place due to the fact that a CP by itself is unable to check case features, which implies that the case assigner needs to assign its case feature to a DP. Now, my view about these two possible hypotheses is that the former view can be considered applicable to Icelandic if we only extend our analysis to DPs, but the latter could be a better interpretation when we add clausal complements to the general picture. Let us start taking a closer look at the possibility that it is the DP itself that triggers case assignment. In the literature, the so-called *Case Filter* accounts for this hypothesis:

(81) *Case Filter*

\*NP, if NP has phonetic content and has no Case (Chomsky 1981:49)

The Case Filter indicates that a nominal category cannot be licensed if it is not assigned case, which implies that the mechanism is performed due to the fact that all DPs need case in order to emerge. The Case Filter appears to be valid in Icelandic if we consider examples like the following:

- (82) a. \*María er hrædd **hunda**  
María is scared dogs<sub>ACC</sub>  
Intended: 'María is scared of dogs.'
- b. María er hrædd **við hunda**  
María is scared at dogs<sub>ACC</sub>  
'María is scared of dogs.'

The noun *hunda* 'dogs', by virtue of being a nominal, is supposed to require case, but the adjective *hrædd* 'scared' is unable to do so. Therefore, the derivation crashes. It can only converge if we add the preposition *við* 'at', which is selected by the adjective and which is able to assign case to *hunda*.<sup>25</sup>

Now, things become more problematic for the idea that DPs are the trigger of case when we take a look at lexical case. We have seen that verbs like *fagna* 'rejoice' and *sakna* 'miss' require *það* before their clausal complements, as illustrated in the following examples (note that I will discuss structural case in the following paragraph):

<sup>25</sup>Note that the outcome will be the same even if the noun is assigned dative or genitive (*hundum* and *hunda* respectively).

- (83) a. **Ég fagnaði \*(því) að hafa fengið styrkinn**  
 I rejoiced that<sub>DAT</sub> to have received grant.the<sub>ACC</sub>  
 'I was happy to receive the grant.'
- b. **Ég sakna \*(þess) að sjá Suðurströndina**  
 I miss that<sub>GEN</sub> to see south.coast.the<sub>ACC</sub>  
 'I miss having seen the South Coast.'

An important structural implication of these examples is that, if *það* is mandatory, the clausal complement can only be an extraction island. However, we have already seen in our discussion on structural removal that extraction can violate pronominal obligatoriness. Interestingly, extraction takes place in a way that indicates that dative and genitive case assignment in the examples above is triggered by V and is not required by the clausal argument itself. Consider, for instance, wh-movement. Let us transform the direct objects embedded into the clausal arguments, i.e. *styrkinn* 'the grant' and *Suðurströndina* 'the South Coast', into interrogative pronouns and move them to matrix Spec,C. Unexpectedly, they tend to emerge in dative and genitive case respectively (i.e. the case assigned by the matrix verb) and not in accusative case (at least according to the majority of native speakers I asked in this regard).<sup>26</sup> At the same time, *það* is disallowed:<sup>27</sup>

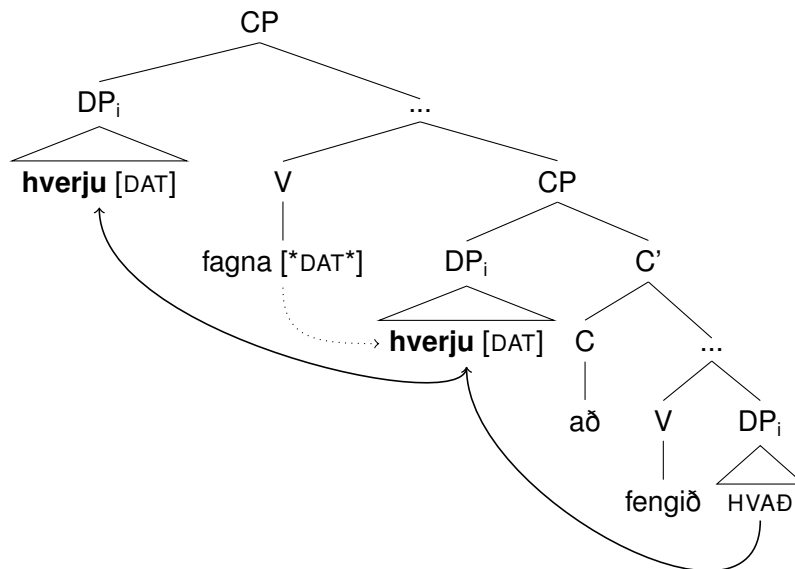
- (84) a. **Hverju<sub>i</sub>/%Hvað<sub>i</sub> fagnaðir þú \*(því) að hafa fengið \_\_\_\_<sub>i</sub> ?**  
 what<sub>DAT/ACC</sub> rejoiced you that<sub>DAT</sub> to have received  
 'What were you happy of having received?'
- b. **Styrkinn**  
 the.grant<sub>ACC</sub>  
 'The grant.'
- (85) a. **Hvers<sub>i</sub>/%Hvað<sub>i</sub> saknar þú \*(þess) að sjá \_\_\_\_<sub>i</sub> ?**  
 what<sub>GEN/ACC</sub> miss you that<sub>GEN</sub> to see  
 'What do you miss having seen?'
- b. **Suðurströndina**  
 south.coast.the<sub>ACC</sub>  
 'The South Coast.'

Because of extraction, we can safely assume that the clausal complement surfaces as a CP, which is unable to be assigned case based on *Case Resistance Principle* (Stowell (1981); cf. Roussou (1991)). The extracted element, at this point, takes up the case that the clausal complement itself cannot be assigned (the same, however, does not occur when an elliptical answer is given to the questions in (84) and (85), as one can see in the (b) examples; this suggests that the DP in the elliptical answer has not moved from its position). This is exemplified in a clearer fashion in the tree diagram below:

<sup>26</sup>This issue is not a novelty in the cross-linguistic literature. A recent account can be found in Abramovitz (2020), who observed that, in Koryak (a Chukotko-Kamchatkan language), ergative overwrites the case of absolutive subjects moving successive-cyclically, and dative overwrites the case of absolutive objects.

<sup>27</sup>see Footnote 18 in this chapter. Also note that the elliptical answer, by contrast, is still grammatical in accusative case for all the speakers I contacted.

(86)



The interrogative pronoun is merged as the complement of the verb *fá* 'get' in past participle. The verb *fagna* needs to assign its dative case and, since the complement of *fagna* is a CP due to extraction and consequent structural removal, case cannot be assigned. The crash in the derivation is prevented by the movement of the extracted item, which is promoted to the edge of subordinate Spec,C, where it can be assigned dative case, and it is promoted further to matrix Spec,C to check the [+wh] feature in C.

The idea that DPs are the trigger of case assignment is problematic with such a configuration. If it were correct to assume that DPs are the triggers of case assignment, then we should expect them to display accusative case marking in the examples above, as they are merged in a position where they would already be assigned structural case. The fact that they emerge in dative or genitive after movement indicates that it is a case assigner that triggers case (in our examples, the matrix verbs *fagna* and *sakna*). This, at least, appears to be correct with instances of lexical case. We will see in the next subsection that structural case displays a different configuration.

#### 4.4.4 Nominative and accusative are not syntactic case features

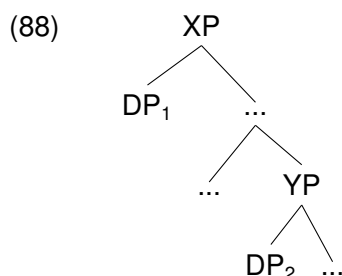
As we have just seen in relation to dative and genitive case, we can assume that they are assigned by the relevant lexical item to a DP. However, the question is how nominative and accusative are configured in this regard, especially considering the fact that the data on clausal nominalization in Icelandic indicate a clear distinction between structural case and lexical case, where the former is correlated to a much lower frequency of NomCIs. To answer this question, I believe it is important to find out at what level of representation nominative and accusative are assigned. Let us discuss this specific issue in more detail. Traditional case theory entails that both structural and lexical case are features checked in narrow syntax. Take, for instance, the case of Modern Greek I have already described in §2.5.5 in relation to the *Case-marking Hypothesis*. Roussou (1991) claimed that NomCI subjects in Modern Greek surface obligatorily due to the fact that nominative case

(assigned by T) must be assigned, as *Case Resistance Principle* (Stowell (1981)) entails that a CP is unable to check case by itself. However, such a theory is inadequate to analyze the results of clausal nominalization in Icelandic, first of all because structural case clearly differs from lexical case on the basis of the results in Chapter 3. Secondly, let us recall that, with nominative subjects, *það* is preferably dropped if the clausal subject is moved to Spec,C, but not if it is in Spec,T. The approach adopted by Roussou, if applied to Icelandic, cannot predict the existence of examples of nominative subjects where *það* can be dropped:

- (87) a. (**það**) að læra á hljóðfæri er eins og að læra að hjóla  
 that<sub>NOM</sub> to learn to instrument is same as to learn to ride a bike  
 'Learning to play an instrument is like learning to ride a bike.' (Garofalo 2023:106)
- b. Er \*(**það**) að læra á hljóðfæri eins og að læra að hjóla?  
 is that<sub>NOM</sub> to learn to instrument same as to learn to ride a bike  
 'Is learning to play an instrument like learning to ride a bike?' (Garofalo 2023:106)

If the traditional theory of case was valid for Icelandic, we should at least expect *það* to be mandatory with nominative clausal subjects independently of its position. The same objection can be raised, for example, for accusative clausal direct objects in Icelandic, as they are never obligatorily nominalized.

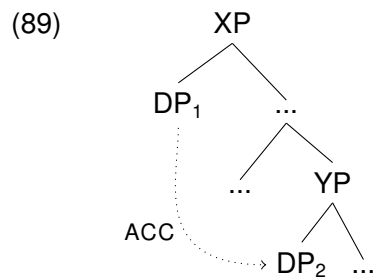
At this point, the data from clausal nominalization suggest that only a theory of case that makes a distinction between how structural case and lexical case are assigned can account for the distribution of NomCIs and non-NomCIs. In this regard, one possible theory that can be applied to Icelandic is the so called *Dependent Case Theory* (see e.g. a preliminary discussion in Yip et al. (1987); also Marantz (2000), Preminger (2011), Baker (2015), Wood (2017a) among many others), which redefines the concept of structural case and views case assignment from a configurational perspective. In many accounts in favor of DCT, lexical case still remains the product of feature checking just as in the traditional theory of case.<sup>28</sup> However, nominative and accusative case are assigned differently. Accusative, in particular, is the product of a relationship between nominals in the same phase (cf. Baker (2014)). Let us consider the following structure to understand this better:



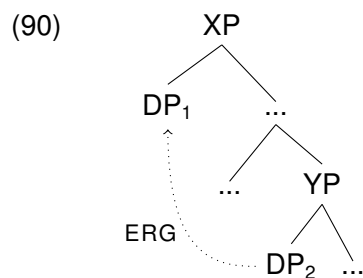
Let us suppose that DP<sub>1</sub> and DP<sub>2</sub> are nominals that have not been assigned lexical case by any

<sup>28</sup>There are, however, some accounts that consider all instances of morphologically marked case as the result of a relationship between DPs rather than between a head and a DP (e.g. Levin and Preminger (2015)), but the idea that there is a distinction between lexical case and dependent/unmarked case in DCT appears to be the most common view.

lexical item during the derivation and are yet to receive case, and let us also assume that they are inside the same phase, e.g. a CP. DCT posits that one of the two DPs is assigned a case that is dependent on the relationship itself between these two DPs, i.e. *dependent case*. The relationship that triggers dependent case is c-command, i.e. DP<sub>1</sub> c-commands DP<sub>2</sub> within the same phase. In nominative-accusative languages, the dependent case is accusative, which is assigned to the lower DP (see in particular Baker (2015:48-49) for the underlying conditions of dependent case assignment):



DCT is also applicable to ergative-absolutive languages like Hindi or Basque, where it is the higher DP that is assigned dependent case, i.e. ergative:



Those DPs that remain after dependent case is assigned take up nominative morphology (or absolutive in the case of ergative languages). To put it clearly, DCT proposes that case assignment functions according to the following hierarchy (see e.g. Marantz (2000)):

- (91) lexical case > dependent case > unmarked case > default case

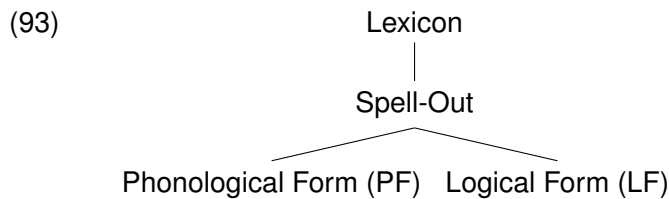
The first DPs that receive case are the ones that are assigned lexical case by lexical items like V. After this operation, if two DPs in the same domain are not assigned case yet, since DP<sub>2</sub> is c-commanded by DP<sub>1</sub>, DP<sub>2</sub> takes dependent case in nominative-accusative languages, which manifests itself via accusative case morphology after *Spell-Out*. Note that, if the condition that two DPs are yet to be assigned case in the same spell-out domain is not satisfied, dependent case is not assigned. Consider the case of nominative objects:

- (92) **Mér** leiðist **þessi fundur**  
 me<sub>DAT</sub> is bored this<sub>NOM</sub> meeting<sub>NOM</sub>  
 'I'm bored of this meeting.'

In this example, the subject is already assigned lexical case, which means that the relationship between DPs cannot be established for dependent case to be assigned. The object, therefore,

surfaces in nominative case.<sup>29</sup>

After the relevant DPs are assigned dependent case (if any), if there are any remaining DPs that still have to be assigned case, they get unmarked case after *Spell-Out*, which is morphologically expressed as nominative. A default case can differ from unmarked case, as it emerges in instances where the relevant DP is in a fragmented sentence, e.g. the pronoun *me* instead of *I* in *me too* in English (see Marantz (2000)). In other words, within DCT, nominative and accusative do not correspond to case features that need to be checked in narrow syntax, unlike lexical case assignment. Recall the Minimalist Y-model:



In DCT, lexical case assignment remains an operation performed before *Spell-Out*, while nominative and accusative are manifested at PF (cf. e.g. Marantz (2000); Baker (2015)). If it is correct to affirm that nominative and accusative do not correspond to features to check in narrow syntax, one might ask why there is still a distinction between the two. The distinction is due to the fact that both are the result of different processes: accusative case in canonical argument positions is the result of a relation between DPs in the same phase, while nominative is the morphological manifestation of unmarked case, which does not depend on a relation between DPs.

<sup>29</sup>It is true, however, that certain Icelandic verbs like *vanta* 'lack' are problematic for case theory, due to the fact that both the subject and the object are assigned accusative (in Icelandic, transitive verbs with an accusative subject generally take an accusative object as well, not to mention that *vanta* also displays some variation with the case of the subject, which can also be dative; cf. Thráinsson (2007:224)):

- (i) **Mig** vantar **penna**  
 me<sub>ACC</sub> lacks pen<sub>ACC</sub>  
 'I need a pen.'

Wood (2017b) proposed, following DCT, that both DPs are assigned dependent case, due to the presence of a covert clitic which is originally merged higher than the other two DPs. His proposal is based on similarities between verbs like *vanta* and existential accusatives, but there is no independent evidence in Wood (2017b) in support of the existence of a higher clitic. As for what clausal nominalization can tell us about this type of verb and the nature of double accusative assignment, I believe the object of *vanta* is assigned dependent case (see the example below), as *það* is not mandatory, in line with our analysis on accusative direct objects. Such a view would be compatible with Wood's insights. The subject, however, cannot be tested as its role is the one of an experiencer, which is semantically incompatible with clausal arguments:

- (ii) Mig vantar (**það**) að lesa sjötta kaflann og svo er ég búinn  
 me<sub>ACC</sub> lacks that<sub>ACC</sub> to read sixth chapter.the and so am I finished  
 'I have the sixth chapter to read and then I'm done.'

#### 4.4.5 The problem of prepositional complements

Within an approach like DCT, we can make more sense of the distribution of NomCIs and non-NomCIs as there is a clear distinction between lexical case and structural case. It is true, however, that DCT generally analyzes case distribution in canonical argument positions. Therefore, the question is how we can interpret case in the distribution of NomCIs and non-NomCIs in other syntactic positions. The data from Chapter 3 on NomCIs suggest in this regard that accusative case is more likely to be manifested at PF in PPs rather than to be the product of case feature checking, as *pað* is generally not preferable with accusative PP complements in a similar way to accusative direct objects, whereas dative and genitive prepositional complements tend to display a high frequency of *pað* – or at least a higher frequency than NomCIs in accusative case – similarly to dative in genitive in direct object position.

The idea of a *default* case in PPs is not a novelty in the cross-linguistic literature, as Van Riemsdijk (2007) proposed, for example, that dative is the default case in prepositional complements in German. German provides a morphological case system with four cases like Icelandic (further discussion about German is offered in §4.5.3) and displays other similarities with it like the presence of double case prepositions (which assign accusative and dative). More specifically, Van Riemsdijk proposed that dative case in prepositional complements is a default case in German, whereas accusative is assigned to phrases where a route component is involved, especially those that include a goal. Consider the following examples from Van Riemsdijk (2007:8):

- (94) a. Er hat [**die** halbe/ganze Nacht] geschlafen  
he has the<sub>ACC</sub> half/whole night slept  
'He has slept for half of the night / the entire night.'
- b. Er ist [**den** halben/ganzen Berg hinauf] gerannt  
he has the<sub>ACC</sub> half/whole<sub>ACC</sub> mountain up run  
'He covered half the distance up the mountain / the whole distance running.'
- c. Er ist [auf **den** Berg halb/ganz hinauf] gegangen  
he has onto the<sub>ACC</sub> mountain wholly up gone  
'He has gone halfway / completely up onto the mountain.'
- GERMAN

The examples show that, in those phrases where a (space or time) distance is covered, accusative appears to be common. This also applies to PPs like (94c), which contains a goal as well as a distance covered (in other words, a route). This view is also supported by the fact that there are route-expressing prepositions like *durch* 'through' or *entlang* 'along' that assign accusative case. By contrast, dative is not related to a specific meaning and appears to be a default option whenever neither goal nor route is expressed. The fact that dative is a default case is also supported by the fact that it can surface in appositions that should be assigned genitive or accusative, as illustrated in the following examples (Van Riemsdijk 2007:13–14):

- (95) a. Sie war im Besitz **zweier** Kleidungsstücke der Ermordeten,  
 she was in possession two<sub>GEN</sub> clothing.pieces<sub>GEN</sub> the<sub>GEN</sub> murdered-woman<sub>GEN</sub>  
**einem** Persianermantel und **einem** roten Kimono ...  
 a<sub>DAT</sub> fur coat and a<sub>DAT</sub> red kimono  
 'She owned two pieces of clothing of the murdered woman, a fur coat and a red kimono...'
- b. Der König kam ohne **Krone** und **Zepter**, **den** wichtigsten Symbolen  
 the king came without crown<sub>ACC</sub> and scepter<sub>ACC</sub> the<sub>DAT</sub> most important symbols  
 seiner Macht und Würde  
 his power and dignity  
 'The king came without crown and scepter, the most important symbols of his power and dignity.'
- GERMAN

As we can see in the first example, the number *zweier* 'two' is in genitive case, but the apposition can emerge in dative case (genitive is, of course, also possible; see Van Riemsdijk (2007)), whereas the apposition in the second example can surface in dative even if the correlated nouns *Krone* 'crown' and *Zepter* 'scepter' are in accusative case (*ohne* 'without' assigns accusative in German).

Now, the question is whether Icelandic also displays a default case in PPs and whether there are any independent data that confirm the possibility expressed by the empirical data in Chapter 3 that this default case is accusative. The set of prepositions that select accusative complements is more complex than in German, as there are some which do not express a distance covered or a goal, e.g. double case prepositions like *með* 'with' or *við* 'at', as illustrated below:

- (96) a. Ég er með **blá** **augu**  
 I am with blue<sub>ACC</sub> eyes<sub>ACC</sub>  
 'I have blue eyes.'
- b. Ég sit við **borðið**  
 I sit at table.the<sub>ACC</sub>  
 'I sit at the table.'

This means that we cannot easily apply the proposal in Van Riemsdijk (2007) to Icelandic since accusative case in PPs is not necessarily correlated to covering distances or a goal. As for the apposition test presented earlier for German, appositions in Icelandic normally follow the case of the noun they refer to, so we cannot use Van Riemsdijk's argument to find out the possible default case of prepositional complements in Icelandic:

- (97) a. Við ætlum að tala um Arnald Indriðason, **rithöfund**/\***rithöfundi**  
 we intend to talk about Arnaldur<sub>ACC</sub> Indriðason<sub>ACC</sub> writer<sub>ACC/DAT</sub>  
 'We are going to talk about Arnaldur Indriðason, writer.'
- b. Ég sakna Whitney Houston, **einnar**/\***einni** af bestu söngkonum síðustu  
 I miss Whitney Houston one<sub>GEN/DAT</sub> of best singer.women last<sub>GEN</sub>  
 áratuga  
 decades<sub>GEN</sub>  
 'I miss Whitney Houston, one of the best female singers of the last decades.'

There are, however, other clues in favor of accusative as the default case in Icelandic PPs. For instance, adverbs like *upp* ‘up’, *niður* ‘down’ or *út* ‘out’, which are generally not supposed to assign case, can sometimes act as prepositions, as illustrated in the following examples. Interestingly, the noun that follows is in accusative case. Also note that it is possible to topicalize these so-called adverbs together with the noun they follow, which suggests that together they form a PP (the examples are adapted from Práinsson (2005:115); also cf. the discussion offered in Práinsson (2009:42ff.)):

- (98) a. Nokkrir útlendingar gengu **upp/niður stigann**  
 some foreigners went up/down stairs.the<sub>ACC</sub>  
 ‘Some foreigners went upstairs/downstairs.’
- b. Stór skúta sigldi **út fjörðinn**  
 big sailboat sailed out fjord.the<sub>ACC</sub>  
 ‘A big sailboat sailed all the way out of the fjord.’
- (99) a. [**Upp/Niður stigann**] gengu nokkrir útlendingar  
 up/down stairs.the<sub>ACC</sub> went some foreigners  
 ‘Some foreigners went upstairs/downstairs.’
- b. [**Út fjörðinn**] sigldi stór skúta  
 out fjord.the<sub>ACC</sub> big sailboat sailed  
 ‘A big sailboat sailed all the way out of the fjord.’

The reader might point out, however, that these kind of examples match the description given by Van Riemsdijk (2007) of accusative PPs in German, as they express a distance covered or a goal. However, it is not always the case in Icelandic (unlike German) that, whenever a PP expresses a distance covered (whether in time or space), a route or a goal, the complement is in accusative, therefore we cannot take for granted that Van Riemsdijk’s argument applies to Icelandic as well. The preposition *meðfram* ‘along’, for instance, takes a dative complement instead of one with accusative case:

- (100) Hann labbaði meðfram **ströndinni**  
 he walked along coast.the<sub>DAT</sub>  
 ‘He walked along the coast.’

Furthermore, another piece of evidence that suggests that accusative is a default case in PPs in Icelandic is the fact that certain prepositions assigning genitive case display case variation. Consider the preposition *innan* ‘within, inside’, which assigns genitive as well as accusative case. Práinsson (2005:119ff.) reported that *innan* assigns accusative case in instances of movement, otherwise genitive:

- (101) a. Hestarnir komu **innan** dalinn  
 horses.the came inside valley.the<sub>ACC</sub>  
 ‘The horses came inside the valley.’

- b. Það er meiri skógur **innan** girðingarinnar  
 it is bigger woods within fence.the<sub>GEN</sub>  
 'There are bigger woods within the fence.' (Práinsson 2005:120)

This difference between cases, however, is not to be generalized, as *innan* can also refer to a timeframe and both accusative and genitive are possible, without any difference in meaning. Interestingly, dative case is not felicitous compared to accusative, as illustrated in the following example:

- (102) Innan **tveggja daga** /  **tvo daga** / ?\***tveimur dögum**  
 within two<sub>GEN</sub> days<sub>GEN</sub> two<sub>ACC</sub> days<sub>ACC</sub> two<sub>DAT</sub> days<sub>DAT</sub>  
 'Within two days'

Now, if a) we assume that genitive is more marked than accusative or dative (see, for example, analyses like Caha (2009) or Bayer et al. (2001)) and b) we suppose that accusative and genitive are interchangeable with *innan* on the basis of (102), then the fact that it is accusative rather than dative that emerges as an alternative case to genitive suggests that it could be a default case in PPs.

In this regard, from a cross-linguistic perspective, the data from Faroese, which is the closest language to Icelandic, offer some insights into these kind of prepositions. As shown in Thráinsson et al. (2012), genitive case is not productive like in Icelandic, and the great majority of genitive forms of nouns are found in fixed expressions. Hence, prepositions like *innan* 'within', *uttan* 'outside' and *millum* 'between', which historically were followed by genitive word forms, are now followed by accusative forms, as shown below (see Thráinsson et al. 2012:176ff.). Given that Icelandic and Faroese are very similar languages, the changes in the case system of the latter corroborate the possibility that the default case in PPs might be accusative rather than dative in both languages:

- (103) a. innan **garðs**  
 within wall<sub>GEN</sub>  
 'Inside the walled-off area' (fixed expression)
- b. innan **ein mánað**  
 within one<sub>ACC</sub> month<sub>ACC</sub>  
 'Within one month'
- (104) a. uttan **veggja**  
 outside walls<sub>GEN</sub>  
 'Outside' (fixed expression)
- b. uttan **mat og drekka**  
 without food<sub>ACC</sub> and drink<sub>ACC</sub>  
 'Without food and drink'
- (105) a. millum **fjals og fjøru**  
 between mount<sub>GEN</sub> and shore<sub>GEN</sub>  
 'Between mountains and shore' (fixed expression)

- b. millum **oyggjarnar**  
 between islands.the<sub>ACC</sub>  
 'Between the islands'

FAROESE

It is true, however, that there are some pieces of evidence in favor of dative as the default case for PPs in Icelandic, which appear to contrast with the results from clausal nominalization and our discussion so far. Let us take a look at the preposition *á* 'on' in particular when it is selected by nouns. Generally speaking, nouns take by default a genitive DP as complement. However, alternatively, many nouns in Icelandic can select *á* followed by a DP complement in dative case (cf. Thráinsson 2007:94–95):

- (106) a. Lýsingin **á myndinni**  
 description.the on picture.the<sub>DAT</sub>  
 'The description of the picture'
- b. Verðið **á þessari vöru**  
 price.the on this<sub>DAT</sub> good<sub>DAT</sub>  
 'The price of this good'

The problem in this regard is that, since *á* + DAT is the most frequent pattern among PP complements of nouns, it is puzzling that it is dative and not accusative that emerges after the preposition, which begs the question as to whether dative is the actual default case in PPs.

Despite this counterargument, however, there are some data – unfortunately not independent, as they are based on clausal nominalization, but interesting nonetheless – related to the preposition *á* that point once again to the possibility of accusative as the default case in PPs. We can find in Icelandic certain nouns that select *á* as a functional preposition but are followed by a noun in accusative case, like *gagnrýni* 'criticism':

- (107) Gagnrýni **á kerfið**  
 criticism on system.the<sub>ACC</sub>  
 'The criticism of the system'

Now, if we compare the distribution of NomCIs and non-NomCIs between nouns followed by *á* + DAT and those followed by *á* + ACC, the typical contrast between dative and accusative we observed in Chapter 3 emerges once again. Let us take into account nouns like *lýsing* 'description' and *skilgreining* 'definition', which both take *á* + DAT as complement, and *gagnrýni* which selects *á* + ACC as mentioned above. A quick search in RMH returns the results presented in Table 4.3, in which non-NomCIs are more preferable when accusative case is assigned while NomCIs are more preferable in instances of dative case, corroborating once again the possibility that accusative does not correspond to a case feature to check:

| Noun                                     | -það | +það | %það          |
|------------------------------------------|------|------|---------------|
| <i>lýsing</i> ‘description’ á + DAT      | 34   | 150  | <b>81.52%</b> |
| <i>skilgreining</i> ‘definition’ á + DAT | 107  | 515  | <b>82.79%</b> |
| <i>gagnrýni</i> ‘criticism’ á + ACC      | 354  | 60   | <b>14.49%</b> |

Table 4.3: Data from RMH on *lýsing*, *skilgreining* and *gagnrýni*

Overall, it appears there are various pieces of evidence that suggest that the possibility that accusative is the default case in prepositional complements in Icelandic is valid, although it is unclear why á + DAT is a frequent pattern among complements of nouns.

#### 4.4.6 Beyond case: gender and number

Now that we have completed our discussion on case assignment, let us focus on gender and number. Recall that clausal nominalization in postcopular clauses is not preferable when það surfaces in nominative neuter singular, whereas the frequency of það increases as soon as the gender or the number is non-default:

(108) a. *A NomCl is preferable:*

Afleiðingin er (**sú**) að Bandaríkin eru í alvarlegri skuldakreppu  
 consequence.the<sub>F-PL</sub> is that<sub>F-PL</sub> that USA.the are in serious debt.crisis  
 ‘The consequence is that the USA is in a serious debt crisis.’

b. *A non-NomCl is preferable:*

Vandamálið er (**það**) að fólk vill ekki láta bólusetja sig  
 problem.the<sub>N-SG</sub> is that<sub>N-SG</sub> that people wants not let vaccinate oneself  
 ‘The problem is that people do not want to be vaccinated.’

On the basis of our discussion on case and, in particular, of the contrast between nominative and accusative on the one hand and dative and genitive on the other, we have a similar contrast here between neuter singular and the other combinations of gender and number features, which suggests that neuter and singular could be a default gender and number respectively. This view aligns with other accounts like Rögnvaldsson (1990:52ff.) in relation to the fact that singular and neuter are considered the unmarked traits of number and gender respectively as grammatical categories. However, the question is whether there is any independent evidence in support of the possibility that neuter and singular are default gender and number features in Icelandic and do not actually correspond to features to check in syntax, similarly to nominative and accusative case. A first piece of evidence in favor of this possibility can be found in predicative adjectives. Consider the following examples:

(109) a. Karl er **skemmtilegur**  
 Karl<sub>M-SG-NOM</sub> is funny<sub>M-SG-NOM</sub>  
 ‘Karl is fun.’

- b. Strákarnir eru **skemmtilegir**  
 boys.the<sub>M-PL-NOM</sub> are funny<sub>M-PL-NOM</sub>  
 'The boys are funny.'

Generally speaking, as commonly known, a predicative adjective in Icelandic agrees in gender, number and case with the relevant subject. However, whenever agreement cannot be established, e.g. when the subject is not in structural case, the adjective defaults to neuter singular, as illustrated in the following examples, where the predicative adjective cannot agree with the gender and number properties of the dative subject (cf. e.g. Thráinsson (2007:238ff.)):

- (110) a. Strákunum er **kalt**  
 boys.the<sub>M-PL-DAT</sub> is cold<sub>N-SG-NOM</sub>  
 'The boys are (feeling) cold.'
- b. \*Strákunum er(u) **kaldir**  
 boys.the<sub>M-PL-DAT</sub> is/are cold<sub>M-PL-NOM</sub>  
 Intended: 'The boys are (feeling) cold.'
- c. \*Strákunum er(u) **köldum**  
 boys.the<sub>M-PL-DAT</sub> is/are cold<sub>M-PL-DAT</sub>  
 Intended: 'The boys are (feeling) cold.'

Furthermore, another piece of evidence can be found in gender and number agreement between conjoined DP subjects, where predicative adjectives become less systematic if the nouns involved are abstract. This is captured in the literature, for instance, by the so-called *Individuation hierarchy* (see Thorvaldsdóttir (2019); cf. other analyses like Dixon (1994) and Audring (2009)):

- (111) *Individuation Hierarchy*  
 human > animal > countable object > uncountable abstract

Thorvaldsdóttir (2019) argued that high individuation (e.g. animacy) tends to correlate with a syntactic (see (112)) or semantic resolution (see (113)) in the predicative adjective between the gender and the number of the DPs involved, while low individuation (e.g. abstract nouns) correlates with partial agreement of one of the two DPs with the adjective (as in (114)) or with a semantic default neuter singular independently of the gender of the nouns (see (115)); the examples are from Thorvaldsdóttir (2019:27–28); further examples on conjoined DPs can also be found in Friðjónsson (1989)):

- (112) a. Diskurinn og bollinn eru **bláir**  
 plate.the<sub>M-SG</sub> and cup.the<sub>M-SG</sub> are blue<sub>M-PL</sub>  
 'The plate and the cup are blue.'
- b. Diskurinn og bollinn eru **blá**  
 plate.the<sub>M-SG</sub> and cup.the<sub>M-SG</sub> are blue<sub>N-PL</sub>  
 'The plate and the cup are blue.'

- (113) Ófríska skáldið og Jóna eru **ánægðar**  
 pregnant<sub>N-SG</sub> poet.the<sub>N-SG</sub> and Jóna<sub>F-SG</sub> are pleased<sub>F-PL</sub>  
 ‘The pregnant poet and Jóna are pleased.’
- (114) a. **Ánægja** og **árangur** er **mikilvægur** í lífinu  
 joy<sub>F-SG</sub> and success<sub>M-SG</sub> is important<sub>M-SG</sub> in life.the  
 ‘Joy and success is important in life.’  
 b. **Ánægja** og **árangur** er **mikilvæg** í lífinu  
 joy<sub>F-SG</sub> and success<sub>M-SG</sub> is important<sub>F-SG</sub> in life.the  
 ‘Joy and success is important in life.’
- (115) **Ánægja** og **árangur** er **mikilvægt** í lífinu  
 joy<sub>F-SG</sub> and success<sub>M-SG</sub> is important<sub>N-SG</sub> in life.the  
 ‘Joy and success is important in life.’

From a semantic perspective, the DPs that are semantically closer to NomCIs and non-NomCIs are abstract nouns, which are characterized by low individuation, as they represent an event or a state of affairs (see Masuoka (2007); Hara et al. (2013)). With this category of nouns, we can observe that the outcome of the predicative adjective can either be in agreement with one of the two conjoined DPs or default to neuter singular. This aspect is also evident with conjoined clausal subjects. According to McCloskey (1991), there are cases in English where the verb can emerge in the plural when there are two or more clausal arguments coordinated:

- (116) a. That the president will be reelected and that he will be impeached **are** equally likely at this point  
 b. That the march should go ahead and that it should be canceled **have** been argued by the same people at different times

In Icelandic, however, the verb remains in the singular even if there are two coordinate clauses that form the subject. This corroborates the idea that neuter and singular are default gender and number respectively:

- (117) Að John deyi í lok þáttarins og að hann lifi af **er/\*eru** ekki  
 that John dies in end episode.the and that he survives is/are not  
**mikilvægt/\*mikilvæg** í söguþræðinum  
 important<sub>N-SG/N-PL</sub> in plot.the  
 ‘That John dies at the end of the episode and that he survives are not important in the plot.’

Considering these facts, it is plausible that neuter and singular as well as nominative and accusative do not correspond to features to check, and cause *það* to be less preferable due to the fact that the purpose of *það* is specifically to check features that a CP would otherwise be unable to check by itself.

## 4.5 Cross-linguistic analysis

### 4.5.1 Introduction

In this third section we will focus on comparing the distribution of NomCIs and non-NomCIs in Icelandic with other languages. Cross-linguistically speaking, the distribution of NomCIs is quite extensive: we have seen in the previous chapters examples from English (see Rosenbaum (1967); Kaltenböck (2003); Kallulli (2006); Kim and Sag (2005); Shahar (2008); Gentens (2016) among many others), Modern Greek (e.g. Roussou (1991)), Persian (see Farudi (2007); Jahromi (2011)), Swedish (e.g. Josefsson (2006); Faarlund (2019)) and Russian (see Hartman (2012); Knyazev (2016)), but the same phenomenon can also be encountered in many other languages, like German (e.g. Sudhoff (2016); Axel-Tober et al. (2016); Zimmermann (2016)), other Mainland Scandinavian languages (see Faarlund (2019)), Polish (e.g. Hartman (2012)), Ndebele (see Pietraszko (2019)), Spanish (e.g. Delicado Cantero (2013); Picallo (2002)), Hungarian (see De Cuba and Ürögdi (2010)), Hindi (e.g. Subbarao (1984)) and Hebrew (Zuckermann (2006); Kastner (2015)). It would be interesting to observe how CNs function in all of these languages. However, for reasons of space, it is impossible to discuss them all in this dissertation. Therefore, I picked a small sample of languages which display clausal nominalization on the basis of specific criteria, which are the following: a) whether the language is a V2 language (marked in the list below with the trait [ $\pm$ V2]); b) whether it displays a morphological case system (marked with the trait [ $\pm$ mcase]); c) whether the language is very close to Icelandic or not (marked with the trait [ $\pm$ relative]):

(118) *Languages*

- a. Persian: [-V2], [-mcase], [-relative]
- b. Russian: [-V2], [+mcase], [-relative] (cases: Nom, Acc, Dat, Gen, Ins, Prep)
- c. Swedish: [+V2], [-mcase], [+relative]
- d. German: [+V2], [+mcase], [+relative] (cases: Nom, Acc, Dat, Gen)

This categorization, on the one hand, is based on our discussion on mandatory and optional nominalization in nominative clausal subjects. The Icelandic data have shown that clausal subjects can emerge as CPs in first position, but not after the finite verb when the whole sentential subject surfaces in Spec,T. However, not all languages that adopt clausal nominalization are V2 languages. Therefore, the question is whether there is any difference as to how clausal subjects are derived in V2 and non-V2 languages and, if there is, how the lack of V2 structure can account for this difference. Alternatively, if there is no difference in how they are derived, we have the possibility of extending the nominalization mechanism proposed for Icelandic to non-V2 languages as well, i.e. clausal subjects must be DPs by virtue of D-feature checking.

As for the second criterion, the empirical data from Icelandic in Chapter 3 have shown that the distribution of clausal nominalization is mainly triggered by case marking. The questions to answer in this case are two: a) whether languages that display a morphological case system

behave in a similar way as in Icelandic and b) which factors determine clausal nominalization in absence of a morphological case system.

Regarding the third criterion, the main question is whether Icelandic and its close relatives share similar characteristics when it comes to the distribution of CNs and, at the same time, whether distant languages present different characteristics compared to Icelandic clausal nominalization (and why that might be), or whether they still have common characteristics in this regard.

We are going to compare the four languages listed above with Icelandic by taking a look at canonical argument positions, postcopular clauses and prepositional phrases. I will skip the rest of the syntactic positions that can be affected by nominalization in Icelandic to keep the analysis simple, also due to the fact that some of them are rare (see indirect objects and complements of adjectives).

## 4.5.2 Swedish

Let us start taking a look at Swedish, which is the closest language to Icelandic on the list in (118). Swedish, like Icelandic, is a V2 language, but it does not display a morphological case system. On the basis of CNM proposed for Icelandic, what do we expect to find in a language like Swedish? If we consider the fact that it is a V2 language, we should expect nominative subjects to surface in two positions, before and after the finite verb. The position after the finite verb is Spec,T, so that nominalization should be mandatory there. As for the first position, we should be able to find an optional determiner instead.<sup>30</sup> The data from Josefsson (2006:footnote 12) confirms our predictions, as the determiner *det* ‘that’ is mandatory after the finite verb, but not before it (I follow Faarlund (2019) in relation to the idea that *det* before clauses is a demonstrative):

- (119) a. Därför är **\*(det)** att professorn avgår just nu ganska olyckligt  
 therefore is that<sub>NOM</sub> that professor.the resigns just now quite unfortunate
- b. **(Det)** att professorn avgår just nu är därför ganska olyckligt  
 that<sub>NOM</sub> that professor.the resigns just now is therefore quite unfortunate  
 ‘Therefore it is unfortunate that the professor is resigning right now.’ SWEDISH

Let us now move to clausal direct objects. Due to the fact that Swedish does not provide a morphological case system, we can expect direct objects to surface in accusative case only. Under these circumstances, CNM helps to predict that we can generally find instances of optional *det* in direct object position. Here are some examples from Brandtler (2013:110), which confirm our predictions:

<sup>30</sup>In this cross-linguistic analysis, I will not follow the same methodology as in Chapter 3 (i.e. administering questionnaires and gathering a wide sample of data from corpora) for each of the four languages listed in (118), as such a work would be outside the scope of this dissertation.

- (120) a. Jag beklagar (**det**) att jag sårade dig  
 I regret that that I hurt you  
 'I'm sorry that I hurt you.'
- b. Jag sa (**det**) att han förmodligen skulle förlora  
 I said that that he probably would lose  
 'I told him he would probably lose.'

SWEDISH

I could not find examples of Swedish verbs that are obligatorily followed by a NomCI, which suggests that transitive verbs that can select a clausal complement optionally allow *det* to emerge as a general rule.

By observing structural case positions, we can clearly see that Icelandic and Swedish are very similar, as nominative (see Spec,C) and accusative case are correlated to pronominal optionality, which indicates that structural case does not correspond to features to check. This is also confirmed by postcopular clauses, where the pronoun is optional, as illustrated in the following quote (note that, similarly to Icelandic, *det* can also be inflected for gender and number in accordance with the subject of the matrix clause; the quote below contains a plural CN):

- (121) „Konsekvenserna är (**de**) att en koncern inte ska missgynnas p.g.a. den  
 consequences.the are those that a concern not shall be disadvantaged due to the  
 valda koncernformen för sin verksamhet“  
 chosen concern form.the for its activity  
 'The consequences are that a concern will not be disadvantaged due to the chosen  
 concern form for its activity.' (from Johansson (2014:8))

SWEDISH

Let us focus on prepositional phrases. In the case of Icelandic, we have seen in Chapter 3 that case determines whether *það* is more or less preferable (or even mandatory) after functional prepositions. Lexical prepositions, on the other hand, tend to require *það* independently of case, which indicates that they need to be followed by a DP. Since case variety is also excluded in prepositional complements in Swedish, we can expect, in light of our observations in Icelandic, that functional prepositions would tend to select *det* optionally, while lexical prepositions would select *det* obligatorily. However, Swedish appears to not follow the expected pattern. Contrarily to Icelandic, where we find a clear distinction between functional and lexical prepositions, Swedish applies optional nominalization to both categories. In other words, Swedish lexical prepositions do not subcategorize for DPs only:

- (122) a. Hon är stolt över (**det**) att hennes son fick jobbet  
 she is proud over that<sub>ACC</sub> that her son got job.the  
 'She is proud that her son got the job.' (adapted from Faarlund (2019:59))
- b. Vi tror på (**det**) att tillstånd skapar transformationer  
 we believe on that<sub>ACC</sub> that posture creates transformations  
 'We believe that posture causes transformations.' (adapted from vindoga.nu)

- (123) a. Hitta svar på alla praktiska frågor omkring **(det)** att flytta till ett annat nordiskt land  
 find answer on all practical questions around that<sub>ACC</sub> to move to one other Nordic country  
 'Find an answer to all practical questions about moving to another Nordic country.'  
 (adapted from nordfokus.org)
- b. Att inte se trots **(det)** att vi tittar  
 to not see despite that<sub>ACC</sub> that we look  
 'Not seeing despite the fact that we look.'<sup>31</sup>

SWEDISH

The comparison between Swedish and Icelandic indicates that, independently of whether a certain language displays a morphological case system or not, case in structural case positions does not correspond to features to check, as nominative and accusative case are correlated to pronominal optionality. Moreover, Swedish indirectly supports my hypothesis proposed on the basis of the Icelandic data that lexical case is a trigger of either a high frequency of CNs or pronominal obligatoriness, as CNs in Swedish are optional as a general rule due to a lack of lexical case features to check, unlike Icelandic. An important question, on the other hand, is whether it is a general cross-linguistic trend that lexical case triggers obligatory nominalization or, alternatively, a high frequency of NomCIs. We will verify whether this is the case in German and Russian in the next subsections.

### 4.5.3 German

Let us now observe German, another Germanic language which is both characterized by V2 as well as a morphological case system like Icelandic. In Chapter 2, in relation to our discussion on Thráinsson (1979) and his factivity hypothesis (§2.5.2), we said that there are two instances of *það* which can at times be ambiguous, especially in the first position when clausal extraposition has occurred: expletive *það* and *það* as CN. In German, by contrast, there are two different pronouns that can precede clauses, the personal pronoun *es* 'it', which also plays the role of a subject expletive, and the distal demonstrative *das* 'that' (see e.g. Axel-Tober et al. (2016)). They differ both in relation to their function and distribution. Pütz (1975) and in particular Axel-Tober et al. (2016), who compared the two pronouns in an experiment involving left discourse accessibility, showed that *das* is discourse-anaphoric as it needs to access left discourse, while *es* can simply be a placeholder. As for their distribution, let us take a look at both pronouns in action. The first difference between these pronouns is that *es* only precedes extraposed clauses, while *das* is correlated to optional extraposition (see in particular Axel-Tober et al. 2016:52):

<sup>31</sup>This example is adapted from the title of an episode of *Spanarna* (a Swedish podcast series). The whole title is *Att inte se trots att vi tittar, Från dialog till monolog & Sluta filma!* 'Not seeing despite the fact that we look, From dialogue to monologue and Stopping filming'.

- (124) a. Er hat **das**/\***es**, dass er schwarz gewählt hat, bedauert  
 he has it/that that he black voted has regretted  
 'He regretted having voted black.'
- b. Er hat **das/es** bedauert, dass er schwarz gewählt hat  
 he has it/that regretted that he black voted has  
 'He regretted having voted black.'

GERMAN

Let us take a closer look at *es*. It is commonly known that *es* is mandatory both before and after the finite verb. This differs from Icelandic, as expletive *það* is not phonetically realized after the finite verb:<sup>32</sup>

- (125) a. \*(**Es**) ist schwierig, diese Aufgabe zu lösen  
 it is difficult this assignment to complete  
 'It is difficult to solve this assignment.'
- b. Ist \*(**es**) schwierig, diese Aufgabe zu lösen?  
 is it difficult this assignment to complete  
 'Is it difficult to solve this assignment?'

GERMAN

However, it is also possible to find instances where *es* is optional or even ungrammatical. In Sudhoff (2016:38), we find a couple of such examples, which are shown in (126):

- (126) a. Niemanden interessiert (**es**), dass Marie wegfährt  
 no-one<sub>ACC</sub> is interested in it that Marie away.goes  
 'No one is interested about the fact that Marie left.'
- b. In der Zeitung steht (\***es**), dass ein Unwetter kommen wird  
 in the newspaper is said it that a thunderstorm come will  
 'In the newspaper it is said that a thunderstorm will come.'

Interestingly, if these examples are rendered as yes/no-questions, *es* becomes mandatory in both instances (see the examples below). This confirms that Spec,T is, generally speaking, a position where *es* must appear, in line with D-feature checking as proposed for Icelandic. Nevertheless, it is possible that other competing derivations cause *es* to be optional with verbs like *interessieren* 'interest', in a similar way as we have seen with Icelandic in instances of extraposition (see §4.3), where *það* was optional. Although the case of *stehen* 'stand, be said' remains mysterious due to the fact that the results with topicalization and yes/no-questions are diametrically opposite, it is still a fact that *es* is still mandatory in yes/no-questions, which resonates with the examples from Icelandic (see §4.3):<sup>33</sup>

<sup>32</sup>This contrast can be clearly seen with weather verbs, as is commonly known:

- (i) a. **Það** rignir / Rignir (\***það**)?  
 it rains rains it  
 'It is raining. / Is it raining?'
- b. **Es** regnet / Regnet \*(**es**)?  
 it rains rains it  
 'It is raining. / Is it raining?'

ICELANDIC

GERMAN

<sup>33</sup>Things complicate further if we add, for example, an adverb like *klar* 'clearly' to the example with the verb *stehen*. The speakers of German I contacted consider *es* optional rather than ungrammatical, which is a similar result to

- (127) a. Interessiert \*(**es**) dich, dass Marie wegfährt?  
 interests it you<sub>ACC</sub> that Marie away.goes  
 'Are you interested about the fact that Marie left?'  
 b. Steht \*(**es**) in der Zeitung, dass ein Unwetter kommen wird?  
 is said it in the newspaper that a thunderstorm come will  
 'Is it said in the newspaper that a thunderstorm will come?' GERMAN

Let us now move to the discourse-anaphoric *das*, which is relevant for our analysis due to the fact that it behaves exactly like Icelandic *það* when it comes to its distribution with clausal subjects (see the examples below). This suggests that, in German, nominative does not correspond to a feature to check in syntax and also that D-feature checking in Spec,T could be the trigger of mandatory nominalization, in line with the CNM proposed on the basis of the Icelandic data:

- (128) a. (**Das**), dass er zu spät gekommen ist, hat mich nicht überrascht  
 that<sub>NOM</sub> that he too late come is has me not suprised  
 'The fact that he came too late did not surprise me.'  
 b. Hat \*(**das**), dass er zu spät gekommen ist, dich nicht überrascht?  
 has that<sub>NOM</sub> that he too late come is you not suprised  
 'Did the fact that he came too late not surprise you?' GERMAN

Postcopular clauses also confirm the same hypothesis in relation to nominative, as the determiner is unsurprisingly optional according to the native speakers of German I contacted (note that only *das* is possible in the example below, and not *es*):

- (129) Die Konsequenz ist (**die/\*es**), dass wir mehr Geld schulden  
 the consequence<sub>FEM-SG</sub> is that<sub>FEM-SG</sub>/it that we more money owe  
 'The consequence is that we owe more money.' GERMAN

In relation to accusative clausal direct objects, on the other hand, Sudhoff (2016) made a distinction between German verbs that take an optional *es*, like *bedauern* 'regret', and verbs that subcategorize for CPs only, like *behaupten* 'assert', which cannot accept *es* (shown below).<sup>34</sup> This indicates that *es* only surfaces in DP positions, in a similar way as *það* does in Icelandic. At the same time, considering the fact that accusative case does not appear to be correlated to

---

the one with *interessieren*:

- (i) In der Zeitung steht (**es**) klar, dass ein Unwetter kommen wird  
 in the newspaper is said it clearly that a thunderstorm come will  
 'In the newspaper it is clearly said that a thunderstorm will come.'

<sup>34</sup>It is true, however, that the discourse anaphor *das* can surface with verbs like *behaupten* if a previous context has been mentioned (see Sudhoff (2016:40)):

- (i) a. *Wer behauptet, dass Marie wegfährt?*  
 'Who asserts that Marie is going away?'  
 b. Peter behauptet (**das**), dass Marie wegfährt  
 Peter asserts that<sub>ACC</sub> that Marie away.goes  
 'Peter asserts that Marie is going away.'

pronominal obligatoriness, we can also suppose that structural case in German does not correspond to features to check:

- (130) a. Peter bedauert (**es**), dass Marie wegfährt  
 Peter regrets it that Marie away.goes  
 'Peter regrets that Marie is going away.' (Sudhoff (2016:33))
- b. Peter behauptet (**\*es**), dass Marie wegfährt  
 Peter asserts it that Marie away.goes  
 'Peter asserts that Marie is going away.' (Sudhoff (2016:33)) GERMAN

As for lexical case, the pronoun that surfaces before clauses is *dem* for dative and *dessen* for genitive, as illustrated in the following examples (cf. Zimmermann (2016); also Frey (2016:footnote 6)):

- (131) a. Der Beweis hat **dem** widersprochen, dass er sie ermordet hat  
 the evidence has that<sub>DAT</sub> contradicted that he her murdered has  
 'The evidence has contradicted the fact that he murdered her.'
- b. Ich kann mich **dessen** nicht entsinnen, dass er seine Mutter angelogen hat  
 I can me<sub>ACC</sub> that<sub>GEN</sub> not remember that he his mother lied has  
 'I cannot recall him ever lying to his mother.' GERMAN

From a topological perspective, these pronouns appear to be an instance of *das* rather than *es*, due to the fact that clausal extraposition is not mandatory (*es* is correlated to obligatory movement of the CP).<sup>35</sup>

- (132) a. Der Beweis hat **dem**, dass er sie ermordet hat, widersprochen  
 the evidence has that<sub>DAT</sub> that he her murdered has contradicted  
 'The evidence has contradicted the fact that he murdered her.'
- b. Ich kann mich **dessen**, dass er seine Mutter angelogen hat, nicht entsinnen  
 I can me<sub>ACC</sub> that<sub>GEN</sub> that he his mother lied has not remember  
 'I cannot recall him ever lying to his mother.' GERMAN

In other words, assuming the analysis in Axel-Tober et al. (2016), German displays a pattern in canonical argument positions with structural case where *es* introduces clausal subjects and objects by default, although clausal extraposition co-occurs obligatorily, while *das* is adopted as a discourse anaphor. As for lexical case positions, it is *das* that surfaces and not *es*.

The question, at this point, is whether lexical case is correlated to pronominal obligatoriness in German. If this expectation is satisfied, then we can also infer that: a) the structural vs. lexical case dichotomy of Icelandic and German is similar when it comes to complement clauses; and b) from a cross-linguistic perspective, the presence of a morphological case system can be correlated to obligatory nominalization in lexical case positions. The data from dative and genitive clausal complements align to much extent with our expectations, except when extraposition takes place (see the case of genitive in (134); also Frey (2016)):

<sup>35</sup>I differ here from accounts like Zimmermann (2016), who assumes in her analysis that *dem* and *dessen* are dative and genitive entries for *es*.

- (133) a. Der Beweis hat **\*(dem)**, dass er sie ermordet hat, widersprochen  
 the evidence has that<sub>DAT</sub> that he her murdered has contradicted  
 'The evidence has contradicted the fact that he murdered her.'
- b. Ich kann mich **\*(dessen)**, dass er seine Mutter angelogen hat, nicht entsinnen  
 I can me<sub>ACC</sub> that<sub>GEN</sub> that he his mother lied has not remember  
 'I cannot recall him ever lying to his mother.'
- (134) a. Der Beweis hat **\*(dem)** widersprochen, dass er sie ermordet hat  
 the evidence has that<sub>DAT</sub> contradicted that he her murdered has  
 'The evidence has contradicted the fact that he murdered her.'
- b. Ich kann mich **(dessen)** nicht entsinnen, dass er seine Mutter angelogen hat  
 I can me<sub>ACC</sub> that<sub>GEN</sub> not remember that he his mother lied has  
 'I cannot recall him ever lying to his mother.' GERMAN

It is possible that the reason why *dessen* is optional with clausal extraposition is that the matrix verb *entsinnen* is also able to subcategorize for a CP, which, in turn, must be extraposed as CPs normally cannot remain in the middle field in German (see e.g. Sudhoff (2003)).<sup>36</sup> As a result, *dessen* in (133b) turns out to be mandatory because the non-NomCl counterpart must be extraposed.

Despite some differences between German and Icelandic, in a general sense German also appears to make a distinction between structural case and lexical case when it comes to clausal nominalization. The fact that this distinction is not displayed, by contrast, in languages close to Icelandic without a morphological case system like Swedish, suggests that the existence of such a system is a necessary underlying condition for nominalization to be triggered on the basis of case feature checking. In absence of it, the only feature that triggers clausal nominalization is a D-feature, as structural case does not appear to correspond to a feature that can trigger clausal nominalization.

The last point we can make about German before moving to the next language to study is about prepositional phrases. In this case, there is an important distinction between German and Icelandic. In German, the determiner is replaced with the morpheme *da(r)-*, that is combined with the relevant functional preposition (note that the clausal complement can be extraposed). There is no difference as to whether the case assigned is dative or accusative. The morpheme still surfaces, as shown in the following examples with *über* 'over', which selects a complement in accusative case and *vor* 'before, of' which selects a complement in dative case (cf. Zimmermann (2016) among others):

- (135) a. Ich freue mich **darüber**, dass er kommt  
 I rejoice myself it.over that he comes  
 'I'm happy that he is coming.'
- b. Ich freue mich **über** die Einladung  
 I rejoice myself over the<sub>ACC</sub> invitation  
 'I'm happy about the invitation.'

<sup>36</sup>There are, however, some exceptions to this rule, as shown by Frey (2016:79) with the verb *bedauern*.

- (136) a. Ich habe **davor** Angst, dass er die Wahrheit sagt  
 I have it.of fear that he the truth says  
 'I'm afraid he tells the truth.'
- b. Ich habe Angst **vor** dem Tod  
 I have fear of the<sub>DAT</sub> death  
 'I'm afraid of death.'
- GERMAN

Also note that the *da(r)+P* compound forms one constituent with the relevant CP, as shown in the following examples. Furthermore, *da(r)+P* cannot be removed if extraposition does not take place, in a similar way as with NomCIs with lexical case (also cf. Frey (2016)):

- (137) a. [**(Darüber)**, dass wir Peter einladen], hat sich Maria sehr gewundert  
 it.about that we Peter invite has herself Maria a lot wondered  
 'Mary was very surprised about our inviting Peter.' (Frey 2016:88–89)
- b. Maria hat sich [**(darüber)**, dass wir Peter einladen], sehr gewundert  
 Maria has herself it.about that we Peter invite a lot wondered  
 'Mary was very surprised about our inviting Peter.'
- GERMAN
- (adapted from Frey (2016:89))

This system differs from Icelandic due to the fact that, in order to suppress the determiner, the preposition should be removed as well. In other words, nominalization is mandatory if the preposition must surface, otherwise both the preposition and the morpheme together are removed, as in the following example (adapted from Schwabe et al. (2016:1)):

- (138) Max hat sich (**darüber**) gefreut, dass Lea wieder gesund war  
 Max has himself it.about enjoyed that Lea again well was  
 'Max was glad about the fact that Lea was well again.'
- GERMAN

Hence, the distribution of NomCIs and non-NomCIs in German in prepositional complements does not seem to depend on case, but rather on whether the preposition is mandatory or optional. This is an aspect of nominalization that could not be captured in Icelandic, as the contrast in the examples above is not between DPs and CPs in the complement position of P, but rather between CPs and PPs in the complement position of the matrix verb.<sup>37</sup> This particular rule of German, however, does not apply to all prepositions. Lexical prepositions, in fact, still allow the classic nominalization we have seen in Icelandic. Here is an example from the preposition *aufgrund*

<sup>37</sup>It is true, however, that CPs and PPs as possible competing variants in the complement position of a verb do exist in Icelandic, and specifically with verbs like *þvinga* 'force', *hjálpa* 'help' (see §2.4.4), *langa* 'want' (see §3.3.4) or *hvetja* 'encourage' among others. The verbs of this class either take a CP complement in a non-DP position or a PP introduced by *til* 'to'. The inner complement of the PP can also be nominalized:

- (i) a. Ég hjálpaði honum að flytja  
 I helped him to move out  
 'I helped him to move out'
- b. Ég hjálpaði honum til (**þess**) að flytja  
 I helped him to that<sub>GEN</sub> to move out  
 'I helped him to move out'

'based on', which assigns genitive. In this case, the determiner is mandatory:

- (139) Aufgrund **\*(dessen)**, dass ...  
 based that<sub>GEN</sub> that  
 'Based on the fact that...'GERMAN

In sum, German confirms for the most part the structural-lexical case dichotomy that we found in Icelandic and it also corroborates the possibility that mandatory nominalization is triggered by lexical case in those languages that display a morphological case system. Moreover, the German data also tell us that there might be more rules that determine the distribution of NomCIs and non-NomCIs, in particular the existence of a different form of nominalization in prepositional phrases, where CPs and PPs might compete as complements of the matrix verb rather than DPs and CPs.

#### 4.5.4 Persian

Let us now move to a language that is a very distant relative of Icelandic. Persian is an SOV and *pro*-drop language without a morphological case system and without V2 (see e.g. Soheili (2024) among many others). The determiner commonly used in Persian to introduce clauses is the proximal demonstrative pronoun *in* 'this', although the distal demonstrative *an* 'that' can also be used in written language (see Jahromi (2011)).<sup>38</sup> I will only present examples with *in* in this discussion for convenience. Focusing on clausal subjects, since Persian is not a V2 language, there is only one position available where they can surface in absence of extraposition, i.e. the first position of the sentence. As we can see in the following examples, the determiner is mandatory:

- (140) a. **\*(In)** ke Maryam raft ma'alum e  
 this<sub>NOM</sub> that Maryam left clear is  
 'It is clear that Maryam left.' (Hartman 2012:36)  
 b. **\*(In)** ke zamin gerd ast Julia ra moteajjeb kard  
 this<sub>NOM</sub> that Earth round is Julia OBJ surprised  
 'It surprised Julia that the Earth is round.' (Jahromi 2011:48)PERSIAN

Now, the question is whether obligatory nominalization derives from D-feature checking, as I have claimed for Icelandic, or whether it derives from structural case assignment, as proposed, for example, by Roussou (1991) in relation to Modern Greek (see §2.5.5). When the relevant CP is extraposed, *in* becomes optional:

- (141) a. **(In)** ma'alum e [ke Maryam raft]  
 this<sub>NOM</sub> clear is that Maryam left  
 'It is clear that Maryam left.'  
 b. **(In)** Julia ra moteajjeb kard ke zamin gerd ast  
 this<sub>NOM</sub> Julia OBJ surprised that Earth round is  
 'It surprised Julia that the Earth is round.' (Jahromi 2011:49)PERSIAN

<sup>38</sup>Yousef (2018:302), however, had a different view on the interchangeability of *an* and *in* and explained that clausal subjects need to be introduced by *in* rather than *an*. I will not delve into this issue here as it is not relevant for our discussion.

If structural case corresponded to a case feature that needs to be checked, then we should expect *in* to be mandatory independently of extraposition. Since that is not the case, we cannot consider nominative case to be a trigger of obligatory clausal nominalization in Persian when the clause is not extraposed, which leaves D-feature checking a more valid option. There is, however, another piece of evidence that suggests that D-feature checking is responsible for obligatory nominalization in (140). Jahromi (2011:49) observed that if the clause is moved rightwards together with *in*, nominalization becomes optional as well:

- (142) Julia *ra* moteajjeb kard [(*in*) ke zamin gerd ast]  
 Julia OBJ surprised this<sub>NOM</sub> that Earth round is  
 'It surprised Julia that the Earth is round.' PERSIAN

This indicates that if the clausal subject escapes Spec,T, it can still emerge as a non-NomCl, in line with the data from Icelandic. It also suggests that, despite the difference in the word order between non-V2 and V2 languages, the D-feature in T remains a trigger of obligatory nominalization for both language groups.

Let us now observe direct objects more closely. Contrary to Icelandic, Swedish or German, direct objects are not correlated to pronominal optionality *in situ* as we might expect, but to pronominal obligatoriness. Here is an example:

- (143) Julia [*\*(in)* ke zamin gerd ast **ra**] midanad  
 Julia this<sub>ACC</sub> that earth round is OBJ know-3sg  
 'Julia knows that the earth is round.' (Jahromi 2011:50) PERSIAN

*In* is accompanied by the object marker *ra*, which is exclusively assigned to the accusative argument. The marker can be found at the end of the direct clausal object, as shown above, or, alternatively, immediately after *in*. It can also be reduced to *-o* (see e.g. the Persian examples in Toosarvandani (2008); I will only focus, however, on *ra* to simplify things):

- (144) Julia [*\*(in ra/-o)* ke zamin gerd ast] midanad  
 Julia this<sub>ACC</sub> OBJ that earth round is know-3sg  
 'Julia knows that the earth is round.' PERSIAN

When the clausal argument is *in situ*, *in* and *ra* have to be both visible (Jahromi (2011)). At the surface, this can be due to two possible causes that we need to explore: a) the object marker makes a requirement for a DP, which is independent of case (e.g. perhaps due to a [*•D•*]-feature); and b) accusative in Persian is a case feature that must be checked, unlike Icelandic accusative and, therefore, the CN surfaces together with the object marker as a result of case-marking. We can use clausal extraposition as a diagnostic test in order to verify whether the status of *in* changes. Interestingly, also in this case, *in* becomes optional after extraposition, but it is still dependent on the marker; if the marker surfaces, then *in* must emerge as well, otherwise it is disallowed.<sup>39</sup>

<sup>39</sup>These facts on clausal subjects and objects share similarities with the data from German: if the clause is not

- (145) Julia (**in ra**) midanad [ke zamin gerd ast]  
 Julia this<sub>ACC</sub> OBJ know-3sg that earth round is  
 ‘Julia knows that the earth is round.’ (Jahromi 2011:50)

Since extraposition causes *in ra* to be optional, then we can suppose that accusative case does not trigger clausal nominalization, otherwise we should expect it to surface obligatorily. If this assumption is correct, then the fact that the determiner must surface *in situ* means that there is a requirement for a DP due to the presence of an object marker. However, if the clause escapes that position via extraposition, non-NomCIs can still emerge, in line with the idea that structural case does not correspond to a feature to check. These facts also tell us something new about structural case positions that we could not observe in Icelandic. If a morphological element must be realized in direct object position, like a marker, then nominalization must surface.

In the other languages we have observed so far, in order to support the idea that structural case does not correspond to a formal feature to check, we have also taken a look at postcopular clauses. This specific type of clause constitutes a more complex issue in Persian, due to the fact that various speakers would tend to extrapose the CP. Note that in Persian these clauses are pre-copular as they are merged to the left of the copula (see (147)). With extraposition, *in* is optional, which confirms that nominative case does not trigger nominalization:

- (146) Zarurat (**in**) ast [ke pul bedim]  
 necessity this<sub>NOM</sub> is that money give-1pl  
 ‘It is necessary that we pay.’ PERSIAN

However, I also wanted to verify any possibility of keeping the clause *in situ*. Some native speakers from Tehran I contacted considered the following example possible, although some other speakers (who speak in particular a mixed variant of Dari and Iranian Persian) still prefer extraposition of the clause.<sup>40</sup>

- (147) Mushkil [**in**] ke emtehan ra nadadid] bud  
 problem this<sub>NOM</sub> that exam OBJ not.gave-2pl was  
 ‘The problem was that you did not take the exam.’ PERSIAN

In the grammar of those speakers who accept the embedded clause *in situ*, *in* happens to be obligatory. In other words, in Persian, it seems that the positions occupied by clausal arguments

---

nominalized it needs to be extraposed obligatorily.

<sup>40</sup>There is a further complication, however, specifically with the noun *mushkil* ‘problem’, which is homophonous with an adjective that means ‘difficult’. By applying extraposition, Persian speakers would interpret *mushkil* as an adjective rather than a noun when *in* does not surface, which causes nominalization to be mandatory for the sake of maintaining a nominal reading of *mushkil*:

- (i) Mushkil **in** bud [ke emtehan ra nadadid]  
 problem this<sub>NOM</sub> was that exam OBJ not.gave-2pl  
 ‘The problem was that you did not take the exam.’ PERSIAN

The noun *zarurat* ‘necessity’, on the other hand, is not affected by this ambiguity with an adjective. However, native speakers reported a higher degree of difficulty with this noun to keep the clausal complement *in situ*.

are only-DP positions. However, if an embedded clause can escape such positions via extraposition, non-NomCIs can still emerge, with nominalization becoming optional as a result. These data are, in a way, diametrically opposite to the ones from Swedish. Neither language provides a morphological case system, but Swedish generally displays optional nominalization (except in Spec,T), while Persian requires DPs, i.e. NomCIs, unless the embedded clause is extraposed.

Let us move to the last syntactic position, namely the one of prepositional complements. Interestingly, this is the only syntactic position where nominalization is obligatory independently of whether the clause is extraposed or not, as illustrated in the following examples from Farudi (2007:9):

- (148) a. Sara [az \*(in) ke dar in mosabeqe barande nashode] xeili  
 Sara from this<sub>OBL</sub> that in this competition winner NEG.became-3sg very  
 sharmande ast  
 ashamed is  
 'Sara is very embarrassed at the fact that she did not win the competition.'
- b. Sara az \*(in) xeili sharmande ast [ke dar in mosabeqe barande  
 Sara from this<sub>OBL</sub> very ashamed is that in this competition winner  
 nashode]  
 NEG.became-3sg  
 'Sara is very embarrassed at the fact that she did not win the competition.'

PERSIAN

I believe this fact suggests that prepositions in Persian require DPs obligatorily, as Persian does not provide a morphological case system. This is also due to the fact that there is no preposition in Persian that can avoid nominalization of the embedded clause, see e.g. *dar \*(in) ke* 'in the fact that', *be \*(in) ke* 'to the fact that', *baraye \*(in) ke* 'for the fact that, in order to' (cf. in particular chapter 13 in Yousef (2018)). This also happens when prepositions surface after the matrix verb. See, for instance, the following example with the preposition *baraye* 'for' (adapted from Yousef (2018:295)):

- (149) Baraye \*(in) ke to ra bebinad be inja amad  
 for this that you OBJ see-3sg to here came-3sg  
 'He came here to see you.'

PERSIAN

In sum, the data from Persian suggest, in the first place, that all positions inside the matrix clause are positions that require nominalization, albeit for different reasons. Subjects need to have their D-feature checked by a DP, the direct object has a morphological marker that requires nominalization, and prepositions can only be merged with a DP complement. It is possible, therefore, that languages without a morphological case system can be oriented towards two main opposite directions: they can make nominalization mostly optional as in Swedish, or mandatory *in situ*, as in Persian. As for clausal subjects, D-feature checking remains a key factor in obligatory nominalization of Persian sentential subjects (like in Icelandic as well as in Swedish and German). However, non-NomCIs can still emerge if embedded clauses escape canonical argument positions via ex-

trapolation.

#### 4.5.5 Russian

The last language we are going to discuss in our cross-linguistic analysis is Russian, a Slavic language which does not display a V2 structure but has a very rich inflectional system. The case system of Russian is composed of six cases (see e.g. Bailyn (2012:123)): nominative, accusative, dative, genitive, instrumental and prepositional. Based on our observations on Icelandic and German, we would expect to see a difference between structural and lexical case where instances of the latter trigger clausal nominalization while the former does not correspond to a case feature to check. Secondly, in relation to clausal subjects, we also expect non-NomCIs to be allowed to emerge outside Spec,T. However, the distribution of NomCIs and non-NomCIs in Russian is more complex than in the languages we have observed in this section. Let us discuss, first of all, the distribution of nominative clausal subjects. We have already seen in Knyazev (2016) that clausal subjects need to be preceded by a CN in the first position, as shown in the following example (see §2.4.2):

- (150) \***(To)**, čto on èto skazal, dokazyvaet ego nevinovnost  
that<sub>NOM</sub> that he<sub>NOM</sub> this<sub>ACC</sub> said proves his innocence<sub>ACC</sub>  
'The fact that he said this proves his innocence.' (Knyazev 2016:23) RUSSIAN

However, when a nominative clausal subject is low, it displays an optional CN. Compare, for instance, the following examples. In the first, the CN *to* 'that' in nominative case is optional, but, in the second, it becomes mandatory if it is moved to first position:<sup>41</sup>

- (151) a. Menja udivljaet (**to**), čto on priexal  
me<sub>ACC</sub> surprises that<sub>NOM</sub> that he came  
'The fact that he came surprises me.' (adapted from Knyazev 2016:23, 25)  
b. \***(To)**, čto on priexal, menja ne udivljaet  
that<sub>NOM</sub> that he came me<sub>ACC</sub> not surprises  
'That he came does not surprises me.' (Knyazev 2016:40, also footnote 19) RUSSIAN

These data align with the ones from Icelandic: if the subject lands into a position that is not Spec,T, it is possible to drop the pronoun. This issue could be apparently explained by the fact that the argument that moves to Spec,T must check the D-feature in T, but there is a major complication in this regard: explaining mandatory nominalization by D-feature checking is based on the assumption that NomCIs and non-NomCIs are DPs and CPs respectively at the surface (as I have proposed for Icelandic). But in §2.4.2, we have seen that non-NomCIs can also emerge as DPs in Russian, because a) extraction from a clausal argument is not allowed (see Dyakonova

<sup>41</sup>These examples from Knyazev only show the negation *ne* 'not' with the variant containing a NomCI subject. There was no specific discussion in his dissertation about the use of the negation in these examples, so I assume that they are still comparable examples.

(2009)) and b) non-NomClIs as clausal complements of verbs cannot even emerge if accusative is not assigned (this depends on whether the subject is an agent or not). In these conditions, the fact that *to* must surface even if NomClIs and non-NomClIs are both DPs is explained by Knyazev (2016) with the idea that the head of the DP emerging in Spec,T must be overt, in line with Landau (2007). This is an important difference between Icelandic and Russian. Consequently, the difference between NomClIs and non-NomClIs in structural case positions in Russian is determined by case visibility or phonological realization of the D-head. However, despite this structural difference between Icelandic and Russian, it is still a fact that it is not nominative case that triggers the presence of a CN, since embedded clauses that are not introduced by *to* can emerge outside Spec,T. At least this point is in line with our expectations.

Regarding post-copular clauses, CNs are also optional in Russian, which confirms the idea that nominative case does not trigger clausal nominalization. This is illustrated in the following example:

- (152) Fakt (**tot**), čto Lenin v samom dele ochen' mnogim obyazan finnam  
 fact<sub>M-SG</sub> that<sub>M-SG</sub> that Lenin actually very much owe Finns  
 'The fact is that Lenin owes very much to the Finns.'  
 (from the *Russian National Corpus*) RUSSIAN

In relation to clausal objects in structural case, accusative is hardly visible, as shown in the example below. This is due to the fact that a non-NomCl as DP can already be licensed in an accusative case position (see also §2.4.2; §2.5.5). In other words, nominative and accusative do not trigger the occurrence of *to* (in a stricter way than in Icelandic, as the CN is not optional):

- (153) Maša skazala (?\*to), čto ona ustala  
 Masha<sub>NOM</sub> said that<sub>ACC</sub> that she<sub>NOM</sub> tired  
 'Masha said that she is tired.' (Knyazev 2016:25) RUSSIAN

At this point, the question is how lexical case emerges in Russian and whether it differs from structural case when it comes to clausal nominalization. In this particular regard, agentivity has an important impact on the distribution of NomClIs and non-NomClIs. When the subject is an agent, the CN in lexical case is surprisingly optional. By contrast, if the subject is not an agent, then nominalization is mandatory:

- (154) a. Načal'nik grozit Maše (**tem**), čto ona budet uvolena  
 boss<sub>NOM</sub> threatens Masha that<sub>INS</sub> that she will be fired  
 'The boss threatens Masha with dismissal.' (Knyazev 2016:64)  
 b. Opozdanie grozit Maše \*(**tem**), čto ona budet uvolena  
 being late threatens Masha that<sub>INS</sub> that she will be fired  
 'Being late threatens Masha with being fired.' (Knyazev 2016:3) RUSSIAN

In order to explain this contrast, we need to add more information on Knyazev's proposal. As we have just discussed in relation to direct objects, a non-NomCl as DP can be licensed if the subject is an agent and the clause receives accusative case. However, there is a further alternative

condition for a non-NomCI to be licensed as DP in Russian (Knyazev 2016:5):

(155) *The Case requirement of čto-clauses*

A *čto*-clause complement has to be licensed by Case. This can be realized in one of the following two ways:

- a. the sentential complement is assigned structural Case;
- b. the sentential complement is licensed by (a possibly silent) P.

The condition in question is (b). When the subject is an agent, non-NomCI complements can be licensed with verbs like *grozit'* 'threaten', see (154). However, they are not licensed by virtue of accusative case, as non-clausal DPs in accusative case are disallowed, as shown in the following example (Knyazev 2016:64):

- (156) \*Načal'nik grozit Maše uvol'nenie  
 boss<sub>NOM</sub> threatens Masha<sub>DAT</sub> dismissal<sub>ACC</sub>  
 Intended: 'The boss threatens Masha with dismissal.' RUSSIAN

Therefore, Knyazev proposed that, with verbs like *grozit'*, the clausal complement is embedded in a PP where the preposition is not phonetically realized. If we assume that this is a correct analysis of Russian, then the only instances of lexical case we have observed here are the ones with a NomCI that is assigned instrumental case (in particular (154b)), which indicates that lexical case triggers mandatory nominalization as we expected. The other construction with a silent P (where non-NomCIs are able to emerge) is actually subsumed within prepositional objects rather than bare arguments of the verb. Now, if we now take a closer look at prepositional objects, we can find a clear distinction between overt prepositions, where nominalization is required (see §2.4.2; §2.5.5) and silent prepositions, with which clausal DPs with a silent CN can emerge (as we have seen in (154) in the agentive construction). Examples of an overt preposition are shown below:

- (157) a. Maša nadeetsja na \*(to), čto u nee budet mnogo svobodnogo vremeni  
 Masha<sub>NOM</sub> hopes on that<sub>ACC</sub> that at her<sub>GEN</sub> will be a lot free time<sub>GEN</sub>  
 'Masha hopes that she will have a lot of free time.' (Knyazev 2016:22)
- b. Vanja nastajvaet na \*(tom), čto žit' v Moskve deševu  
 Vanya<sub>NOM</sub> insists on that<sub>PREP</sub> that to live in Moscow<sub>PREP</sub> cheap  
 'Vanya insists that it is cheap to live in Moscow.' (Knyazev 2016:22) RUSSIAN

In this regard, prepositional objects in Russian are more systematic than in Icelandic. While in Icelandic the distribution of *það* is differentiated based on the relevant case, Russian applies a general rule where overt prepositions are followed by a NomCI, independently of case (consider, for example, accusative case in (157a)), while silent prepositions do not require the CN to emerge. What is more, given that the clausal complement does not show any case morphology when it is preceded by a silent P (as *to* does not emerge), there might be a possibility that the case assigned by the covert preposition is some sort of a default case.

In sum, the data from Russian differ to some extent from the Icelandic data in relation to the

trigger of mandatory nominalization of subjects as well as the distribution of NomCl and non-NomCl in prepositional objects, due to the fact that both variants emerge as DPs. However, it is also true that we found some important similarities: structural case does not trigger clausal nominalization whereas lexical case does. This confirms our assumption that lexical case can only trigger nominalization in languages that display a morphological case system.

## 4.6 Summary

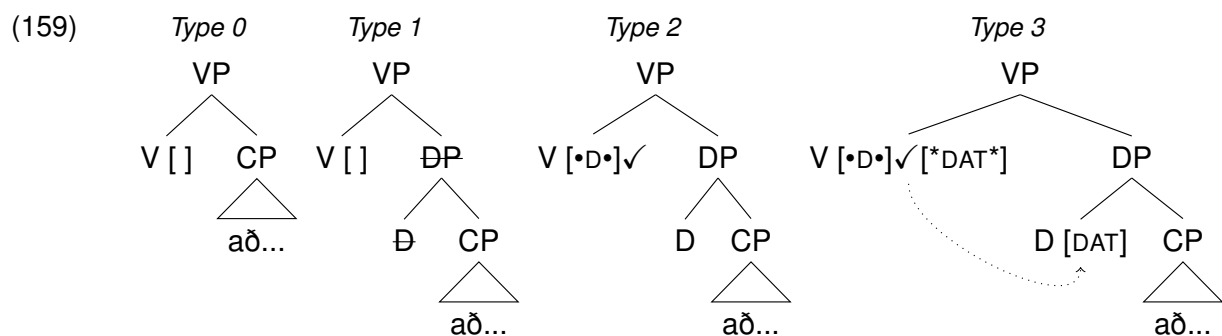
In this chapter, I have discussed some important theoretical as well as cross-linguistic implications on the basis of the *Feature Checking Hypothesis* I presented in Chapter 3. Our analysis of clausal subjects led us to the conclusion that NomCl and non-NomCl are both merged as DPs in Icelandic. However, a clausal DP undergoes structural removal if *það* does not check any feature, yielding a non-NomCl. Following the notation proposed in Heck and Müller (2007), where bullet [ $\bullet$ D $\bullet$ ] features cause Merge of a DP, I proposed that FCH is a mechanism where *það* emerges whenever a feature needs to be checked by a clausal complement. This is displayed in a refined version of FCH, which I called *Clausal Nominalization Mechanism*:

(158) *Clausal Nominalization Mechanism (CNM)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and can only surface if (it is enough to satisfy one condition):

- a. CN checks at least one D-,  $\varphi$ - or case feature, or
- b. the clause merges with an item bearing a [ $\bullet$ D $\bullet$ ] feature

We also came to the conclusion that four main possible configurations exist that account for clausal complementation. *Type 0* is the only configuration valid for non-DP positions, where CPs are merged as such (as with clausal complements of the verbs *hjálpa* ‘help’ and *þvinga* ‘force’; see Chapter 2). The other three, *Type 1*, *Type 2* and *Type 3*, cover DP positions. They are illustrated below once again for the ease of the reader (the tree diagrams show an example of a clausal complement selected by a verb, but these configurations can also be applied to other syntactic positions):



*Type 1* is the one involving structural removal when *það* has no feature to check locally. *Type 2* is the configuration where the DP-shell remains by virtue of the [ $\bullet$ D $\bullet$ ] feature, which triggers Merge of a DP. *Type 3* covers those instances where *það* checks at least one case, gender,

number or D-feature and the DP-shell still remains. These configurations often compete in the same position, e.g. *Type 1* and *Type 2* in structural case positions (except Spec,T), and *Type 1* and *Type 3* with clausal complements that are assigned dative or genitive. *Type 0* can also compete with other configurations in instances of stylistic fronting, copula+adjective constructions and passivized constructions with extraposition (see in particular §4.3.7; §4.3.8).

Regarding case assignment, the data from Icelandic indicate that, while lexical case corresponds to a feature to check in syntax, structural case does not, as non-NomCIs are preferable in those instances. This indicates that nominative and accusative are assigned in a different fashion, and aligns with configurational approaches like *Dependent Case Theory* (see e.g. Marantz (2000) or a preliminary discussion in Yip et al. (1987)). Within DCT, nominative is the morphological manifestation of unmarked case while accusative is the product of a relationship between DPs, i.e. dependent case (see §4.4.4).

In relation to clausal nominalization outside canonical argument positions, there is evidence that points to the possibility that accusative represents the default case in prepositional complements in Icelandic (see §4.4.5). This is supported by the fact that adverbs that are at times able to take up the role of prepositions like *upp* 'up', *niður* 'down' are followed by an accusative complement. Moreover, other prepositions like *innan* 'within', which generally assign genitive case, also tend to be followed by accusative. The data from Faroese, which is the closest language to Icelandic, also shows that accusative is appearing in PPs in those instances where genitive is not productive anymore and is only used in fixed expressions.

As for postcopular clauses, structural removal is only preferable when the subject is neuter singular. This suggests that neuter and singular do not correspond to features to check in syntax, similarly to nominative and accusative case (also cf. Garofalo (2023)). This hypothesis is also supported by instances of a lack of agreement between predicative adjectives and non-nominative subjects as well as with coordinated subjects involving abstract nouns or clauses (see §4.4.6). Furthermore, from a structural perspective, I have also shown that clauses introduced by *sá* and *sú* are structured slightly differently from the rest of NomCIs (i.e. as a variant of *Type 3*), due to the fact that these CNs can be located in the specifier position of a silent DP projection. This allows them to be promoted to Spec,T leaving the rest of the constituent behind. This is a property that other forms of CNs do not have due to the fact that they are homophonous with personal pronouns, which could cause them to be merged as D-heads instead (see §4.3.9).

Lastly, cross-linguistic data (see §4.5) have confirmed that structural case does not correspond to features to check in syntax, similarly to what happens in Icelandic: nominative and accusative case, in fact, do not trigger clausal nominalization. By contrast, lexical case triggers clausal nominalization in those languages that display a morphological case system (like German and Russian). On the other hand, by observing Persian and Swedish, we have discovered that the lack of a morphological case system can have a notable impact on clausal nominalization, which can become optional as a general rule across syntactic positions (as in the case of Swedish, apart from Spec,T) or mandatory across syntactic positions when extraposition is not

involved (as in Persian). Overall, cross-linguistic data have also shown that Spec,T is a position in which clausal subjects must be nominalized. Non-NomCIs, however, can still emerge if the clausal subject is able to escape Spec,T.



## 5 Conclusions

We have reached the final chapter of this dissertation, in which I am going to summarize the results on clausal nominalization that I discussed in the previous chapters. This work has offered the most extensive investigation of clausal nominalization in contemporary Icelandic, as it encompasses all syntactic positions where NomCIs can surface. So far, in the linguistic literature, the phenomenon has been analyzed to some extent, but the existing accounts only dealt with subjects, direct objects, postcopular clauses and prepositional complements (see Thráinsson (1979) in the case of subjects and objects; Garofalo (2015; 2020; 2023)) or just other aspects related to clausal nominalization like movement or case-marking (see e.g. Ingason (2018); Ott (2014); Wood (2012)).

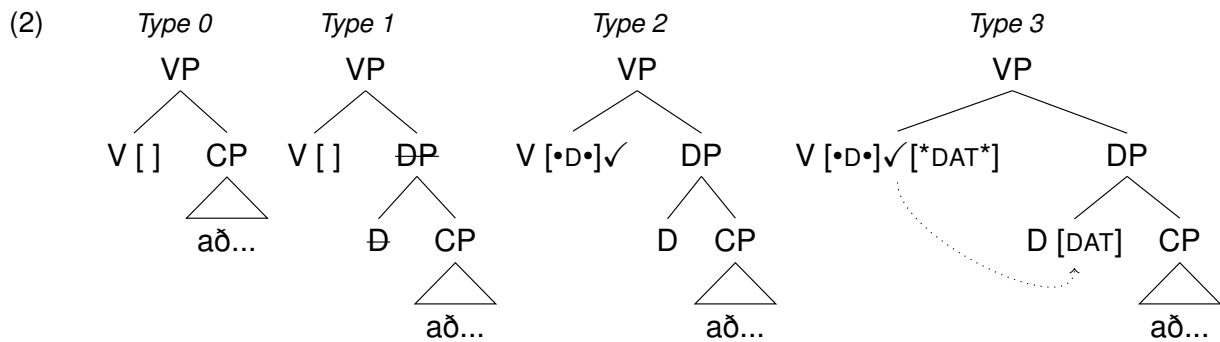
By including all the relevant syntactic positions where NomCIs can surface, the results presented provide a more accurate picture of the actual trigger and the distribution of clausal nominalization compared to previous accounts. The purpose of *það*, as discussed in this dissertation, is to check features (case, gender, number, etc.) that a CP is unable to check. This is clearly described in CNM, illustrated below (see §4.3.7):

(1) *Clausal Nominalization Mechanism (CNM)*

A CN surfaces to check syntactic features for the CP it introduces. Its distribution is restricted to DP positions and can only surface if (it is enough to satisfy one condition):

- a. CN checks at least one D-,  $\varphi$ - or case feature, or
- b. the clause merges with an item bearing a [ $\bullet$ D $\bullet$ ] feature

I argued that NomCIs and non-NomCIs are both merged as DPs in DP positions, although CPs can only merge as such in non-DP positions (see §2.4; §4.3.5). However, non-NomCIs undergo a process of removal of the DP projection (see §4.3.5), which causes them to emerge in the end as CPs. In light of this hypothesis, I argued for the possible existence of four main configurations which account for the complex distribution of NomCIs and non-NomCIs in Icelandic. One of these configurations covers non-DP positions where CPs are merged as such (*Type 0*), while the other three account for DP positions and differ on the basis of whether a feature must be checked by the clausal argument/complement or not. In absence of a feature to check (*Type 1*), the DP shell is removed. This is more clearly illustrated below (see also §4.3.7). The tree diagrams display the case of an embedded clause selected by a verb, but it applies in a similar way to other syntactic positions as well:



In relation to case assignment in particular, the empirical data have clearly shown a contrast between structural case and lexical case in Icelandic, in which only the latter appears to trigger clausal nominalization. These results suggest that nominative and accusative do not correspond to formal features that need to be checked in narrow syntax (see §4.4.4). This contrast, as we have seen, can be accounted for by configurational theories of case like *Dependent Case Theory* (Yip et al. (1987); Marantz (2000); Preminger (2011) among others), where nominative and accusative are realized at PF: accusative is the morphological manifestation of dependent case, i.e. of a relationship between two DPs that are yet to receive case, and nominative is the morphological manifestation of unmarked case (see §4.4.4).

Furthermore, I have argued that accusative case could be a default case in prepositional objects (see §4.4.5). The data from clausal nominalization show that non-NomCIs are systematically preferable with clausal complements of functional accusative-assigning prepositions, similarly to what happens with accusative case in the direct object position. There is some independent evidence in support of the idea that accusative is a default case in Icelandic PPs: a) accusative rather than dative is adopted by adverbs like *upp* ‘up’ or *út* ‘out’ when they play the role of prepositions; b) prepositions like *innan* ‘within’, which normally assign genitive case, can also assign accusative, while dative is not felicitous; and c) Faroese, which is the closest language to Icelandic, has taken up accusative with former genitive-assigning prepositions, which corroborates the possibility that also Icelandic accusative case is default in PPs (see §4.4.5).

In relation to post-copular clauses, the data from clausal nominalization suggest that neuter and singular also do not correspond to formal features to check in narrow syntax, based on the fact that non-NomCIs are only preferable when the subject with which the CN agrees is neuter singular (cf. Garofalo (2023)). Further data about gender and number agreement with coordinate subjects as well as the lack of agreement between non-nominative subjects and predicative adjectives corroborate this possibility (see §4.3.9). From a structural perspective, however, we found out that the mandatory extraposition constructions with CNs like *sú* (§2.2.4) were only apparently instances of extraposition (if we intend it as right-adjunction). I have shown that NomCIs introduced by *sá* and *sú* are structured slightly differently from the other NomCIs (i.e. as a variant of *Type 3*), in the sense that these CNs are most likely located in the specifier position of a silent DP projection rather than in the D-head (cf. e.g. Brugé (2002) in relation to the position of demonstratives). This allows them to be promoted to Spec,T leaving the rest of the constituent behind.

Other forms of CNs do not have this property, plausibly because they are merged as D-heads due to their homophony with personal pronouns (see §4.3.9).

Overall, my account on clausal nominalization offers a view of *það* as CN that notably differs from other analyses in the past literature on the topic. In particular, contrary to Thráinsson (1979), who mainly attributed the use of CN *það* before clausal subjects and objects to factivity (see §2.5.2), the data I presented in this dissertation have shown that clausal nominalization mainly depends on syntax and, more specifically, feature checking rather than semantic factors. In light of the findings in §4.3.5, my analysis partially differs from Thráinsson's idea that NomClIs and non-NomClIs are NPs (see §2.4.2), as it also includes structural removal and accounts for positions where CPs must be merged as such (i.e. non-DP positions). The account I presented also argues against bi-clausal analyses of clausal nominalization like Garofalo (2015) (also Ott (2014)) as *það* and the CP clearly form together one constituent (see §2.5.6.1), and differs from analyses like Ingason (2018) in that clausal complements are not directly case-marked, but rather require *það* if a formal case feature is assigned (see §2.5.5).

The hypothesis I presented in my dissertation has strong empirical underpinnings as it is corroborated by two main types of data, i.e. corpora and questionnaires. The use of both has proven to be effective to map the distribution of NomClIs and non-NomClIs, especially by virtue of the fact that the results show a high degree of systematicity. Canonical argument positions (except the case of non-extraposed clausal subjects in Spec,T) tend to disprefer *það* (§3.3.1; §3.3.2.2), while lexical case positions tend to highly prefer it or to require it (§3.3.2.3). Also prepositional complements showed a systematic distinction between accusative and dative case which mirrors to a great extent the results from canonical argument positions (§3.3.4). As for post-copular clauses, the data from RMH offer important insights into the distinction between *það* in neuter singular and the other CNs with a different gender or number (§3.3.3). Without either set of data, it would have been impossible to correctly map the distribution of NomClIs and non-NomClIs. In the case of indirect objects, on the other hand, interviews were crucial to determine that Spec,Appl is a position in which nominalization must occur, although a passivized accusative clausal indirect object can still emerge as a non-NomCl if it is moved to Spec,C (see §3.3.7).

As for the impact of this dissertation on the discussion on clausal nominalization from a cross-linguistic perspective, the comparison between Icelandic and Swedish, German, Persian and Russian (§4.5) has shown that lexical case only triggers nominalization when a language has a rich morphological case system (this is the case of Icelandic, German and Russian; see §4.5.3; §4.5.5). As for structural case, the languages I studied indicate that nominative and accusative case do not trigger clausal nominalization, independently of whether the language provides a morphological case system or not. However, a requirement for a DP (independently of case assignment) can affect the distribution of NomClIs and non-NomClIs in structural case positions. This could be observed, in particular, by comparing the data from Persian and Swedish, which do not have a morphological case system. While Swedish tends to allow clausal nominalization to be optional (see §4.5.2), Persian essentially requires DPs independently of case, unless the clause is

extraposed (see §4.5.4). However, across all the languages studied, nominalization is mandatory in Spec,T except when the sentential subject escapes that position. In that case, non-NomCIs are possible. In other words, D-feature checking is, cross-linguistically speaking, a possible trigger of clausal nominalization. In this regard, however, Russian is an exception as non-NomCIs also emerge as DPs (see §2.4.2; §4.5.5), which suggests that a CN must surface in subject position by virtue of the fact that the D-head of the subject must be visible (see Knyazev (2016); cf. Landau (2007)).

Although the hypothesis I presented in this dissertation is supported by a wide sample of empirical data, there are still some issues that have remained mysterious and that I have left open for further research. The first important issue worth mentioning is the impossibility of analyzing intra-speaker variation (see §3.2.1). Given the extension of clausal nominalization across syntactic positions, it would have been very difficult to assess the whole distribution of NomCIs and non-NomCIs by administering multiple questionnaires to a specific set of participants, at the risk of not gathering enough data if participants had not been available to answer questionnaires at a certain point in my research. Therefore, as already mentioned in §3.2.1, I only focused on inter-speaker variation in order to get at least a general picture of how non-NomCIs and NomCIs are distributed across syntactic positions and which variant is preferable in a specific position. Further research can assess how *það* is distributed in the grammar of single individuals and whether the existence of the configurations I have proposed is also confirmed in this regard.

A further important issue that is left open for further research is a possible counterargument against the idea that accusative case is a default case in PPs. Various nouns can subcategorize for a PP where the preposition *á* ‘on’ assigns dative. This structure is highly common and suggests that the default case for Icelandic might be dative rather than accusative, contrary to what I propose in this dissertation. I am not aware of why it is dative that emerges in these constructions rather than accusative, but, as already shown in §4.4.5, clausal nominalization with *á* and accusative vs. dative case shows once again that non-NomCIs are preferable when accusative case is assigned, but not when dative is assigned, which mirrors the idea that accusative might not be checked as a formal case feature. Further research can assess why it is dative that emerges in these constructions, provided that accusative is the default case in PPs.

# A Data from corpora

## A.1 Introduction

This appendix contains additional information on the data retrieved from RMH presented in Chapter 3 and explains how special queries have been constructed. It also describes how the data have been presented in their relevant online spreadsheets, which can be found on Github (the following link also contains the spreadsheets on questionnaires and interviews, see Appendix B). The name of the repository is *IceNomCl* (Icelandic Nominalized Clauses). The folder for the data from RMH is called *Corpora*:

(1) <https://github.com/mirkogarofalo/icenomcl/>

## A.2 Further information on methodology

### A.2.1 Special queries for clausal subjects

The queries for clausal subjects were more difficult to construct compared to the others (a general description of how queries have been constructed is offered in §3.2.3), both in the case of clausal subjects in the first position and after the finite verb. This was mainly due to the fact that a clausal subject is not preceded by a lexical item selecting it.

Regarding nominative clausal subjects in first position, in order to look for examples, I had to set the conjunction/infinitive marker or the interrogative pronoun first (or the CN in the case of NomClIs). Moreover, I also had to set the query in such a way that only instances of *að* (infinitive marker or conjunction) and interrogative pronouns that had a capital letter would be counted, which would be a signal that a new sentence is starting. In order to do this, while classifying the first item of the query as a conjunction, infinitive marker or interrogative pronoun in the case of non-NomCl variants, I also specified as word form the regular expression `[A-Z] . * . $`, which means that the first character of the word form should be a capital letter, no matter what the next letters are and no matter how long the word is.<sup>1</sup>

However, setting exclusively *að* or the interrogative pronoun into the query to look for instances of clausal subjects in the first position is not enough. By setting, for instance, *að* only (as a conjunction or infinitive marker), we could gather a very high number of false positives, which might often be clausal complements that belong to a previous utterance but where the conjunction

---

<sup>1</sup>The reader might object to this method by pointing out that it is possible via RMH to set the query in such a way that the first item to look for must be at the beginning of a sentence (instead of resorting to a regular expression). The problem is that, while trying to do this, I gathered surprisingly very few examples. This could be probably due to a bug in the system at the time (in 2020). By using a regular expression instead of following the default way provided by RMH, many more examples could be retrieved.

happens to start with a capital letter. The same applies to interrogative pronouns, as the query would look for sentences where the interrogative pronoun has been moved to first position via *wh*-movement (as in matrix clauses), which is definitely not what we are looking for. In order to solve these issues, I set the query in the following way. Let us consider the following examples:

- (2)
- a. Að læra íslensku er skemmtilegt  
to learn Icelandic is fun  
'Learning Icelandic is fun.'
  - b. Að hann borðaði pítsuna er ólíklegt  
that he ate pizza.the is unlikely  
'It is unlikely that he ate the pizza.'
  - c. Hvernig hann týndi lyklunum er hlægilegt  
how he lost keys.the is ridiculous  
'It's ridiculous how he lost his keys.'

The structure of these examples is very simple and minimal. In the case of an infinitive subject, we have to expect there to be *að*, then the infinitive verb and a direct object, perhaps a negation as well, and then the finite matrix verb. In the case of *that*-clauses and embedded questions, we also have to consider the presence of a subject inside the clause. Based on these indications, I built a set of queries to look for structures similar to the ones in (2), adding as well some flexibility in order to gather some more complex structures (although this could also open the way for some false positives to show up, but one can exclude false positives manually later on provided that they are not too many). Therefore, I applied the following queries to clausal subjects in first position:

- (3)
- a. *að* (infinitive marker) - infinitive verb - 1 to 2 words in-between (where the direct object and maybe the negation can surface) - finite matrix verb
  - b. *að* (conjunction) - 1 to 2 words in-between (where the subject of the embedded clause can surface) - finite verb - 1 to 2 words in-between (for any direct object and perhaps negation) - finite matrix verb
  - c. Interrogative pronoun - 1 to 2 words in-between (where the subject of the embedded clause can surface) - finite verb - 1 to 2 words in-between (for any direct object and perhaps negation) - finite matrix verb

As for nominative subjects after the finite verb, I based the relevant queries on (3), making sure that the finite matrix verb comes first in the sequence this time. I also applied the regular expression `[A-Z].*.$` onto the verb since it was the first item of the query. However, I gathered many false positives (in total, 14684 tokens were downloaded but about 99% thereof were not valid). This is most probably due to the fact that the presence of both *það* and the clausal argument in subject position is not so common. The remaining examples, however, were valid.

As for clausal subjects in lexical case, which are more difficult to find, I focused specifically on the predicate *verða vart* 'be noticed', which takes a subject in genitive case (see also §3.3.1.2). This time, I could only look for tokens with this predicate by applying extraposition, as I could find

no example at all by letting the sentence start with *þess* + CP. Moreover, only *that*-clauses can be clausal subjects with this predicate. Therefore, I constructed the queries in the following way:

- (4) a. *þess* + *verða* + *vart* + *að* (conjunction) > for first position (NomCI)
- b. *það* + *verða* + *vart* + *að* (conjunction) > for first position (non-NomCI)
- (5) a. *verða* + *þess* + *vart* + *að* (conjunction) > for Spec,T (NomCI)
- b. *verða* + *vart* + *að* (conjunction) > for Spec,T (non-NomCI)

In the case of genitive subjects in the first position, I alternated the NomCI extraposed variant with a non-NomCI extraposed variant, which involves expletive *það*. As for genitive subjects in Spec,T, I did not adopt the expletive with extraposed bare clausal subjects due to the fact that expletive *það* is not possible after the finite verb.

## A.2.2 Exclusions from the analysis

I will enumerate here the restrictions I applied to my observations in RMH after I retrieved all the necessary data. Firstly, in relation to argument positions, I only excluded clausal indirect objects from my analysis of RMH, as they are quite rare and generally very difficult to find among naturalistic data considering how RMH queries are structured.

Secondly, in relation to the text categories contained in RMH, I excluded some of them from the query. The texts I did not take into account in my analysis all belong to the category called *Stjórnýsla*, which contains law texts and political speeches. Although we find examples in the linguistic literature on Icelandic where political speeches have turned out to be fruitful in the analysis of syntactic variation (see in particular Stefánsdóttir and Ingason (2018) for Icelandic), it is also true that these types of text tend to adopt a highly formal linguistic register, which might not reflect the typical distribution of non-NomCIs and NomCIs that one would commonly find in linguistic production within society in general. Hence, there is a possibility that including this category would distort to some extent the actual picture of clausal nominalization.

Thirdly, based on how queries are constructed on RMH, it was challenging to create queries where extraposed clauses emerge, especially because the number of words between *það* and the CP may vary and can lead to an enormous quantity of false positives. Therefore, to simplify things, clausal extraposition was not targeted in the queries, unless there was no room for ambiguity and the extraposed construction could be easily found in RMH (as I did with genitive subjects with the verb *verða* *vart*). However, if examples of extraposition incidentally showed up among the examples retrieved and were both valid and non-ambiguous examples of one of the two variants, they were taken into account in the analysis (see in particular §3.3.1 and §3.3.1.2 for instances of passivized extraposed clausal objects).

Fourthly, I expected other tokens downloaded from RMH to also be false positives of various kinds. I applied the following criteria to exclude them from the analysis, especially considering the fact that many invalid examples of this kind emerged across datasets:

- A token is invalid if it is impossible for the relevant clause to be preceded by *það* due to syntactic restrictions that have nothing to do with clausal nominalization
- A token is invalid if the clausal argument is not at all the argument type we are observing (e.g. an alleged clausal subject is not a subject at all or not even a clausal argument)
- A token is invalid if it is unclear or ambiguous
- A token is invalid if there is another token in the same dataset that includes the same relevant sentence (for instance, a quote containing clausal nominalization that is published in two or more articles of different newspapers) or if the example is simply a duplicate

For the ease of the reader, we will take a look at some examples where these criteria have been applied, apart from the fourth criterion, which is quite obvious. Here are three examples extracted from the *spá* dataset:

- (6) a. Það hefur því orðið mikil umbreyting í efnahagslífinu eins og við  
 there has therefore become great transformation in financial life as we  
 höfðum reyndar spáð [**því**] að yrði  
 had actually predicted that<sub>DAT</sub> that would become  
 ‘There has been therefore a great transformation in financial life as we had already  
 predicted there would have been.’
- b. En það eru fleiri en Þorvaldur sem hafa spáð **að undanfögnu**  
 but there are more than Þorvaldur that have predicted previously  
 ‘There are more people than Þorvaldur that have predicted [it] previously.’
- c. **Því** hefur verið spáð að þau séu núna við það að ná hámarki  
 ? has been predicted that they are now at that<sub>ACC</sub> to reach limit  
 ‘(Therefore?) it has been predicted that they are now reaching the limit.’

In the first example, the pronoun *því* in dative case can never precede the clausal argument. This is not caused by the matrix verb, but by the fact that NomCIs are islands as discussed in Chapter 2 (the DP *mikil umbreyting* ‘great transformation’ cannot cross the DP projection where *það* is positioned). The second example is invalid as *að undanfögnu* ‘previously’ is simply a PP adjunct and not a clausal complement, while the third example is invalid because *því* can be a CN but also an adverb.

### A.2.3 Values attributed to each token

After excluding all the relevant items based on the criteria described in §A.2.2, I organized the data into different spreadsheets.<sup>2</sup> In each of them, I assigned various values to every single token as shown in the figure below (see columns S to Y):

<sup>2</sup>The spreadsheets are not directly editable online, in order to safeguard the integrity of the analysis, but the reader can download them from the repository on Github and edit them separately for further tests.

|       | D            | E                                                          | S     | T   | U    | V    | W      | X     | Y              |
|-------|--------------|------------------------------------------------------------|-------|-----|------|------|--------|-------|----------------|
| 1     | match        | right_context                                              | valid | það | verb | case | clause | voice | token from     |
| 12430 | spáir því að | jafntefli veriði niðurstaðan . Þetta eru tveir þjálfarar s | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12431 | spáir því að | jafnvægi náist á markaðinum eftir 2 til 3 ár . Þegar að    | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12432 | spáir því að | jafnvægi náist á milli framboðs og eftirspurnar á olíun    | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12433 | spáir því að | jafnvægi náist ekki fyrr en árið 2012 . Karl velti upp ýr  | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12434 | spáir því að | jafnvægi náist í 130 til 135 vísitölustigum . EFTIR að     | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12435 | spáir því að | jafnvægi verði náð í búferlafutningum árið 2012 eftir      | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12436 | spáði því að | jafnvel 20.000 friðargæsluliðar á vegum S.Þ. myndu         | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12437 | spá því að   | jafnvel gæti orðið skortur á nautakjöti þegar liði á árið  | 0     |     |      |      |        |       | spá því að+THT |
| 12438 | spá því að   | jafnvel gæti orðið skortur á nautakjöti þegar liði á árið  | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12439 | spáir því að | jafnvel hún gæti náð flugi vegna þess aðhaldsleysis s      | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12440 | spá því að   | jafnvel núverandi samsteypustjórn haldi velli eða hva      | 1     | 1   | spá  | d    | tht    | p     | spá því að+THT |
| 12441 | spáði því að | jafnvel þótt gjaldeyrishöftin yrðu afnumin í náinni fram   | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12442 | spáir því að | Jagúarbillinn muni eiga erfitt uppdráttar um helgina .     | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12443 | spáir því að | Jairo Riedewald verði óvænt í byrjunarliði gegn Íslan      | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |
| 12444 | spáir því að | Jafnvel gæti orðið skortur á nautakjöti þegar liði á árið  | 1     | 1   | spá  | d    | tht    | a     | spá því að+THT |

Figure A.1: Example of token list (from the *spá* dataset)

The following is a description of the variables presented in Figure A.1 with their respective values:

- *valid*: In this variable, I established the validity of the token. 1 is assigned if the example is valid, 0 if it is not valid. If a token is assigned 0, the following columns will receive no value at all (see e.g. row 12437 in Figure A.1). The criteria based on which a token is invalid are described in §A.2.2.
- *það*: In this variable, I assigned 1 if the token is an example of clausal nominalization, 0 if the pronoun does not surface.
- *verb/noun/prep/adj*: The value I assigned here is the lemma of the item that selects the clausal complement. In the picture above, *spá* is the verb selecting the clausal argument, therefore the value assigned is the infinitive of the verb itself.
- *case*: Here I specified the case assigned to the clause by applying the following values: *n* (nominative), *a* (accusative), *d* (dative) and *g* (genitive).
- *clause*: Here I declared the type of clausal complement I found in the token. Its values are *tht* (*that*-clause), *inf* (infinitive clause) or *que* (embedded question).
- *voice* (only for verbs taking a direct object and subjects): here I declared the voice of the verb that selects the clause. The values here are *a* (active), *p* (passive) and *m* (middle).<sup>3</sup> For impersonal passives like *verið er að spá* ‘it is predicted’ or *búið er að spá* ‘it has been predicted’, however, I still assigned *a* as value since the clausal object would surface in object position like in active voice.

<sup>3</sup>Data on middle voice are only presented in nominative subjects in first position and Spec,T. In these datasets, there are many matrix verbs as possible values of the variable *verb*, many of which are in middle voice. As for the other datasets, especially the ones related to clausal direct objects, I focused exclusively on active and passive voice, even in the case of *krefjast* ‘demand’, which clearly has a middle voice morphology. The reason why I opted for this method is that the verbs selecting direct objects that have been included in our analysis are generally only found in either active and passive constructions, also with *krefjast*.

- *token from*: This variable only signals the set of data the token has been downloaded from. In the picture above, *spá því að+THT* refers to examples of nominalized *that*-clauses of the verb *spá*. The values are generally like the following (the example below is for the adjective *háður* ‘dependent’):

- (7) *háður því að+THT* > nominalized *that*-clauses  
*háður því að+INF* > nominalized infinitives  
*háður því hv* > nominalized embedded questions  
*háður að+THT* > non-nominalized *that*-clauses  
*háður að+INF* > non-nominalized infinitives  
*háður hv* > non-nominalized embedded questions

Moreover, with verbs that take a dative or a genitive clausal direct object, i.e. *spá* ‘predict’, *fresta* ‘postpone’, *fagna* ‘rejoice, celebrate’, *krefjast* ‘demand’ and *sakna* ‘miss’, I also set an additional variable called *Spec, T*. This variable is specifically for tokens where the clausal object of these lexical items has been passivized and moved either to first position or to *Spec, T*, with extraposition of the CP involved as well (see §3.3.1.2). Three possible values have been assigned to this variable: 0 if the passivized clausal object surfaces in first position, 1 if it surfaces after the finite verb, x if the token is ambiguous (and, therefore, could not be presented in the data on passivized clausal objects). As for verbs assigning accusative case, i.e. *harma* ‘regret’, *gagnrýna* ‘criticize’ and *opinbera* ‘disclose, reveal’, the only instances of passivized objects I considered were the ones where the subject was supposed to surface after the finite verb and not before it (in first position, an extraposed CP would have made *það* ambiguous because it is homophonous with the expletive).

In the case of nominative subjects in first position, I also noted the matrix main verb involved, in order to assess whether there was any difference among verbs in relation to the frequency of *það*. Moreover, when the main matrix verb was *vera* ‘be’, I also took note on the spreadsheet of what type of constituent followed *vera*, i.e. whether it was a DP, a CP, an adjective, etc.

Lastly, all the other columns of the spreadsheet documents have been left unchanged after the CSV files have been imported, except for some spreadsheets which were too heavy. In that case, some columns containing additional (but not quite relevant) metadata have been deleted or emptied.

## A.2.4 Statistical summary

For each spreadsheet document I also created a statistical summary that presents all the major data. Here is an example of the interface the reader can see while examining a spreadsheet:<sup>4</sup>

<sup>4</sup>Note that the layout might differ depending on the application with which the spreadsheet is opened. The images presented below are from Google Sheets.



Figure A.2: Statistical summary (example from *opinbera*)

In the green box on the top left, there is a summary of all the items that have been retrieved from the queries. Let us take a closer look at an example of this summary, shown in Figure A.3, which contains the data on the clausal direct object of the verb *opinbera* 'reveal':

| Data                | Tokens | Valid | Invalid | Remaining | True tokens | Incorrect set |
|---------------------|--------|-------|---------|-----------|-------------|---------------|
| opinbera að+THT     | 392    | 337   | 55      | 0         | 326         | 11            |
| opinbera að+INF     | 18     | 11    | 7       | 0         | 11          | 0             |
| opinbera hv         | 223    | 182   | 41      | 0         | 181         | 1             |
| opinbera það að+TH  | 106    | 96    | 10      | 0         | 89          | 7             |
| opinbera það að+INI | 7      | 7     | 0       | 0         | 6           | 1             |
| opinbera það hv     | 3      | 3     | 0       | 0         | 3           | 0             |

Figure A.3: Green box - Statistical summary (example from *opinbera*)

The content of the box appears to be quite obvious as it represents a general summary on the data that have been retrieved, which tells us how many tokens have been downloaded and how many are valid or invalid. But perhaps only the column with the title *Incorrect set* can appear to be unclear. This column refers to certain tokens that are valid but that have been downloaded from the wrong subset. For instance, if a token is downloaded from the subset *opinbera að+INF* but its clausal argument, by closer examination, is actually a *that*-clause, the token is assigned the value *t.ht* in the variable *clause* as we have seen earlier. This discrepancy is automatically displayed in the column *Incorrect set* of the green box.

In the boxes on the bottom left, all the valid items are counted and classified based on four criteria:

- presence or absence of *það* in general (red box)
- presence or absence of *það* based on the voice (active, passive, middle) of the selecting category (light green box); this box is only present in the case of direct objects and subjects

- presence or absence of *það* based on clause type (blue box)
- presence or absence of *það* based on both clause type and voice of the selecting item (yellow box); this box is only present in the case of direct objects and subjects

The results shown in these four boxes are also represented graphically in the four graphs on the right side.

Lastly, I also put into the statistical summary a pink box next to the graphs. This box contains a list of all the sources the tokens have been downloaded from. Every source has a tick button which can be checked or unchecked in case one wants to take a look at whether certain sources tend to display more frequently than others instances of clausal nominalization. These data on the frequency of *það* based on sources are also shown in much detail in another tab I created in each spreadsheet. The tab's name is *Trends* and looks as shown in Figure A.4:

| Source                         | Trends (% of það) |         |         |         |         |         |
|--------------------------------|-------------------|---------|---------|---------|---------|---------|
|                                | general           | tht     | inf     | que     | active  | passive |
| <a href="#">433.pressan.is</a> | 25.00             | 50.00   | #DIV/0! | 0.00    | 20.00   | 33.33   |
| <a href="#">Andriki.is</a>     | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Bæjarins besta                 | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Bændablaðið                    | 0.00              | 0.00    | #DIV/0! | #DIV/0! | 0.00    | #DIV/0! |
| <a href="#">Bleikt.is</a>      | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Bónði.is                       | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Bylgjan                        | 50.00             | #DIV/0! | #DIV/0! | 50.00   | 50.00   | #DIV/0! |
| <a href="#">DV.is</a>          | 20.00             | 23.53   | 50.00   | 7.14    | 21.62   | 15.38   |
| Eiðfaxi                        | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Eyjafréttir.is                 | 0.00              | 0.00    | #DIV/0! | #DIV/0! | #DIV/0! | 0.00    |
| Eyjan                          | 16.67             | 33.33   | #DIV/0! | 0.00    | 16.67   | #DIV/0! |
| <a href="#">Eyjar.net</a>      | 0.00              | #DIV/0! | #DIV/0! | 0.00    | #DIV/0! | 0.00    |
| Fiskifréttir                   | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Fjarðarfréttir.is              | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Fjarðarpósturinn               | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Fótbolti.net                   | 28.81             | 37.35   | 40.00   | 3.33    | 35.39   | 8.62    |
| Fréttablaðið                   | 0.00              | 0.00    | #DIV/0! | 0.00    | 0.00    | 0.00    |
| Fréttatíminn                   | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Fréttavefur suðu               | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Heimur                         | 0.00              | 0.00    | #DIV/0! | #DIV/0! | 0.00    | #DIV/0! |
| Húnahornið                     | 0.00              | 0.00    | #DIV/0! | #DIV/0! | 0.00    | #DIV/0! |
| Jónas.is                       | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Kaffið.is                      | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
| Kjarninn                       | #DIV/0!           | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |

Figure A.4: Example of a *Trends* tab

The *Trends* tab is divided into six columns, which show for each source how the general distribution of *það* is, and also how it is on the basis of the clausal argument following as well as the voice.

### A.3 Data

The data from RMH are going to be presented here, including a breakdown for each dataset based on clause type.

### A.3.1 Subjects in first position

| GENERAL      |               | CLAUSE TYPE   |               |            |
|--------------|---------------|---------------|---------------|------------|
| <b>Valid</b> | 9478          | <i>tht</i>    | <i>inf</i>    | <i>que</i> |
| <b>-pað</b>  | 7047          | 628           | 6419          | 0          |
| <b>+pað</b>  | 2431          | 991           | 1440          | 0          |
| <b>%pað</b>  | <b>25.65%</b> | <b>61.21%</b> | <b>18.32%</b> | –          |

Table A.1: Nominative subjects in first position

| GENERAL         |                   |                         |                      |
|-----------------|-------------------|-------------------------|----------------------|
|                 | <i>verða vart</i> | <i>krefjast (pass.)</i> | <i>sakna (pass.)</i> |
| <b>Valid</b>    | 21                | 1422                    | 16                   |
| <b>-pað</b>     | 0                 | 5                       | 0                    |
| <b>+pað</b>     | 21                | 1417                    | 16                   |
| <b>%pað</b>     | <b>100.00%</b>    | <b>99.64%</b>           | <b>100.00%</b>       |
| CLAUSE TYPE     |                   |                         |                      |
| <i>tht -pað</i> | 0                 | 4                       | 0                    |
| <i>tht +pað</i> | 21                | 1408                    | 5                    |
| <i>tht %pað</i> | <b>100.00%</b>    | <b>99.71%</b>           | <b>100.00%</b>       |
| <i>inf -pað</i> | 0                 | 1                       | 0                    |
| <i>inf +pað</i> | 0                 | 9                       | 11                   |
| <i>inf %pað</i> | –                 | <b>90.00%</b>           | <b>100.00%</b>       |
| <i>que -pað</i> | 0                 | 0                       | 0                    |
| <i>que +pað</i> | 0                 | 0                       | 0                    |
| <i>que %pað</i> | –                 | –                       | –                    |

Table A.2: Genitive subjects in first position

| Verb                    | -pað | +pað | %pað          |
|-------------------------|------|------|---------------|
| <i>vera</i> 'be'        | 4232 | 963  | <b>18.54%</b> |
| <i>pýða</i> 'mean'      | 293  | 209  | <b>41.63%</b> |
| <i>hafa</i> 'have'      | 164  | 121  | <b>42.46%</b> |
| <i>gera</i> 'do'        | 141  | 116  | <b>45.14%</b> |
| <i>gefa</i> 'give'      | 132  | 68   | <b>34.00%</b> |
| <i>koma</i> 'come'      | 95   | 60   | <b>38.71%</b> |
| <i>skipta</i> 'change'  | 94   | 34   | <b>26.56%</b> |
| <i>taka</i> 'take'      | 97   | 22   | <b>18.49%</b> |
| <i>fela</i> 'hide'      | 87   | 32   | <b>26.89%</b> |
| <i>snúast</i> 'revolve' | 82   | 29   | <b>26.13%</b> |

Table A.3: Most common matrix verbs with a nominative clausal subject in first position

| Constituent                                | -það | +það | Total | %það          |
|--------------------------------------------|------|------|-------|---------------|
| <i>jafn</i> + adj. + <i>og</i> 'as...as'   | 22   | 9    | 31    | <b>29.03%</b> |
| QP                                         | 131  | 48   | 179   | <b>26.82%</b> |
| PP                                         | 231  | 77   | 308   | <b>25.00%</b> |
| DP                                         | 1467 | 440  | 1907  | <b>23.07%</b> |
| Ambiguous or unclear                       | 14   | 4    | 18    | <b>22.22%</b> |
| Adjective                                  | 1216 | 255  | 1471  | <b>17.34%</b> |
| <i>álíka</i> + adj. + <i>og</i> 'as...as'  | 7    | 1    | 8     | <b>12.50%</b> |
| CP                                         | 455  | 72   | 527   | <b>13.66%</b> |
| <i>sama og</i> 'same as'                   | 8    | 1    | 9     | <b>11.11%</b> |
| AdvP                                       | 19   | 2    | 21    | <b>9.52%</b>  |
| <i>eins</i> + (adj.) + <i>og</i> 'as...as' | 618  | 46   | 662   | <b>6.93%</b>  |
| <i>sem</i> 'as'                            | 3    | 0    | 3     | <b>0.00%</b>  |

Table A.4: Constituents following *vera* and distribution of *það*

### A.3.2 Subjects in Spec,T

| GENERAL |               | CLAUSE TYPE   |               |                |
|---------|---------------|---------------|---------------|----------------|
| Valid   |               | <i>tht</i>    | <i>inf</i>    | <i>que</i>     |
| -það    | 5             | 1             | 4             | 0              |
| +það    | 66            | 14            | 52            | 5              |
| %það    | <b>92.96%</b> | <b>93.33%</b> | <b>92.86%</b> | <b>100.00%</b> |

Table A.5: Nominative subjects (not extraposed) in Spec,T

| Verb                             | -það | +það | %það           |
|----------------------------------|------|------|----------------|
| <i>vera</i> 'be'                 | 1    | 31   | <b>96.88%</b>  |
| <i>duga</i> 'be enough'          | 2    | 3    | <b>60.00%</b>  |
| <i>hjálpa</i> 'help'             | 0    | 4    | <b>100.00%</b> |
| <i>ganga</i> 'go'                | 1    | 3    | <b>75.00%</b>  |
| <i>tíðkast</i> 'become frequent' | 0    | 3    | <b>100.00%</b> |
| <i>hafa</i> 'have'               | 0    | 3    | <b>100.00%</b> |
| <i>virka</i> 'function'          | 1    | 1    | <b>50.00%</b>  |
| <i>teljast</i> 'be considered'   | 0    | 2    | <b>100.00%</b> |
| <i>sannast</i> 'be proven'       | 0    | 2    | <b>100.00%</b> |
| <i>veita</i> 'provide'           | 0    | 1    | <b>100.00%</b> |

Table A.6: Most common matrix verbs with a nominative clausal subject after the finite verb

| <b>GENERAL</b>         |                         |                      |                         |
|------------------------|-------------------------|----------------------|-------------------------|
|                        | <i>gagnrýna (pass.)</i> | <i>harma (pass.)</i> | <i>opinbera (pass.)</i> |
| <b>Valid</b>           | 1233                    | 140                  | 174                     |
| <b>-það</b>            | 1011                    | 96                   | 162                     |
| <b>+það</b>            | 222                     | 44                   | 12                      |
| <b>%það</b>            | <b>18.00%</b>           | <b>31.43%</b>        | <b>6.90%</b>            |
| <b>CLAUSE TYPE</b>     |                         |                      |                         |
| <i>tht -það</i>        | 1001                    | 96                   | 68                      |
| <i>tht +það</i>        | 216                     | 44                   | 11                      |
| <b><i>tht %það</i></b> | <b>17.75%</b>           | <b>31.43%</b>        | <b>13.92%</b>           |
| <i>inf -það</i>        | 4                       | 0                    | 0                       |
| <i>inf +það</i>        | 1                       | 0                    | 0                       |
| <b><i>inf %það</i></b> | <b>20.00%</b>           | –                    | –                       |
| <i>que -það</i>        | 6                       | 0                    | 94                      |
| <i>que +það</i>        | 5                       | 0                    | 1                       |
| <b><i>que %það</i></b> | <b>45.45%</b>           | –                    | <b>1.05%</b>            |

Table A.7: Passivized accusative clausal objects in Spec,T

| <b>GENERAL</b>         |                   |                         |                      |                    |                       |                      |
|------------------------|-------------------|-------------------------|----------------------|--------------------|-----------------------|----------------------|
|                        | <i>verða vart</i> | <i>krefjast (pass.)</i> | <i>sakna (pass.)</i> | <i>spá (pass.)</i> | <i>fresta (pass.)</i> | <i>fagna (pass.)</i> |
| <b>Valid</b>           | 23                | 4577                    | 13                   | 1772               | 39                    | 501                  |
| <b>-það</b>            | 0                 | 115                     | 1                    | 258                | 7                     | 4                    |
| <b>+það</b>            | 23                | 4462                    | 12                   | 1514               | 32                    | 497                  |
| <b>%það</b>            | <b>100.00%</b>    | <b>97.48%</b>           | <b>92.30%</b>        | <b>85.44%</b>      | <b>82.05%</b>         | <b>99.20%</b>        |
| <b>CLAUSE TYPE</b>     |                   |                         |                      |                    |                       |                      |
| <i>tht -það</i>        | 0                 | 109                     | 1                    | 258                | 4                     | 0                    |
| <i>tht +það</i>        | 23                | 4433                    | 6                    | 1514               | 492                   | 4                    |
| <b><i>tht %það</i></b> | <b>100.00%</b>    | <b>97.60%</b>           | <b>85.71%</b>        | <b>85.44%</b>      | <b>99.19%</b>         | <b>100.00%</b>       |
| <i>inf -það</i>        | 0                 | 6                       | 0                    | 0                  | 0                     | 7                    |
| <i>inf +það</i>        | 0                 | 29                      | 6                    | 0                  | 4                     | 28                   |
| <b><i>inf %það</i></b> | –                 | <b>82.85%</b>           | <b>100.00%</b>       | –                  | <b>100.00%</b>        | <b>80.00%</b>        |
| <i>que -það</i>        | 0                 | 0                       | 0                    | 0                  | 0                     | 0                    |
| <i>que +það</i>        | 0                 | 0                       | 0                    | 0                  | 1                     | 0                    |
| <b><i>que %það</i></b> | –                 | –                       | –                    | –                  | <b>100.00%</b>        | –                    |

Table A.8: Genitive and dative subjects in Spec,T

### A.3.3 Clausal objects

| GENERAL                |               |                |                 |               |                 |
|------------------------|---------------|----------------|-----------------|---------------|-----------------|
|                        | <i>líka</i>   | <i>leiðast</i> | <i>gagnrýna</i> | <i>harma</i>  | <i>opinbera</i> |
| <b>Valid</b>           | 246           | 577            | 4241            | 2492          | 462             |
| <b>-það</b>            | 212           | 553            | 3739            | 1897          | 364             |
| <b>+það</b>            | 34            | 24             | 502             | 595           | 98              |
| <b>%það</b>            | <b>13.82%</b> | <b>4.16%</b>   | <b>11.84%</b>   | <b>23.88%</b> | <b>21.21%</b>   |
| CLAUSE TYPE            |               |                |                 |               |                 |
| <i>tth</i> <b>-það</b> | 21            | 25             | 3639            | 1620          | 265             |
| <i>tth</i> <b>+það</b> | 8             | 2              | 466             | 492           | 89              |
| <i>tth</i> <b>%það</b> | <b>27.59%</b> | <b>7.41%</b>   | <b>11.35%</b>   | <b>23.30%</b> | <b>25.14%</b>   |
| <i>inf</i> <b>-það</b> | 187           | 520            | 57              | 266           | 12              |
| <i>inf</i> <b>+það</b> | 26            | 22             | 22              | 99            | 6               |
| <i>inf</i> <b>%það</b> | <b>12.21%</b> | <b>4.06%</b>   | <b>27.85%</b>   | <b>27.12%</b> | <b>33.33%</b>   |
| <i>que</i> <b>-það</b> | 4             | 8              | 43              | 11            | 87              |
| <i>que</i> <b>+það</b> | 0             | 0              | 14              | 4             | 3               |
| <i>que</i> <b>%það</b> | <b>0.00%</b>  | <b>0.00%</b>   | <b>24.56%</b>   | <b>26.67%</b> | <b>3.33%</b>    |

Table A.9: Nominative and accusative clausal objects

| GENERAL                |               |               |               |                 |                |
|------------------------|---------------|---------------|---------------|-----------------|----------------|
|                        | <i>spá</i>    | <i>fagna</i>  | <i>fresta</i> | <i>krefjast</i> | <i>sakna</i>   |
| <b>Valid</b>           | 12865         | 8926          | 1602          | 32593           | 5228           |
| <b>-það</b>            | 1045          | 157           | 154           | 278             | 124            |
| <b>+það</b>            | 11820         | 8769          | 1448          | 32315           | 5104           |
| <b>%það</b>            | <b>91.88%</b> | <b>98.24%</b> | <b>90.39%</b> | <b>99.15%</b>   | <b>97.63%</b>  |
| CLAUSE TYPE            |               |               |               |                 |                |
| <i>tth</i> <b>-það</b> | 972           | 128           | 2             | 240             | 8              |
| <i>tth</i> <b>+það</b> | 11786         | 7512          | 138           | 28544           | 675            |
| <i>tth</i> <b>%það</b> | <b>92.38%</b> | <b>98.32%</b> | <b>98.57%</b> | <b>99.17%</b>   | <b>98.83%</b>  |
| <i>inf</i> <b>-það</b> | 12            | 27            | 152           | 38              | 116            |
| <i>inf</i> <b>+það</b> | 4             | 1210          | 1310          | 3771            | 4415           |
| <i>inf</i> <b>%það</b> | <b>25.00%</b> | <b>97.82%</b> | <b>89.60%</b> | <b>99.00%</b>   | <b>97.44%</b>  |
| <i>que</i> <b>-það</b> | 61            | 2             | 0             | 0               | 0              |
| <i>que</i> <b>+það</b> | 30            | 47            | 0             | 0               | 14             |
| <i>que</i> <b>%það</b> | <b>32.97%</b> | <b>95.92%</b> | –             | –               | <b>100.00%</b> |

Table A.10: Dative and genitive clausal objects

### A.3.4 Post-copular clauses

| GENERAL                |               |                  |                 |               |                    |                |
|------------------------|---------------|------------------|-----------------|---------------|--------------------|----------------|
|                        | <i>galli</i>  | <i>afleiðing</i> | <i>vandamál</i> | <i>kostir</i> | <i>niðurstöður</i> | <i>rök</i>     |
| <b>Valid</b>           | 325           | 659              | 721             | 39            | 54                 | 142            |
| <b>-það</b>            | 97            | 123              | 678             | 19            | 20                 | 32             |
| <b>+það</b>            | 228           | 536              | 43              | 20            | 34                 | 110            |
| <b>%það</b>            | <b>70.15%</b> | <b>81.34%</b>    | <b>5.96%</b>    | <b>51.28%</b> | <b>62.96%</b>      | <b>77.46%</b>  |
| CLAUSE TYPE            |               |                  |                 |               |                    |                |
| <i>tht -það</i>        | 94            | 122              | 654             | 17            | 20                 | 32             |
| <i>tht +það</i>        | 228           | 535              | 43              | 20            | 34                 | 109            |
| <b><i>tht %það</i></b> | <b>70.81%</b> | <b>81.43%</b>    | <b>6.17%</b>    | <b>54.05%</b> | <b>62.96%</b>      | <b>77.30%</b>  |
| <i>inf -það</i>        | 1             | 1                | 20              | 2             | 0                  | 0              |
| <i>inf +það</i>        | 0             | 1                | 0               | 0             | 0                  | 1              |
| <b><i>inf %það</i></b> | <b>0.00%</b>  | <b>50.00%</b>    | <b>0.00%</b>    | <b>0.00%</b>  | –                  | <b>100.00%</b> |
| <i>que -það</i>        | 2             | 0                | 4               | 0             | 0                  | 0              |
| <i>que +það</i>        | 0             | 0                | 0               | 0             | 0                  | 0              |
| <b><i>que %það</i></b> | <b>0.00%</b>  | –                | <b>0.00%</b>    | –             | –                  | –              |

Table A.11: Post-copular clauses

### A.3.5 Prepositional phrases

| GENERAL                |                           |                  |                  |
|------------------------|---------------------------|------------------|------------------|
|                        | <i>vera meðvitaður um</i> | <i>fjalla um</i> | <i>í kringum</i> |
| <b>Valid</b>           | 5401                      | 3145             | 215              |
| <b>-það</b>            | 4617                      | 2474             | 25               |
| <b>+það</b>            | 784                       | 671              | 190              |
| <b>%það</b>            | <b>14.52%</b>             | <b>21.34%</b>    | <b>88.37%</b>    |
| CLAUSE TYPE            |                           |                  |                  |
| <i>tht -það</i>        | 3361                      | 1256             | 5                |
| <i>tht +það</i>        | 611                       | 352              | 56               |
| <b><i>tht %það</i></b> | <b>15.38%</b>             | <b>21.89%</b>    | <b>91.80%</b>    |
| <i>inf -það</i>        | 496                       | 409              | 15               |
| <i>inf +það</i>        | 55                        | 144              | 114              |
| <b><i>inf %það</i></b> | <b>9.98%</b>              | <b>26.04%</b>    | <b>88.37%</b>    |
| <i>que -það</i>        | 760                       | 809              | 5                |
| <i>que +það</i>        | 118                       | 175              | 20               |
| <b><i>que %það</i></b> | <b>13.44%</b>             | <b>17.78%</b>    | <b>80.00%</b>    |

Table A.12: Prepositions assigning accusative case only

| <b>GENERAL</b>     |                   |                  |                 |
|--------------------|-------------------|------------------|-----------------|
|                    | <i>komast hjá</i> | <i>frétta af</i> | <i>gagnvart</i> |
| <b>Valid</b>       | 4941              | 1092             | 1144            |
| <b>-það</b>        | 262               | 6                | 4               |
| <b>+það</b>        | 4679              | 1086             | 1140            |
| <b>%það</b>        | <b>94.70%</b>     | <b>99.45%</b>    | <b>99.65%</b>   |
| <b>CLAUSE TYPE</b> |                   |                  |                 |
| <i>tth -það</i>    | 14                | 5                | 0               |
| <i>tth +það</i>    | 533               | 1065             | 514             |
| <i>tth %það</i>    | <b>97.44%</b>     | <b>99.53%</b>    | <b>100.00%</b>  |
| <i>inf -það</i>    | 248               | 0                | 1               |
| <i>inf +það</i>    | 4145              | 0                | 579             |
| <i>inf %það</i>    | <b>94.35%</b>     | –                | <b>99.83%</b>   |
| <i>que -það</i>    | 0                 | 1                | 3               |
| <i>que +það</i>    | 1                 | 21               | 47              |
| <i>que %það</i>    | <b>100.00%</b>    | <b>95.45%</b>    | <b>94.00%</b>   |

Table A.13: Prepositions assigning dative case only

| <b>GENERAL</b>     |                  |                   |                   |               |
|--------------------|------------------|-------------------|-------------------|---------------|
|                    | <i>langa til</i> | <i>ætlast til</i> | <i>mælast til</i> | <i>utan</i>   |
| <b>Valid</b>       | 15770            | 7002              | 2199              | 1230          |
| <b>-það</b>        | 13255            | 2643              | 270               | 513           |
| <b>+það</b>        | 2515             | 4359              | 2199              | 717           |
| <b>%það</b>        | <b>15.95%</b>    | <b>62.25%</b>     | <b>87.72%</b>     | <b>58.29%</b> |
| <b>CLAUSE TYPE</b> |                  |                   |                   |               |
| <i>tth -það</i>    | 241              | 2405              | 260               | 267           |
| <i>tth +það</i>    | 110              | 4125              | 1902              | 501           |
| <i>tth %það</i>    | <b>31.34%</b>    | <b>63.17%</b>     | <b>87.97%</b>     | <b>65.23%</b> |
| <i>inf -það</i>    | 13014            | 238               | 10                | 54            |
| <i>inf +það</i>    | 2405             | 234               | 27                | 212           |
| <i>inf %það</i>    | <b>15.60%</b>    | <b>49.58%</b>     | <b>72.97%</b>     | <b>79.70%</b> |
| <i>que -það</i>    | 0                | 0                 | 0                 | 192           |
| <i>que +það</i>    | 0                | 0                 | 0                 | 4             |
| <i>que %það</i>    | –                | –                 | –                 | <b>2.04%</b>  |

Table A.14: Prepositions assigning genitive case only

| <b>GENERAL</b>     |                        |                             |                    |
|--------------------|------------------------|-----------------------------|--------------------|
|                    | <i>styttast í+acc.</i> | <i>vera tilbúinn í+acc.</i> | <i>pæla í+dat.</i> |
| <b>Valid</b>       | 2964                   | 837                         | 802                |
| <b>-það</b>        | 2773                   | 776                         | 377                |
| <b>+það</b>        | 191                    | 61                          | 425                |
| <b>%það</b>        | <b>6.44%</b>           | <b>7.29%</b>                | <b>52.99%</b>      |
| <b>CLAUSE TYPE</b> |                        |                             |                    |
| <i>tht -það</i>    | 2746                   | 16                          | 20                 |
| <i>tht +það</i>    | 190                    | 4                           | 71                 |
| <i>tht %það</i>    | <b>6.47%</b>           | <b>20.00%</b>               | <b>78.02%</b>      |
| <i>inf -það</i>    | 27                     | 760                         | 200                |
| <i>inf +það</i>    | 1                      | 57                          | 125                |
| <i>inf %það</i>    | <b>3.57%</b>           | <b>6.98%</b>                | <b>38.46%</b>      |
| <i>que -það</i>    | 0                      | 0                           | 157                |
| <i>que +það</i>    | 0                      | 0                           | 229                |
| <i>que %það</i>    | –                      | –                           | <b>59.33%</b>      |

Table A.15: Preposition *í* with accusative and dative case

| <b>GENERAL</b>     |                         |                           |                                |
|--------------------|-------------------------|---------------------------|--------------------------------|
|                    | <i>komast yfir+acc.</i> | <i>gleðjast yfir+dat.</i> | <i>vera svekktur yfir+dat.</i> |
| <b>Valid</b>       | 196                     | 1774                      | 780                            |
| <b>-það</b>        | 147                     | 253                       | 189                            |
| <b>+það</b>        | 49                      | 1521                      | 591                            |
| <b>%það</b>        | <b>25.00%</b>           | <b>85.74%</b>             | <b>75.77%</b>                  |
| <b>CLAUSE TYPE</b> |                         |                           |                                |
| <i>tht -það</i>    | 17                      | 101                       | 39                             |
| <i>tht +það</i>    | 14                      | 1084                      | 241                            |
| <i>tht %það</i>    | <b>45.16%</b>           | <b>91.48%</b>             | <b>86.07%</b>                  |
| <i>inf -það</i>    | 129                     | 147                       | 149                            |
| <i>inf +það</i>    | 33                      | 403                       | 343                            |
| <i>inf %það</i>    | <b>20.37%</b>           | <b>73.27%</b>             | <b>69.72%</b>                  |
| <i>que -það</i>    | 1                       | 5                         | 1                              |
| <i>que +það</i>    | 2                       | 34                        | 7                              |
| <i>que %það</i>    | <b>66.67%</b>           | <b>87.18%</b>             | <b>87.50%</b>                  |

Table A.16: Preposition *yfir* with accusative and dative case

| <b>GENERAL</b>     |                           |                          |                         |                         |
|--------------------|---------------------------|--------------------------|-------------------------|-------------------------|
|                    | <i>komast upp um+acc.</i> | <i>ganga út frá+dat.</i> | <i>vanta upp á+acc.</i> | <i>finna upp á+dat.</i> |
| <b>Valid</b>       | 53                        | 3785                     | 1172                    | 118                     |
| <b>-það</b>        | 48                        | 250                      | 1140                    | 21                      |
| <b>+það</b>        | 5                         | 3535                     | 32                      | 97                      |
| <b>%það</b>        | <b>9.43%</b>              | <b>93.39%</b>            | <b>2.73%</b>            | <b>82.20%</b>           |
| <b>CLAUSE TYPE</b> |                           |                          |                         |                         |
| <i>tht -það</i>    | 5                         | 236                      | 1071                    | 1                       |
| <i>tht +það</i>    | 43                        | 3420                     | 32                      | 6                       |
| <i>tht %það</i>    | <b>10.42%</b>             | <b>93.54%</b>            | <b>2.90%</b>            | <b>85.71%</b>           |
| <i>inf -það</i>    | 0                         | 14                       | 66                      | 20                      |
| <i>inf +það</i>    | 0                         | 114                      | 0                       | 90                      |
| <i>inf %það</i>    | –                         | <b>89.06%</b>            | <b>0.00%</b>            | <b>81.82%</b>           |
| <i>que -það</i>    | 5                         | 0                        | 3                       | 0                       |
| <i>que +það</i>    | 0                         | 1                        | 0                       | 1                       |
| <i>que %það</i>    | <b>0.00%</b>              | <b>100.00%</b>           | <b>0.00%</b>            | <b>100.00%</b>          |

Table A.17: Adverb+preposition constructions

### A.3.6 Complements of nouns

| <b>GENERAL</b>     |                |                  |
|--------------------|----------------|------------------|
|                    | <i>lögmati</i> | <i>hagkvæmni</i> |
| <b>Valid</b>       | 822            | 265              |
| <b>-það</b>        | 12             | 2                |
| <b>+það</b>        | 810            | 263              |
| <b>%það</b>        | <b>98.54%</b>  | <b>99.25%</b>    |
| <b>CLAUSE TYPE</b> |                |                  |
| <i>tht -það</i>    | 0              | 2                |
| <i>tht +það</i>    | 102            | 79               |
| <i>tht %það</i>    | <b>100.00%</b> | <b>97.53%</b>    |
| <i>inf -það</i>    | 2              | 10               |
| <i>inf +það</i>    | 161            | 731              |
| <i>inf %það</i>    | <b>98.77%</b>  | <b>98.65%</b>    |
| <i>que -það</i>    | 0              | 0                |
| <i>que +það</i>    | 0              | 0                |
| <i>que %það</i>    | –              | –                |

Table A.18: Complements of nouns

### A.3.7 Complements of adjectives

| GENERAL         |                   |                    |                      |
|-----------------|-------------------|--------------------|----------------------|
|                 | <i>háður+dat.</i> | <i>feginn+dat.</i> | <i>minnugur+gen.</i> |
| <b>Valid</b>    | 2689              | 5729               | 534                  |
| <b>-það</b>     | 7                 | 4840               | 0                    |
| <b>+það</b>     | 2682              | 889                | 534                  |
| <b>%það</b>     | <b>99.74%</b>     | <b>15.52%</b>      | <b>100.00%</b>       |
| CLAUSE TYPE     |                   |                    |                      |
| <i>tht -það</i> | 5                 | 1811               | 0                    |
| <i>tht +það</i> | 2068              | 501                | 534                  |
| <i>tht %það</i> | <b>99.76%</b>     | <b>21.67%</b>      | <b>100.00%</b>       |
| <i>inf -það</i> | 0                 | 3017               | 0                    |
| <i>inf +það</i> | 489               | 383                | 0                    |
| <i>inf %það</i> | <b>100.00%</b>    | <b>11.26%</b>      | –                    |
| <i>que -það</i> | 2                 | 12                 | 0                    |
| <i>que +það</i> | 125               | 5                  | 0                    |
| <i>que %það</i> | <b>98.43%</b>     | <b>29.41%</b>      | –                    |

Table A.19: Complements of adjectives

### A.3.8 Active, passive and middle voice

| VERB                   | ACTIVE |       |               | PASSIVE |      |               |
|------------------------|--------|-------|---------------|---------|------|---------------|
|                        | -það   | +það  | %það          | -það    | +það | %það          |
| <i>gagnrýna + acc.</i> | 3739   | 502   | <b>11.84%</b> | 1011    | 221  | <b>17.94%</b> |
| <i>harma + acc.</i>    | 1897   | 595   | <b>23.88%</b> | 96      | 44   | <b>31.43%</b> |
| <i>opinbera + acc.</i> | 364    | 98    | <b>21.21%</b> | 162     | 12   | <b>6.90%</b>  |
| <i>spá + dat.</i>      | 1045   | 11820 | <b>91.88%</b> | 342     | 1560 | <b>82.02%</b> |
| <i>fresta + dat.</i>   | 154    | 1448  | <b>90.39%</b> | 7       | 34   | <b>82.93%</b> |
| <i>fagna + dat.</i>    | 157    | 8769  | <b>98.24%</b> | 4       | 494  | <b>99.19%</b> |
| <i>krefjast + gen.</i> | 278    | 32315 | <b>99.15%</b> | 134     | 6393 | <b>97.95%</b> |
| <i>sakna + gen.</i>    | 124    | 5104  | <b>97.63%</b> | 1       | 28   | <b>96.55%</b> |

Table A.20: Active vs. passive

| GENERAL                |                       |                |
|------------------------|-----------------------|----------------|
|                        | <i>First position</i> | <i>Spec, T</i> |
| <b>Valid</b>           | 232                   | 10             |
| <b>-það</b>            | 160                   | 0              |
| <b>+það</b>            | 72                    | 10             |
| <b>%það</b>            | <b>31.03%</b>         | <b>100.00%</b> |
| CLAUSE TYPE            |                       |                |
| <i>tht -það</i>        | 32                    | 0              |
| <i>tht +það</i>        | 35                    | 6              |
| <b><i>tht %það</i></b> | <b>52.24%</b>         | <b>100.00%</b> |
| <i>inf -það</i>        | 128                   | 0              |
| <i>inf +það</i>        | 37                    | 4              |
| <b><i>inf %það</i></b> | <b>22.42%</b>         | <b>100.00%</b> |
| <i>que -það</i>        | 0                     | 0              |
| <i>que +það</i>        | 0                     | 0              |
| <i>que %það</i>        | –                     | –              |

Table A.21: Nominative clausal subjects - Middle voice

### A.3.9 Data on sources based on the frequency of það

| Compl./Pos.            | Lexical item  | Source               | %það          |
|------------------------|---------------|----------------------|---------------|
| PP functional (acc.)   | meðvitaður um | 433.pressan.is       | <b>61.11%</b> |
| PP functional (acc.)   | fjalla um     | Fréttatíminn         | <b>66.67%</b> |
| PP functional (gen.)   | ætlast til    | Jónas.is             | <b>0.00%</b>  |
| PP functional (gen.)   | ætlast til    | Silfur Egils         | <b>28.57%</b> |
| PP functional (gen.)   | ætlast til    | Vísindavefurinn      | <b>29.41%</b> |
| PP functional (gen.)   | ætlast til    | Útgefna bækur        | <b>33.33%</b> |
| PP functional (gen.)   | ætlast til    | Eiðfaxi              | <b>36.36%</b> |
| Adj. complement (dat.) | feginn        | Fiskifréttir         | <b>60.00%</b> |
| PP functional (dat.)   | pæla í        | Silfur Egils         | <b>11.11%</b> |
| PP functional (dat.)   | pæla í        | Sjónvarpsfréttir Rúv | <b>21.43%</b> |
| PP lexical (gen.)      | utan          | Fiskifréttir         | <b>0.00%</b>  |
| PP lexical (gen.)      | utan          | Vísindavefurinn      | <b>6.25%</b>  |
| PP lexical (gen.)      | utan          | Sunnlenska           | <b>37.04%</b> |

Table A.22: Exceptions based on sources

# B Data from questionnaires

## B.1 Introduction

In this appendix, I will present more information on the methodology I applied to the 20 questionnaires I administered online and will also show the main results on all the relevant examples. Filler sentences will not be presented, but they can still be consulted in the spreadsheets online on Github. The address to reach them is the same as the one specified in Appendix A, in the folder *Questionnaires and interviews*.

## B.2 Structure

Each of the 20 questionnaires I administered consisted of 25 sentences, in which I tested about 8 examples of a NomCI or a non-NomCI variant. These 20 questionnaires are 10 pairs in two versions, A and B. The two variants of a certain sentence to test were separated and one put in version A and the other one in version B. Moreover, the order of items in variant B was the opposite of variant A. This is also to make sure that the judgments are not affected by how sentences are distributed (see e.g. Práinsson et al. (2013); Bermel et al. (2018)). I also organized the examples in such a way that there would be a constant alternation of non-NomCIs and NomCIs across all the questionnaires. The two versions looked as follows:

| Version A | Sentence    | Version B | Sentence    |
|-----------|-------------|-----------|-------------|
| 1         | Filler      | ...       | ...         |
| 2         | Non-NomCI 1 | 18        | NomCI 3     |
| 3         | Filler      | 29        | Filler      |
| 4         | Filler      | 20        | Filler      |
| 5         | NomCI 2     | 21        | Non-NomCI 2 |
| 6         | Filler      | 22        | Filler      |
| 7         | Filler      | 23        | Filler      |
| 8         | Non-NomCI 3 | 24        | NomCI 1     |
| ...       | ...         | 25        | Filler      |

Table B.1: Questionnaires - Comparison between version A and B

In questionnaires 2 and 4, I tested more examples due to the fact that I had to re-test some from the previous questionnaires. In these two cases, I decided to distribute the examples differently. Instead of having the first example as sentence no. 2 in Version A (see Table B.1), I set it as sentence no. 1 (which corresponds to no. 25 in version B).

## B.3 Questionnaire template

Here is an example of how the questionnaire looked (in Icelandic, English follows):

Ég heiti Mirko Garofalo og er doktorsnemi í íslenskri málfræði við Háskóla Íslands. Mig langar að biðja þig að taka þátt í þessari könnun sem er aðeins fyrir þátttakendur sem hafa íslensku að móðurmáli. Það tekur í mesta lagi 8 mínútur að svara henni. Könnunin samanstendur af 25 setningum sem þú þarft að meta á skala frá 1 til 7.

1 merkir að setningin er ekki eðlileg samkvæmt þinni máltilfinningu.

7 merkir að setningin er eðlileg samkvæmt þinni máltilfinningu.

Vinsamlegast notaðu 4 (hlutlaust) eins lítið og hægt er þegar þú metur setningarnar.

Þessari könnun er EKKI ætlað að prófa kunnáttu þína í íslensku. Hér er aðeins verið að kanna hvað ÞÉR finnst í lagi að segja samkvæmt þinni máltilfinningu. Könnunin er nafnlaus, en mig langar samt að biðja þig um að tilgreina aldur og kyn. Þakka þér kærlega fyrir þátttökuna.

Umsjónarkennari: Jóhannes Gísli Jónsson - jj@hi.is (Prófessor í íslenskri málfræði við Háskóla Íslands)

Aldur: \_\_\_\_\_

Kyn:

Karl

Kona

Vil ekki svara

Annað...

1. [Sentence that participants have to judge]

|              |                       |                       |                       |                       |                       |                       |                       |         |
|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------|
|              | 1                     | 2                     | 3                     | 4                     | 5                     | 6                     | 7                     |         |
| Ekki eðlileg | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Eðlileg |

...

English translation:

*My name is Mirko Garofalo and I am a doctoral student in Icelandic linguistics at the University of Iceland. I would like to ask you to participate in this questionnaire which is only for those who are native speakers of Icelandic. It will take a maximum of 8 minutes to answer. The questionnaire is made up of 25 sentences that you have to judge on a scale from 1 to 7.*

*1 means that the sentence is not acceptable based on your language feeling.*

*7 means that the sentence is acceptable based on your language feeling.*

*Please choose 4 (unsure) as little as possible when you judge the sentences.*

*This questionnaire has NOT the aim of testing your knowledge of Icelandic. Here, it is what YOU find acceptable to say based on your language feeling that is tested. The questionnaire is anonymous, but I still need to ask you to declare your age and gender. Thank you very much for participating.*

*Supervisor: Jóhannes Gísli Jónsson - jj@hi.is (Professor in Icelandic linguistics at the University of Iceland)*

Age: \_\_\_\_\_

Gender:

Male

Female

Do not want to answer

Other...

1. [Sentence that participants have to judge]

|                       |                       |                       |                       |                       |                       |                       |                       |                   |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------------|
|                       | 1                     | 2                     | 3                     | 4                     | 5                     | 6                     | 7                     |                   |
| <i>Not acceptable</i> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <i>Acceptable</i> |

...

## B.4 Participants

Table B.2 shows some general information about the groups of participants that answered the questionnaires. The number of participants based on age and gender groups are also specified. In total, 1054 native speakers responded online:

| N   | Participants | 21–30 | 41–50 | 61–70 | Men | Women | Other |
|-----|--------------|-------|-------|-------|-----|-------|-------|
| 1A  | 69           | 3     | 6     | 21    | 25  | 43    | 1     |
| 1B  | 126          | 12    | 22    | 26    | 28  | 98    | 0     |
| 2A  | 107          | 8     | 33    | 9     | 33  | 73    | 1     |
| 2B  | 26           | 0     | 6     | 6     | 6   | 19    | 1     |
| 3A  | 126          | 11    | 15    | 28    | 33  | 91    | 2     |
| 3B  | 35           | 0     | 13    | 5     | 11  | 24    | 0     |
| 4A  | 21           | 4     | 2     | 3     | 10  | 11    | 0     |
| 4B  | 36           | 14    | 4     | 0     | 8   | 28    | 0     |
| 5A  | 23           | 8     | 4     | 0     | 2   | 20    | 1     |
| 5B  | 20           | 9     | 2     | 1     | 3   | 16    | 1     |
| 6A  | 21           | 8     | 3     | 1     | 7   | 14    | 0     |
| 6B  | 21           | 5     | 2     | 1     | 10  | 11    | 0     |
| 7A  | 26           | 3     | 5     | 1     | 8   | 18    | 0     |
| 7B  | 23           | 6     | 1     | 0     | 2   | 21    | 0     |
| 8A  | 132          | 14    | 22    | 10    | 24  | 108   | 0     |
| 8B  | 76           | 4     | 24    | 6     | 10  | 66    | 0     |
| 9A  | 65           | 8     | 11    | 11    | 18  | 47    | 0     |
| 9B  | 101          | 3     | 23    | 17    | 22  | 79    | 0     |
| 10A | 30           | 2     | 5     | 6     | 8   | 20    | 0     |
| 10B | 34           | 0     | 8     | 5     | 3   | 31    | 0     |

Table B.2: Participants in questionnaires - overview

## B.5 Spreadsheets

The results of each pair of questionnaires have been presented on a specific spreadsheet online (called Q with the respective number) where the reader can take a closer look at the data. Every spreadsheet contains four tabs. The first two just contain the raw lists of judgments from speakers, for version A and B of the questionnaire respectively. The other two, one for each version of the questionnaire, are an overview of the judgments on all items, as shown in the figure below:

| ID | Sentence                                                              | 1   | 2  | 3  | 4  | 5  | 6  | 7  | Total | Weighted average | Tot. Age 1 | Tot. Age 2 | Tot. Age 3 | Tot. men B | Tot. women B |
|----|-----------------------------------------------------------------------|-----|----|----|----|----|----|----|-------|------------------|------------|------------|------------|------------|--------------|
| 1  | Þetta er bara að versna og ég kvíði fyrir mánaðamótunum.              | 14  | 8  | 15 | 4  | 14 | 24 | 47 | 126   | 5.03             | 11         | 22         | 26         | 28         | 98           |
| 2  | Jón ætlar að afhenda blaðamönnum minnisblaðið strax á morgun.         | 4   | 2  | 5  | 3  | 12 | 25 | 75 | 126   | 6.11             | 11         | 22         | 26         | 28         | 98           |
| 3  | Ýmsir hafa gagnrýnt það að börnum sé leyft að alast upp í fangelsinu. | 15  | 9  | 7  | 5  | 16 | 22 | 52 | 126   | 5.16             | 11         | 22         | 26         | 28         | 98           |
| 4  | Það var því miður rænt mig veskinu á leiðinni til þín.                | 119 | 6  | 1  | 0  | 0  | 0  | 0  | 126   | 1.06             | 11         | 22         | 26         | 28         | 98           |
| 5  | Eftir athöfnina var drifð sig heim til að skipta um fót.              | 92  | 8  | 5  | 3  | 5  | 2  | 11 | 126   | 1.98             | 11         | 22         | 26         | 28         | 98           |
| 6  | Mig langar til að þakka þeim sem sendu umsagnir.                      | 6   | 5  | 3  | 6  | 9  | 30 | 67 | 126   | 5.90             | 11         | 22         | 26         | 28         | 98           |
| 7  | Þetta er raunveruleikinn sem þeir ræddu við hvern annan.              | 62  | 15 | 10 | 8  | 11 | 7  | 13 | 126   | 2.71             | 11         | 22         | 26         | 28         | 98           |
| 8  | Ef hann sé ekki göður við þig þá rasskelli ég hann.                   | 118 | 6  | 0  | 0  | 1  | 0  | 1  | 126   | 1.13             | 11         | 22         | 26         | 28         | 98           |
| 9  | Nú styttist í það að flautað verði til leiks á Laugardalsvelli.       | 11  | 9  | 9  | 2  | 10 | 16 | 69 | 126   | 5.50             | 11         | 22         | 26         | 28         | 98           |
| 10 | Hann segist hafa tilkynnt óhappið starfsmönnum sama dag.              | 50  | 18 | 8  | 14 | 14 | 10 | 12 | 126   | 3.02             | 11         | 22         | 26         | 28         | 98           |
| 11 | Stór hluti barnanna í skólanum hafa smitast af flensunni              | 42  | 10 | 5  | 0  | 9  | 20 | 40 | 126   | 4.14             | 11         | 22         | 26         | 28         | 98           |

Figure B.1: Questionnaires: general overview (example of questionnaire 1B)

The general overview shows all the judgments from the speakers as well as the general weighted average for each token. In relation to age groups, the number of participants is listed in the columns *Tot. Age 1/2/3*. The corresponding number of participants based on gender – only men and women – follows. I did not add any other gender category, as the number of those participants who do not identify as male or female and those who did not answer the question on gender is minimal and could not be considered statistically significant.

## B.6 Statistical tests

The results of the questionnaires were analyzed by comparing the weighted average of judgments on variant A and B of the same token. At the same time, I also investigated through a Mann-Whitney U-test (Mann and Whitney (1947))<sup>1</sup> whether there is any statistical difference between the judgments given to both variants. I used this test to determine whether both variants of a sentence are more or less equally acceptable or whether there is any significant difference among them. The results are shown in §B.7 as well as in Chapter 3.

I applied statistical tests as well to age and gender groups in all syntactic positions. In relation to age, I selected three age groups (21–30, 41–50, 61–70) and compared their judgments on a specific syntactic position through a Kruskal-Wallis test (Kruskal and Wallis (1952)) and a Mann-Whitney U-test. In relation to gender, I compared judgments of men and women with a Mann-Whitney U-test once more, although the number of men answering was quite low compared to the one of women (about 25% of participants on average). The number of those speakers who did not declare their gender or did not identify either as a man or woman is minimal, so, unfortunately, they were not taken into account in the gender group comparison as they are not statistically significant in our analysis. Details on the analysis of age and gender groups are offered in §B.9.

## B.7 Relevant tokens in questionnaires

The tokens of clausal nominalization that were investigated specifically in the questionnaires are presented below. The reader can find in the table the total of participants, the number of participants based on judgment (on a scale from 1 to 7), the weighted average, and the results from the Mann-Whitney U test (see values  $z$ ,  $U$  and  $p$ ) applied to each pair of examples:<sup>2</sup>

<sup>1</sup>I performed all the statistical tests mentioned in this section on the website <https://www.socscistatistics.com/>.

<sup>2</sup>These values have been calculated by using the following website:  
<https://www.socscistatistics.com/tests/mannwhitney/default2.aspx>.

| Q  | N  | Sentence                                                                                                                                                           | 1  | 2  | 3  | 4 | 5  | 6  | 7  | Total | W. Avg. | z      | U       | p     |
|----|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|----|----|---|----|----|----|-------|---------|--------|---------|-------|
| 1A | 02 | Atliðingin er sú að Bandaríkin eru í alvarlegri skuldakreppu                                                                                                       | 1  | 1  | 2  | 3 | 7  | 22 | 33 | 69    | 6.07    | -1.349 | 3838.00 | .177  |
| 1B | 24 | Atliðingin er að Bandaríkin eru í alvarlegri skuldakreppu<br>'The consequence is that the USA is in a serious debt crisis.'                                        | 13 | 7  | 9  | 2 | 11 | 27 | 57 | 126   | 5.38    |        |         |       |
| 1A | 05 | Íbúarnir eru háðir að bændur í nágrenninu sjái þeim fyrir mat                                                                                                      | 46 | 10 | 6  | 3 | 0  | 2  | 2  | 69    | 1.77    | 10.446 | 410.00  | <.001 |
| 1B | 21 | Íbúarnir eru háðir því að bændur í nágrenninu sjái þeim fyrir mat<br>'The inhabitants depend on farmers in the neighbourhood who take care of their food.'         | 3  | 2  | 4  | 0 | 7  | 29 | 81 | 126   | 6.31    |        |         |       |
| 1A | 08 | Við erum meðvítuð um það að ákveðnar stéttir þúi við lág laun                                                                                                      | 23 | 8  | 1  | 4 | 4  | 10 | 19 | 69    | 3.93    | 1.563  | 3757.50 | .118  |
| 1B | 18 | Við erum meðvítuð um að ákveðnar stéttir þúi við lág laun<br>'We are aware that certain society groups live on a low salary.'                                      | 23 | 17 | 7  | 4 | 12 | 24 | 39 | 126   | 4.53    |        |         |       |
| 1A | 11 | Verið er að kanna hagkvæmni að byggja sundlaug undir íþrótahúsinu                                                                                                  | 47 | 14 | 2  | 2 | 2  | 0  | 2  | 69    | 1.64    | 10.139 | 525.50  | <.001 |
| 1B | 15 | Verið er að kanna hagkvæmni þess að byggja sundlaug undir íþrótahúsinu<br>'The practicality of building a swimming pool under the sports centre is being checked.' | 5  | 4  | 7  | 9 | 10 | 23 | 68 | 126   | 5.83    |        |         |       |
| 1A | 14 | Ragnheiður kröfðist þess að Árni bæði starfsmennina afsökunar                                                                                                      | 2  | 2  | 1  | 2 | 5  | 6  | 51 | 69    | 6.30    | -7.007 | 1706.00 | <.001 |
| 1B | 12 | Ragnheiður kröfðist að Árni bæði starfsmennina afsökunar<br>'Ragnheiður demanded Arni to apologize to the staff members.'                                          | 32 | 21 | 15 | 7 | 6  | 18 | 27 | 126   | 3.76    |        |         |       |
| 1A | 17 | Nú stýttist í að flautað verði til leiks á Laugardalsvelli                                                                                                         | 1  | 1  | 0  | 0 | 5  | 11 | 51 | 69    | 6.54    | -2.810 | 3287.50 | .004  |
| 1B | 09 | Nú stýttist í það að flautað verði til leiks á Laugardalsvelli<br>'Now the time for the game to start in Laugardagsvöllur is approaching.'                         | 11 | 9  | 9  | 2 | 10 | 16 | 69 | 126   | 5.50    |        |         |       |
| 1A | 20 | Mig langar til þess að þakka þeim sem sendu umsagnir                                                                                                               | 5  | 2  | 1  | 4 | 7  | 13 | 37 | 69    | 5.80    | 0.180  | 4278.50 | .857  |
| 1B | 06 | Mig langar til að þakka þeim sem sendu umsagnir<br>'I want to thank those who sent comments.'                                                                      | 6  | 5  | 3  | 6 | 9  | 30 | 67 | 126   | 5.90    |        |         |       |
| 1A | 23 | Ymsir hafa gagnrýnt að börnum sé leyft að alast upp í fangelsinu                                                                                                   | 16 | 3  | 2  | 2 | 3  | 17 | 26 | 69    | 4.86    | 0.644  | 4103.50 | .522  |
| 1B | 03 | Ymsir hafa gagnrýnt það að börnum sé leyft að alast upp í fangelsinu<br>'Various people have criticized the fact that children are allowed to grow up in prison.'  | 15 | 9  | 7  | 5 | 16 | 22 | 52 | 126   | 5.16    |        |         |       |
| 2A | 01 | Sindri fagnar því að bændur hafi fengið lækkun á raforkuverði                                                                                                      | 7  | 11 | 11 | 3 | 10 | 16 | 49 | 107   | 5.26    | 4.791  | 546.00  | <.001 |
| 2B | 25 | Sindri fagnar að bændur hafi fengið lækkun á raforkuverði<br>'Sindri is happy that farmers got a reduction of the price of electricity.'                           | 11 | 4  | 3  | 2 | 3  | 1  | 2  | 26    | 2.73    |        |         |       |
| 2A | 04 | Kostirnir eru að kerfið er ódyrt og einfalt í notkun                                                                                                               | 1  | 0  | 1  | 1 | 4  | 16 | 84 | 107   | 6.65    | 0.782  | 1252.50 | .435  |
| 2B | 22 | Kostirnir eru þeir að kerfið er ódyrt og einfalt í notkun<br>'The advantages are that the system is cheap and easy to use.'                                        | 0  | 1  | 1  | 0 | 1  | 5  | 18 | 26    | 6.38    |        |         |       |
| 2A | 07 | Þóra var farin að sakna þess að búa á Íslandi                                                                                                                      | 7  | 1  | 2  | 1 | 4  | 15 | 77 | 107   | 6.24    | 5.979  | 336.50  | <.001 |
| 2B | 19 | Þóra var farin að sakna að búa á Íslandi<br>'Þóra had started to miss living in Iceland.'                                                                          | 14 | 2  | 1  | 1 | 2  | 4  | 2  | 26    | 2.81    |        |         |       |
| 2A | 10 | Mér leiddist hvað markhópur bókarinnar var í raun lítil                                                                                                            | 13 | 10 | 10 | 4 | 10 | 22 | 38 | 107   | 4.93    | 3.015  | 859.00  | .002  |
| 2B | 16 | Mér leiddist það hvað markhópur bókarinnar var í raun lítil<br>'I was sorry about how the target group of the book was actually little.'                           | 10 | 1  | 2  | 4 | 2  | 4  | 3  | 26    | 3.42    |        |         |       |

Table B.3: General results from questionnaires - Part 1

| Q  | N  | Sentence                                                                                                                                | 1  | 2  | 3  | 4 | 5  | 6  | 7  | Total | W. Avg. | z      | U       | p           |
|----|----|-----------------------------------------------------------------------------------------------------------------------------------------|----|----|----|---|----|----|----|-------|---------|--------|---------|-------------|
| 2A | 13 | Einar spáir því að liðð komist í úrslitaleikinn                                                                                         | 6  | 0  | 2  | 4 | 6  | 20 | 69 | 107   | 6.18    | 3.588  | 758.00  | <.001       |
| 2B | 13 | Einar spáir að liðð komist í úrslitaleikinn<br>'Einar predicts that the team will reach the finals.'                                    | 6  | 4  | 1  | 3 | 1  | 3  | 8  | 26    | 4.15    |        |         |             |
| 2A | 16 | Að lifa í núvritund snýst um að njóta augnabliksins                                                                                     | 2  | 1  | 3  | 3 | 3  | 19 | 76 | 107   | 6.41    | 0.643  | 1277.00 | .522        |
| 2B | 10 | Það að lifa í núvritund snýst um að njóta augnabliksins<br>'Living in mindfulness is about enjoying the moment.'                        | 1  | 1  | 0  | 2 | 2  | 3  | 17 | 26    | 6.08    |        |         |             |
| 2A | 19 | Vandamálið er það að fólk vill ekki láta bólusetja sig                                                                                  | 8  | 7  | 7  | 6 | 11 | 16 | 52 | 107   | 5.44    | -1.225 | 1174.50 | .218        |
| 2B | 07 | Vandamálið er að fólk vill ekki láta bólusetja sig<br>'The problem is that people do not want to be vaccinated.'                        | 2  | 1  | 0  | 2 | 2  | 2  | 17 | 26    | 5.88    |        |         |             |
| 2A | 22 | Ég var líka mikið að pæla í að kaupa mér hlutabréf                                                                                      | 14 | 8  | 11 | 5 | 11 | 20 | 38 | 107   | 4.90    | 2.416  | 964.50  | <b>.015</b> |
| 2B | 04 | Ég var líka mikið að pæla í því að kaupa mér hlutabréf<br>'I was also thinking a lot about buying shares.'                              | 5  | 4  | 6  | 3 | 2  | 1  | 5  | 26    | 3.62    |        |         |             |
| 2A | 25 | Hann er mjög meðvitaður um það hvaða áhrif tæki hafa á mann                                                                             | 11 | 8  | 12 | 2 | 18 | 20 | 36 | 107   | 4.98    | 0.604  | 1284.00 | .548        |
| 2B | 01 | Hann er mjög meðvitaður um hvaða áhrif tæki hafa á mann<br>'He is very aware of what influence devices have on people.'                 | 6  | 1  | 3  | 1 | 2  | 4  | 9  | 26    | 4.54    |        |         |             |
| 3A | 02 | Ég þori ekki að spá hvað verður um bækur í framtíðinni                                                                                  | 32 | 12 | 14 | 8 | 18 | 15 | 27 | 126   | 3.96    | -1.584 | 1818.00 | .114        |
| 3B | 24 | Ég þori ekki að spá því hvað verður um bækur í framtíðinni<br>'I do not dare predict what is going to happen to books in the future.'   | 8  | 2  | 2  | 0 | 3  | 9  | 11 | 35    | 4.69    |        |         |             |
| 3A | 05 | Við hörmum það að stofnunin vilji hætta við verkefnið                                                                                   | 6  | 11 | 7  | 6 | 11 | 20 | 65 | 126   | 5.58    | -0.329 | 2124.00 | .741        |
| 3B | 21 | Við hörmum að stofnunin vilji hætta við verkefnið<br>'We regret that the institute wants to cancel the project.'                        | 2  | 1  | 2  | 2 | 4  | 5  | 19 | 35    | 5.74    |        |         |             |
| 3A | 08 | Að vörurnar séu ekki enn komnar skiptir engu máli                                                                                       | 19 | 19 | 11 | 5 | 12 | 22 | 38 | 126   | 4.51    | -2.067 | 1700.00 | <b>.038</b> |
| 3B | 18 | Það að vörurnar séu ekki enn komnar skiptir engu máli<br>'It does not matter that the goods have not arrived yet.'                      | 2  | 5  | 3  | 0 | 1  | 7  | 17 | 35    | 5.34    |        |         |             |
| 3A | 11 | Sum hagkerfi heimsins eru háð því að fá orku úr jarðgasi                                                                                | 6  | 4  | 4  | 7 | 6  | 23 | 76 | 126   | 5.98    | 6.356  | 653.50  | <.001       |
| 3B | 15 | Sum hagkerfi heimsins eru háð að fá orku úr jarðgasi<br>'Some economic systems in the world depend on getting energy from natural gas.' | 18 | 6  | 1  | 0 | 3  | 3  | 4  | 35    | 2.69    |        |         |             |
| 3A | 14 | Ég gagnrýni hvernig þið öfluðuð ykkur þessara upplýsinga                                                                                | 16 | 10 | 7  | 6 | 14 | 27 | 46 | 126   | 5.04    | 0.663  | 2042.50 | .509        |
| 3B | 12 | Ég gagnrýni það hvernig þið öfluðuð ykkur þessara upplýsinga<br>'I criticize how you got this information.'                             | 8  | 2  | 1  | 2 | 5  | 4  | 13 | 35    | 4.66    |        |         |             |
| 3A | 17 | Þeim leiddist það að ég væri alltaf að stoppa og taka myndir                                                                            | 10 | 6  | 9  | 3 | 11 | 24 | 63 | 126   | 5.56    | 2.045  | 1705.50 | <b>.040</b> |
| 3B | 09 | Þeim leiddist að ég væri alltaf að stoppa og taka myndir<br>'They were sorry that I was always stopping and taking pictures.'           | 7  | 2  | 3  | 1 | 5  | 5  | 12 | 35    | 4.66    |        |         |             |
| 3A | 20 | Breytir að ég sé ríkur öllu því sem þú hélt um mig?                                                                                     | 72 | 10 | 15 | 4 | 5  | 7  | 13 | 126   | 2.47    | 0.229  | 2148.50 | .818        |
| 3B | 06 | Breytir það að ég sé ríkur öllu því sem þú hélt um mig?<br>'Does the fact that I'm rich change everything you thought about me?'        | 22 | 2  | 2  | 0 | 2  | 2  | 5  | 35    | 2.54    |        |         |             |

Table B.4: General results from questionnaires - Part 2

| Q  | N  | Sentence                                                                                                                                                                               | 1  | 2 | 3 | 4 | 5 | 6  | 7  | Total | W. Avg. | z      | U      | p     |
|----|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|---|---|---|---|----|----|-------|---------|--------|--------|-------|
| 3A | 23 | Enginn kemst hjá því að kaupa sér mat                                                                                                                                                  | 1  | 0 | 3 | 3 | 5 | 17 | 97 | 126   | 6.57    | 7.168  | 455.50 | <.001 |
| 3B | 03 | Enginn kemst hjá að kaupa sér mat<br>'No one can avoid buying food for themselves.'                                                                                                    | 15 | 6 | 1 | 0 | 5 | 4  | 4  | 35    | 3.06    |        |        |       |
| 4A | 01 | Við erum meðvituð um það að ákveðnir þjóðfélags hópar búi við lág laun                                                                                                                 | 2  | 1 | 2 | 1 | 2 | 5  | 8  | 21    | 5.24    | -0.537 | 345.00 | .589  |
| 4B | 25 | Við erum meðvituð um að ákveðnir þjóðfélags hópar búi við lág laun<br>'We are aware that certain groups of society live with a low salary.'                                            | 8  | 2 | 2 | 3 | 3 | 3  | 15 | 36    | 4.67    |        |        |       |
| 4A | 04 | Ýmsir hafa gagnýnt að flóttafólki sé vísað úr landi                                                                                                                                    | 1  | 1 | 1 | 0 | 0 | 6  | 12 | 21    | 6.00    | 0.115  | 370.00 | .904  |
| 4B | 22 | Ýmsir hafa gagnýnt það að flóttafólki sé vísað úr landi<br>'Various people have criticized the fact that refugees are deported.'                                                       | 4  | 0 | 1 | 0 | 4 | 4  | 23 | 36    | 5.89    |        |        |       |
| 4A | 07 | Mér leiddist það hvað markhópur bókarinnar væri í raun lífill                                                                                                                          | 4  | 6 | 3 | 2 | 1 | 2  | 3  | 21    | 3.38    | -1.695 | 275.00 | .089  |
| 4B | 19 | Mér leiddist hvað markhópur bókarinnar væri í raun lífill<br>'I was sorry about how little was the target group of the book.'                                                          | 18 | 4 | 6 | 0 | 1 | 4  | 3  | 36    | 2.61    |        |        |       |
| 4A | 10 | Ég er að pæla í að kaupa hlutabréf                                                                                                                                                     | 0  | 0 | 1 | 1 | 2 | 5  | 12 | 21    | 6.24    | 0.066  | 373.50 | .944  |
| 4B | 16 | Ég er að pæla í því að kaupa hlutabréf<br>'I'm thinking of buying shares.'                                                                                                             | 4  | 2 | 1 | 0 | 3 | 2  | 24 | 36    | 5.72    |        |        |       |
| 4A | 13 | Jón veitti því enga athygli að Sara væri að gráta                                                                                                                                      | 1  | 0 | 1 | 1 | 3 | 8  | 7  | 21    | 5.71    | -4.971 | 77.00  | <.001 |
| 4B | 13 | Jón veitti enga athygli að Sara væri að gráta<br>'Jón paid no attention to the fact that Sara was crying.'                                                                             | 26 | 4 | 2 | 0 | 1 | 0  | 3  | 36    | 1.83    |        |        |       |
| 4A | 16 | Hann er mjög meðvitaður um hvaða áhrif notkun snjallsíma hefur á börn                                                                                                                  | 1  | 0 | 1 | 0 | 0 | 7  | 12 | 21    | 6.19    | -0.703 | 335.00 | .483  |
| 4B | 10 | Hann er mjög meðvitaður um það hvaða áhrif notkun snjallsíma hefur á börn<br>'He is very aware of the influence the use of smartphones has on children.'                               | 5  | 1 | 0 | 1 | 3 | 7  | 19 | 36    | 5.58    |        |        |       |
| 4A | 19 | KR-ingar voru minnugir þess að tapa 2 - 0 fyrir Völsurum                                                                                                                               | 3  | 0 | 1 | 2 | 1 | 8  | 6  | 21    | 5.19    | -4.756 | 90.00  | <.001 |
| 4B | 07 | KR-ingar voru minnugir að tapa 2 - 0 fyrir Völsurum<br>'KR-players were mindful of losing 2-0 against Valsarar.'                                                                       | 28 | 4 | 2 | 0 | 0 | 1  | 1  | 36    | 1.53    |        |        |       |
| 4A | 22 | Heimstaraldurinn svipir öllu vægi að liðið hefði æft stíft fyrir næsta mót                                                                                                             | 13 | 2 | 2 | 3 | 1 | 0  | 0  | 21    | 1.90    | -0.612 | 340.50 | .541  |
| 4B | 04 | Heimstaraldurinn svipir það öllu vægi að liðið hefði æft stíft fyrir næsta mót<br>'The pandemic deprived the team's hard training for the following tournament of all its importance.' | 27 | 2 | 2 | 0 | 0 | 4  | 1  | 36    | 1.89    |        |        |       |
| 4A | 25 | Er það að læra á hjóðfæri eins og að læra að hjóla?                                                                                                                                    | 0  | 1 | 1 | 3 | 3 | 5  | 8  | 21    | 5.62    | -2.928 | 200.50 | .003  |
| 4B | 01 | Er að læra á hjóðfæri eins og að læra að hjóla?<br>'Is learning to play an instrument like learning to ride a bike?'                                                                   | 11 | 6 | 2 | 1 | 4 | 6  | 6  | 36    | 3.64    |        |        |       |
| 5A | 02 | Ég þarf ekki að pæla í hvað aðrir eru að hugsa                                                                                                                                         | 2  | 3 | 0 | 0 | 6 | 6  | 6  | 23    | 5.04    | -1.862 | 153.00 | .062  |
| 5B | 24 | Ég þarf ekki að pæla í því hvað aðrir eru að hugsa<br>'I don't have to reflect upon what others are thinking.'                                                                         | 0  | 0 | 3 | 2 | 0 | 3  | 12 | 20    | 5.95    |        |        |       |

Table B.5: General results from questionnaires - Part 3

| Q  | N  | Sentence                                                                                                                                                           | 1  | 2 | 3 | 4 | 5 | 6 | 7  | Total | W. Avg. | z      | U      | p     |
|----|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|---|---|---|---|---|----|-------|---------|--------|--------|-------|
| 5A | 05 | Mig langar til þess að allir öðlist hamingju                                                                                                                       | 2  | 3 | 1 | 1 | 4 | 2 | 10 | 23    | 5.09    | 1.46   | 169.50 | .144  |
| 5B | 21 | Mig langar til að allir öðlist hamingju<br>'I want everyone to achieve happiness.'                                                                                 | 4  | 1 | 3 | 3 | 2 | 3 | 4  | 20    | 4.15    |        |        |       |
| 5A | 08 | Vandamálið hjá þér er að vera of kurteis við fólk                                                                                                                  | 3  | 5 | 1 | 1 | 1 | 6 | 6  | 23    | 4.48    | 1.168  | 181.50 | .242  |
| 5B | 18 | Vandamálið hjá þér er það að vera of kurteis við fólk<br>'Your problem is that you are too kind to people.'                                                        | 5  | 2 | 1 | 2 | 3 | 6 | 1  | 20    | 3.90    |        |        |       |
| 5A | 11 | Það hvort hann komi eða ekki skiptir engu máli                                                                                                                     | 3  | 2 | 0 | 3 | 3 | 5 | 7  | 23    | 4.91    | 0.803  | 196.50 | .423  |
| 5B | 15 | Hvort hann komi eða ekki skiptir engu máli<br>'It doesn't matter whether he is coming or not.'                                                                     | 4  | 3 | 2 | 0 | 4 | 1 | 6  | 20    | 4.20    |        |        |       |
| 5A | 14 | Ég sakna að amma muni aldrei framar bjóða mér aðra kökusneið                                                                                                       | 13 | 1 | 2 | 0 | 2 | 2 | 3  | 23    | 2.78    | -1.071 | 185.50 | .284  |
| 5B | 12 | Ég sakna þess að amma muni aldrei framar bjóða mér aðra kökusneið<br>'I miss the fact that my grandma will never offer me another slice of cake again.'            | 7  | 1 | 4 | 0 | 2 | 3 | 3  | 20    | 3.50    |        |        |       |
| 5A | 17 | Margir eru meðvitáðir um það að borða holt og reglulega yfir daginn                                                                                                | 1  | 5 | 1 | 1 | 1 | 6 | 8  | 23    | 5.00    | 0.036  | 228.00 | .968  |
| 5B | 09 | Margir eru meðvitáðir um að borða holt og reglulega yfir daginn<br>'Many people are aware of eating healthy and regularly during the day.'                         | 1  | 0 | 1 | 2 | 8 | 2 | 6  | 20    | 5.30    |        |        |       |
| 5A | 20 | Fyrirtækið kannaði hagkvæmni að vöruhúsi yri byggt við ána                                                                                                         | 16 | 5 | 0 | 0 | 0 | 2 | 0  | 23    | 1.65    | -4.565 | 42.00  | <.001 |
| 5B | 06 | Fyrirtækið kannaði hagkvæmni þess að vöruhúsið yri byggt við ána<br>'The company investigated the practicality of building the storage house close to the river.'  | 1  | 4 | 0 | 1 | 2 | 3 | 9  | 20    | 5.20    |        |        |       |
| 5A | 23 | María er fegin því að prófin séu loksins búin                                                                                                                      | 2  | 0 | 0 | 0 | 4 | 6 | 11 | 23    | 5.87    | 0.45   | 211.00 | .652  |
| 5B | 03 | María er fegin að prófin séu loksins búin<br>'María is happy that the exams are finally finished.'                                                                 | 1  | 0 | 0 | 2 | 3 | 6 | 8  | 20    | 5.80    |        |        |       |
| 6A | 02 | Mér leiddist það að fá enga athygli frá Silju                                                                                                                      | 8  | 4 | 1 | 2 | 1 | 2 | 3  | 21    | 3.10    | -3.119 | 96.00  | .002  |
| 6B | 24 | Mér leiddist að fá enga athygli frá Silju<br>'I was sorry to not get any attention from Silja.'                                                                    | 2  | 1 | 1 | 1 | 1 | 3 | 12 | 21    | 5.62    |        |        |       |
| 6A | 05 | Við erum háð hvað fulltrúar Reykjavíkurborgar vilja gera á þessu svæði                                                                                             | 15 | 3 | 1 | 0 | 0 | 0 | 2  | 21    | 1.81    | -4.439 | 43.50  | <.001 |
| 6B | 21 | Við erum háð því hvað fulltrúar Reykjavíkurborgar vilja gera á þessu svæði<br>'We are dependent on what the representatives of Reykjavík want to do in this area.' | 1  | 0 | 2 | 2 | 1 | 4 | 11 | 21    | 5.76    |        |        |       |
| 6A | 08 | Vopnahléið svipti það öllum heitjúlóma að hermenn fórnðu lífi sínu                                                                                                 | 9  | 5 | 0 | 3 | 2 | 1 | 1  | 21    | 2.57    | 0.955  | 182.00 | .337  |
| 6B | 18 | Vopnahléið svipti öllum heitjúlóma að hermenn fórnðu lífi sínu<br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'                  | 13 | 2 | 2 | 1 | 1 | 2 | 0  | 21    | 2.10    |        |        |       |
| 6A | 11 | Dómarinn krafðist að fá að lesa skýrslurnar                                                                                                                        | 4  | 6 | 2 | 2 | 2 | 2 | 3  | 21    | 3.48    | -4.251 | 51.00  | <.001 |
| 6B | 15 | Dómarinn krafðist þess að fá að lesa skýrslurnar<br>'The judge demanded access to the reports.'                                                                    | 0  | 1 | 0 | 0 | 1 | 1 | 18 | 21    | 6.62    |        |        |       |
| 6A | 14 | Því að Sara væri að gráta var engin athygli veitt                                                                                                                  | 16 | 1 | 2 | 0 | 2 | 0 | 0  | 21    | 1.62    | 0.515  | 199.50 | .603  |
| 6B | 12 | Að Sara væri að gráta var engin athygli veitt<br>'No attention was paid to the fact that Sara was crying.'                                                         | 18 | 1 | 1 | 0 | 0 | 1 | 0  | 21    | 1.38    |        |        |       |

Table B.6: General results from questionnaires - Part 4

| Q  | N  | Sentence                                                                                                                                                                                   | 1  | 2  | 3  | 4  | 5  | 6  | 7  | Total | W. Avg. | z      | U       | p     |
|----|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|----|----|----|----|----|----|-------|---------|--------|---------|-------|
| 6A | 17 | Ég harna hvað gerðist í síðustu viku                                                                                                                                                       | 6  | 1  | 3  | 4  | 1  | 5  | 1  | 21    | 3.57    | -0.1   | 216.00  | .920  |
| 6B | 09 | Ég harna það hvað gerðist í síðustu viku<br>'I regret what happened last week.'                                                                                                            | 7  | 3  | 2  | 2  | 0  | 0  | 7  | 21    | 3.62    |        |         |       |
| 6A | 20 | Þið þurftið að vera minnug þess að ekkert er sjálfgefið í lífinu                                                                                                                           | 2  | 5  | 0  | 3  | 2  | 4  | 5  | 21    | 4.43    | 3.546  | 79.00   | <.001 |
| 6B | 06 | Þið þurftið að vera minnug að ekkert er sjálfgefið í lífinu<br>'You have to be mindful of the fact that nothing is taken for granted in life.'                                             | 15 | 2  | 0  | 1  | 0  | 2  | 1  | 21    | 2.00    |        |         |       |
| 6A | 23 | Pau hata aldrei þætt í að þetta handrit geti verið fölsun                                                                                                                                  | 7  | 2  | 1  | 2  | 3  | 4  | 2  | 21    | 3.57    | 0.654  | 194.00  | .515  |
| 6B | 03 | Pau hata aldrei þætt í því að þetta handrit geti verið fölsun<br>'They never thought that this manuscript could be false.'                                                                 | 10 | 1  | 1  | 0  | 5  | 2  | 2  | 21    | 3.14    |        |         |       |
| 7A | 02 | Ég ætlast ekki til að Bjarna sé sagt upp                                                                                                                                                   | 3  | 2  | 3  | 0  | 1  | 6  | 11 | 26    | 5.15    | -1.282 | 234.50  | .200  |
| 7B | 24 | Ég ætlast ekki til þess að Bjarna sé sagt upp<br>'I don't expect Bjarni to be fired.'                                                                                                      | 2  | 0  | 0  | 0  | 3  | 4  | 14 | 23    | 6.04    |        |         |       |
| 7A | 05 | María var ekki tilbúin í það að reka Sverri úr starfi                                                                                                                                      | 7  | 3  | 2  | 2  | 5  | 4  | 3  | 26    | 3.73    | 0.59   | 269.00  | .555  |
| 7B | 21 | María var ekki tilbúin í að reka Sverri úr starfi<br>'María was not prepared to fire Sverri.'                                                                                              | 9  | 4  | 0  | 2  | 1  | 3  | 4  | 23    | 3.30    |        |         |       |
| 7A | 08 | Við getum ekki komist hjá að verkefnið fari í klessu                                                                                                                                       | 11 | 8  | 1  | 0  | 0  | 3  | 3  | 26    | 2.65    | -3.996 | 99.00   | <.001 |
| 7B | 18 | Við getum ekki komist hjá því að verkefnið fari í klessu<br>'We cannot avoid that the project fails.'                                                                                      | 1  | 0  | 1  | 0  | 5  | 6  | 10 | 23    | 5.87    |        |         |       |
| 7A | 11 | Vopnahléið svipir það að hermenn fórnðu lífi sínu öllum heilujómna                                                                                                                         | 24 | 0  | 0  | 0  | 0  | 1  | 1  | 26    | 1.42    | -0.01  | 298.00  | .992  |
| 7B | 15 | Vopnahléið svipir að hermenn fórnðu lífi sínu öllum heilujómna<br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.'                                          | 21 | 2  | 0  | 0  | 0  | 0  | 0  | 23    | 1.09    |        |         |       |
| 7A | 14 | Gögn þáttakenda eru ekki aðgengileg utan að aldur þeirra er birtur á vefsíðunni                                                                                                            | 14 | 4  | 1  | 0  | 0  | 4  | 3  | 26    | 2.69    | -2.934 | 152.00  | .003  |
| 7B | 12 | Gögn þáttakenda eru ekki aðgengileg utan þess að aldur þeirra er birtur á vefsíðunni<br>'Customer data are not accessible apart from the fact that their age is displayed on the website.' | 2  | 4  | 0  | 3  | 2  | 6  | 6  | 23    | 4.78    |        |         |       |
| 7A | 17 | Þess varð ekki vart að hann væri farinn                                                                                                                                                    | 6  | 3  | 3  | 2  | 6  | 2  | 4  | 26    | 3.81    | 3.636  | 117.00  | <.001 |
| 7B | 09 | Það varð ekki vart að hann væri farinn<br>'The fact that he left went unnoticed.'                                                                                                          | 19 | 2  | 0  | 0  | 0  | 1  | 1  | 23    | 1.57    |        |         |       |
| 7A | 20 | Mér líkar að bókinn sé myndskreytt                                                                                                                                                         | 3  | 2  | 0  | 0  | 4  | 6  | 11 | 26    | 5.38    | 0.631  | 267.00  | .528  |
| 7B | 06 | Mér líkar það að bókinn sé myndskreytt<br>'I like the fact that the book is illustrated.'                                                                                                  | 2  | 3  | 0  | 1  | 5  | 4  | 8  | 23    | 5.09    |        |         |       |
| 7A | 23 | Ákvæðið er sett til verndar því að upplýsingar viðskiptavina verði misnotaðar                                                                                                              | 3  | 3  | 3  | 0  | 2  | 7  | 8  | 26    | 4.85    | 4.227  | 87.50   | <.001 |
| 7B | 03 | Ákvæðið er sett til verndar að upplýsingar viðskiptavina verði misnotaðar<br>'The clause is set to prevent information on customers from being misused.'                                   | 14 | 4  | 2  | 1  | 1  | 1  | 0  | 23    | 1.87    |        |         |       |
| 8A | 02 | Líklega varð þess ekki vart að vefsíðan var ekki lengur virk                                                                                                                               | 49 | 26 | 12 | 8  | 11 | 11 | 15 | 132   | 2.99    | 5.588  | 2679.50 | <.001 |
| 8B | 24 | Líklega varð ekki vart að vefsíðan var ekki lengur virk<br>'Probably it went unnoticed that the website was no longer working.'                                                            | 60 | 10 | 2  | 1  | 1  | 1  | 1  | 76    | 1.42    |        |         |       |
| 8A | 05 | Við eigum að gleðjast yfir að framkvæmdir séu loksins hafnar                                                                                                                               | 18 | 11 | 11 | 10 | 13 | 25 | 44 | 132   | 4.82    | -3.661 | 3485.00 | <.001 |
| 8B | 21 | Við eigum að gleðjast yfir því að framkvæmdir séu loksins hafnar<br>'We have to be happy that the operations have finally started.'                                                        | 5  | 5  | 1  | 1  | 6  | 12 | 46 | 76    | 5.87    |        |         |       |

Table B.7: General results from questionnaires - Part 5

| Q  | N  | Sentence                                                                        | 1  | 2  | 3  | 4  | 5  | 6  | 7  | Total | W. Avg. | z       | U       | p     |
|----|----|---------------------------------------------------------------------------------|----|----|----|----|----|----|----|-------|---------|---------|---------|-------|
| 8A | 08 | Það var mikið ferli í kringum það að byggja þetta hótél                         | 26 | 18 | 14 | 10 | 26 | 17 | 21 | 132   | 3.96    | 2.251   | 4074.50 | .024  |
| 8B | 18 | Það var mikið ferli í kringum að byggja þetta hótél                             | 27 | 9  | 8  | 5  | 10 | 10 | 7  | 76    | 3.26    |         |         |       |
|    |    | 'There was a great process around the construction of this hotel.'              |    |    |    |    |    |    |    |       |         |         |         |       |
| 8A | 11 | Hún er ekki tilbúin í að eignast barn svona ung                                 | 16 | 11 | 13 | 6  | 10 | 22 | 54 | 132   | 5.01    | 1.7488  | 4284.50 | .080  |
| 8B | 15 | Hún er ekki tilbúin í það að eignast barn svona ung                             | 16 | 4  | 9  | 5  | 6  | 14 | 22 | 76    | 4.46    |         |         |       |
|    |    | 'She is not ready to have a child at such a young age.'                         |    |    |    |    |    |    |    |       |         |         |         |       |
| 8A | 14 | Peir eru mjög svektir yfir því að tapa leiknum                                  | 19 | 5  | 6  | 4  | 13 | 28 | 57 | 132   | 5.27    | 4.106   | 3299.00 | <.001 |
| 8B | 12 | Peir eru mjög svektir yfir að tapa leiknum                                      | 18 | 9  | 8  | 6  | 12 | 8  | 15 | 76    | 3.91    |         |         |       |
|    |    | 'They are very annoyed about having lost the game.'                             |    |    |    |    |    |    |    |       |         |         |         |       |
| 8A | 17 | Það hefur orðið vart að ökumenn keyra utan leyfilegra svæða                     | 86 | 17 | 9  | 6  | 2  | 5  | 7  | 132   | 1.97    | -6.266  | 2396.00 | <.001 |
| 8B | 09 | Þess hefur orðið vart að ökumenn keyra utan leyfilegra svæða                    | 16 | 11 | 8  | 5  | 11 | 8  | 17 | 76    | 4.00    |         |         |       |
|    |    | 'It has been noticed that drivers drive outside permissible areas.'             |    |    |    |    |    |    |    |       |         |         |         |       |
| 8A | 20 | Sara sagðist ekki vilja vera viðstödd það að hundurinn væri sprautaður          | 62 | 19 | 16 | 6  | 15 | 3  | 11 | 132   | 2.59    | 4.474   | 3145.00 | <.001 |
| 8B | 06 | Sara sagðist ekki vilja vera viðstödd að hundurinn væri sprautaður              | 63 | 6  | 2  | 1  | 0  | 3  | 1  | 76    | 1.45    |         |         |       |
|    |    | 'Sara said she did not want to be there when the dog would get the injection.'  |    |    |    |    |    |    |    |       |         |         |         |       |
| 8A | 23 | Sýnir að Karl keypti nýjan síma að hann er búinn að fá launahækkun?             | 97 | 15 | 8  | 4  | 5  | 2  | 1  | 132   | 1.60    | -3.866  | 3399.50 | <.001 |
| 8B | 03 | Sýnir það að Karl keypti nýjan síma að hann er búinn að fá launahækkun?         | 33 | 14 | 9  | 6  | 9  | 2  | 3  | 76    | 2.50    |         |         |       |
|    |    | 'Does the fact that Karl bought a new phone show that he got a raise?'          |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 02 | Sóttvarnalæknir mælist til að áhættuþopar njóti forgangs við bólusetningar      | 5  | 5  | 5  | 3  | 7  | 9  | 31 | 65    | 5.35    | 2.942   | 2392.50 | .003  |
| 9B | 24 | Sóttvarnalæknir mælist til þess að áhættuþopar njóti forgangs við bólusetningar | 1  | 1  | 1  | 1  | 11 | 18 | 68 | 101   | 6.43    |         |         |       |
|    |    | 'The epidemiologist asks that groups at risk are prioritized for vaccinations.' |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 05 | Sumt fólk upplifir verkkiða gagnvart því að byrja á verkefnum                   | 8  | 7  | 5  | 5  | 8  | 14 | 18 | 65    | 4.72    | -6.1602 | 1420.00 | <.001 |
| 9B | 21 | Sumt fólk upplifir verkkiða gagnvart að byrja á verkefnum                       | 44 | 26 | 9  | 3  | 8  | 8  | 3  | 101   | 2.42    |         |         |       |
|    |    | 'Some people experience job anxiety at the moment of starting projects.'        |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 08 | Það vantaði lítið upp á að Chelsea næði stigi gegn Arsenal                      | 1  | 2  | 2  | 2  | 9  | 14 | 35 | 65    | 6.05    | -3.1065 | 2343.00 | .002  |
| 9B | 18 | Það vantaði lítið upp á það að Chelsea næði stigi gegn Arsenal                  | 14 | 10 | 8  | 2  | 11 | 21 | 35 | 101   | 4.87    |         |         |       |
|    |    | 'Chelsea was so close to drawing against Arsenal.'                              |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 11 | Það er ekki hægt að ætlast til þess að fá allar sínar óskir uppfylltar          | 3  | 4  | 2  | 1  | 5  | 11 | 39 | 65    | 5.92    | -1.096  | 2950.50 | .271  |
| 9B | 15 | Það er ekki hægt að ætlast til að fá allar sínar óskir uppfylltar               | 12 | 2  | 4  | 2  | 8  | 22 | 51 | 101   | 5.59    |         |         |       |
|    |    | 'It is not possible to expect all wishes to be fulfilled.'                      |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 14 | Eftir slysið komst upp um að bíllinn var ótryggður                              | 32 | 12 | 6  | 2  | 4  | 6  | 3  | 65    | 2.45    | 0.8684  | 3019.50 | .384  |
| 9B | 12 | Eftir slysið komst upp um það að bíllinn var ótryggður                          | 47 | 11 | 12 | 4  | 7  | 9  | 11 | 101   | 2.84    |         |         |       |
|    |    | 'After the accident, it came into light that the car was not insured.'          |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 17 | Börn geta fundið upp á því að troða ótrúlegustu hlutum upp í nefið              | 3  | 1  | 3  | 2  | 3  | 13 | 40 | 65    | 6.08    | -1.3299 | 2880.00 | .183  |
| 9B | 09 | Börn geta fundið upp á að troða ótrúlegustu hlutum upp í nefið                  | 7  | 5  | 0  | 2  | 9  | 29 | 49 | 101   | 5.81    |         |         |       |
|    |    | 'Children can find a way to stick the most incredible stuff into their nose.'   |    |    |    |    |    |    |    |       |         |         |         |       |
| 9A | 20 | Við getum ekki gengið út frá því að sagan í þessari grein sé sönn               | 11 | 6  | 4  | 2  | 9  | 11 | 22 | 65    | 4.74    | 3.8906  | 2106.00 | <.001 |
| 9B | 06 | Við getum ekki gengið út frá því að sagan í þessari grein sé sönn               | 3  | 4  | 1  | 3  | 8  | 21 | 61 | 101   | 6.13    |         |         |       |
|    |    | 'We cannot assume that the story in this article is true.'                      |    |    |    |    |    |    |    |       |         |         |         |       |

Table B.8: General results from questionnaires - Part 6

| Q   | N  | Sentence                                                                                         | 1  | 2  | 3 | 4 | 5  | 6  | 7  | Total | W. Avg. | z       | U     | p     |
|-----|----|--------------------------------------------------------------------------------------------------|----|----|---|---|----|----|----|-------|---------|---------|-------|-------|
| 9A  | 23 | Það er mælist til þess að mæla á flugvölinn þremur tímum fyrir brottför                          | 4  | 2  | 5 | 4 | 4  | 11 | 35 | 65    | 5.69    | -4.9625 | 1782  | <.001 |
| 9B  | 03 | Það er mælist til að mæla á flugvölinn þremur tímum fyrir brottför                               | 27 | 18 | 6 | 6 | 10 | 12 | 22 | 101   | 3.77    |         |       |       |
|     |    | 'It is required to arrive at the airport three hours before the departure.'                      |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 02 | Hann var búinn að fréttu af því að María boðaði ekki fisk                                        | 3  | 1  | 1 | 0 | 3  | 7  | 15 | 30    | 5.67    | -3.4037 | 256.5 | <.001 |
| 10B | 24 | Hann var búinn að fréttu af að María boðaði ekki fisk                                            | 12 | 7  | 2 | 1 | 3  | 2  | 7  | 34    | 3.29    |         |       |       |
|     |    | 'He got to know that Maria did not eat fish.'                                                    |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 05 | Lagið fjallar um að njóta augnabliksins                                                          | 2  | 0  | 0 | 1 | 1  | 1  | 25 | 30    | 6.40    | -1.6480 | 387   | .098  |
| 10B | 21 | Lagið fjallar um það að njóta augnabliksins                                                      | 1  | 0  | 1 | 2 | 3  | 8  | 19 | 34    | 6.12    |         |       |       |
|     |    | 'The song is about enjoying the moment.'                                                         |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 08 | Hún efast um lögmæti þess að loka opinberri stofnun með þessum hætti                             | 1  | 0  | 0 | 1 | 0  | 1  | 27 | 30    | 6.67    | -6.1213 | 54.5  | <.001 |
| 10B | 18 | Hún efast um lögmæti að loka opinberri stofnun með þessum hætti                                  | 9  | 11 | 3 | 4 | 3  | 3  | 1  | 34    | 2.82    |         |       |       |
|     |    | 'She doubts the legality of closing a public institution in this way.'                           |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 11 | Rökin fyrir aukagjaldi eru að starfsmenn í sorphirðu séu undir svo miklu álagi                   | 5  | 4  | 1 | 1 | 2  | 4  | 13 | 30    | 4.83    | 1.1771  | 422   | .238  |
| 10B | 15 | Rökin fyrir aukagjaldi eru þau að starfsmenn í sorphirðu séu undir svo miklu álagi               | 3  | 2  | 2 | 0 | 3  | 5  | 19 | 34    | 5.62    |         |       |       |
|     |    | 'The reasons for an extra charge are that workers for waste collection are under much pressure.' |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 14 | María opinberaði það að hún hefði eytt fjórum milljónum í 750 gjafakort                          | 4  | 4  | 3 | 0 | 3  | 5  | 11 | 30    | 4.77    | 0.6794  | 459   | .496  |
| 10B | 12 | María opinberaði að hún hefði eytt fjórum milljónum í 750 gjafakort                              | 3  | 1  | 1 | 2 | 8  | 6  | 13 | 34    | 5.38    |         |       |       |
|     |    | 'Maria revealed that she had spent four million on 750 gift cards.'                              |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 17 | Breska þingið frestaði að fara í sumarfrí                                                        | 6  | 6  | 5 | 0 | 5  | 5  | 3  | 30    | 3.63    | 4.2782  | 191.5 | <.001 |
| 10B | 09 | Breska þingið frestaði því að fara í sumarfrí                                                    | 1  | 1  | 1 | 1 | 6  | 4  | 20 | 34    | 6.00    |         |       |       |
|     |    | 'The British parliament postponed going on vacation.'                                            |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 20 | Í ræðunni er fjallað um það að ekki megi dæma fólk eftir útliti                                  | 0  | 1  | 1 | 0 | 2  | 8  | 18 | 30    | 6.30    | 0.0874  | 503   | .928  |
| 10B | 06 | Í ræðunni er fjallað um að ekki megi dæma fólk eftir útliti                                      | 0  | 0  | 2 | 3 | 2  | 5  | 22 | 34    | 6.24    |         |       |       |
|     |    | 'In the speech, it is said that one may not judge people based on their appearance.'             |    |    |   |   |    |    |    |       |         |         |       |       |
| 10A | 23 | Enginn getur komist yfir að lesa allar bókmenntir heimsins                                       | 0  | 1  | 0 | 2 | 2  | 6  | 19 | 30    | 6.30    | -1.6615 | 386   | .096  |
| 10B | 03 | Enginn getur komist yfir það að lesa allar bókmenntir heimsins                                   | 0  | 2  | 2 | 4 | 2  | 10 | 14 | 34    | 5.71    |         |       |       |
|     |    | 'No one can manage to read all the world's literature.'                                          |    |    |   |   |    |    |    |       |         |         |       |       |

Table B.9: General results from questionnaires - Part 7

## B.8 Clausal indirect objects: Interviews

Here are the data I gathered from the interviews on clausal indirect objects (the examples I refer to are the ones presented in §3.3.7). The participants I interviewed had to judge both the NomCI and non-NomCI variant, with and without extraposition and also in both active and passive voice (apart from extraposed passivized accusative clausal indirect objects, due to the fact that *pað* would have been ambiguous as also expletive *pað* can occupy that position). I marked judgments from speakers in the following way in the table below:

- AB: both variants are grammatical
- A+: the NomCI variant is better
- B+: the non-NomCI variant is better
- A: only the NomCI variant is grammatical
- B: only the non-NomCI variant is grammatical
- X: neither variant is grammatical

As for extraposition, speakers had to tell whether extraposition is allowed or not, whether it is the only option, or whether it is better to extrapose the CP. All the results are presented below:

| N  | Age | Gender | Dat. (active) |         | Dat. (passive) |         | Acc. (active) |        | Acc. (passive) |
|----|-----|--------|---------------|---------|----------------|---------|---------------|--------|----------------|
|    |     |        | Variant       | Extr    | Variant        | Extr    | Variant       | Extr   | Variant        |
| 1  | 58  | F      | A             | Only    | A              | Only    | A             | No     | AB             |
| 2  | 45  | M      | A             | Allowed | B              | No      | A             | No     | AB             |
| 3  | 23  | F      | A             | Allowed | A              | No      | B             | No     | AB             |
| 4  | 23  | F      | A             | Allowed | A              | Allowed | B+            | No     | AB             |
| 5  | 23  | F      | A             | Allowed | A              | No      | B             | No     | B+             |
| 6  | 38  | F      | A             | Allowed | A              | No      | A+            | No     | A+             |
| 7  | 59  | F      | A             | Better  | AB             | No      | A             | Better | AB             |
| 8  | 55  | F      | A             | Better  | AB             | No      | A+            | Only   | AB             |
| 9  | 64  | M      | A             | Allowed | A              | No      | A             | No     | AB             |
| 10 | 66  | F      | A             | Allowed | A              | No      | A             | No     | AB             |
| 11 | 50  | F      | A             | Allowed | A              | Allowed | A+            | No     | AB             |
| 12 | 39  | M      | A             | Allowed | A              | Allowed | A+            | No     | AB             |
| 13 | 73  | M      | A             | Allowed | A              | Allowed | A             | Only   | AB             |
| 14 | 72  | F      | A             | Only    | A              | Only    | A             | Only   | AB             |
| 15 | 58  | M      | A             | Only    | AB             | Only    | A             | Only   | B+             |
| 16 | 64  | M      | AB            | Only    | A              | Only    | B+            | Only   | AB             |
| 17 | 62  | F      | X             | -       | A              | Only    | B+            | Only   | AB             |

Table B.10: Judgments in interviews

Based on the methodology described in §3.3.7, I classified all judgments on a scale from 1 to 3. Here are all the judgments listed. In Table B.11, I show the judgments on clausal nominalization independently of extraposition. In Table B.12, the judgments are also classified on the basis of extraposition:

| N  | Age | Gender | Dat. Act. |      | Dat. Pass |      | Acc. Act. |      | Acc. Pass. |      |
|----|-----|--------|-----------|------|-----------|------|-----------|------|------------|------|
|    |     |        | +pað      | -pað | +pað      | -pað | +pað      | -pað | +pað       | -pað |
| 1  | 58  | F      | 3         | 1    | 3         | 1    | 3         | 1    | 3          | 3    |
| 2  | 45  | M      | 3         | 1    | 1         | 3    | 3         | 1    | 3          | 3    |
| 3  | 23  | F      | 3         | 1    | 3         | 1    | 1         | 3    | 3          | 3    |
| 4  | 23  | F      | 3         | 1    | 3         | 1    | 2         | 3    | 3          | 3    |
| 5  | 23  | F      | 3         | 1    | 3         | 1    | 1         | 3    | 2          | 3    |
| 6  | 38  | F      | 3         | 1    | 3         | 1    | 3         | 2    | 3          | 2    |
| 7  | 59  | F      | 3         | 1    | 3         | 3    | 3         | 1    | 3          | 3    |
| 8  | 55  | F      | 3         | 1    | 3         | 3    | 3         | 2    | 3          | 3    |
| 9  | 64  | M      | 3         | 1    | 3         | 1    | 3         | 1    | 3          | 3    |
| 10 | 66  | F      | 3         | 1    | 3         | 1    | 3         | 1    | 3          | 3    |
| 11 | 50  | F      | 3         | 1    | 3         | 1    | 3         | 2    | 3          | 3    |
| 12 | 39  | M      | 3         | 1    | 3         | 1    | 3         | 2    | 3          | 3    |
| 13 | 73  | M      | 3         | 1    | 3         | 1    | 3         | 1    | 3          | 3    |
| 14 | 72  | F      | 3         | 1    | 3         | 1    | 3         | 1    | 3          | 3    |
| 15 | 58  | M      | 3         | 1    | 3         | 3    | 3         | 1    | 2          | 3    |
| 16 | 64  | M      | 3         | 3    | 3         | 1    | 2         | 3    | 3          | 3    |
| 17 | 62  | F      | 1         | 1    | 3         | 1    | 2         | 3    | 3          | 3    |
|    |     |        | 2.88      | 1.12 | 2.88      | 1.47 | 2.59      | 1.82 | 2.88       | 2.94 |

Table B.11: Judgments from speakers on a scale from 1 to 3 on indirect objects

| N  | Age | Gender | Dat. Act. -extr |      | Dat. Act. +extr |      | Dat. Pass. -extr |      | Dat. Pass. +extr |      | Acc. Act. -extr |      | Acc. Act. +extr |      | Acc. Pass. -extr |      |
|----|-----|--------|-----------------|------|-----------------|------|------------------|------|------------------|------|-----------------|------|-----------------|------|------------------|------|
|    |     |        | +pað            | -pað | +pað            | -pað | +pað             | -pað | +pað             | -pað | +pað            | -pað | +pað            | -pað | +pað             | -pað |
| 1  | 58  | F      | 1               | 1    | 3               | 1    | 1                | 1    | 3                | 1    | 3               | 1    | 1               | 3    | 3                |      |
| 2  | 45  | M      | 3               | 1    | 3               | 1    | 1                | 3    | 1                | 1    | 3               | 1    | 1               | 1    | 3                | 3    |
| 3  | 23  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 1                | 1    | 3               | 1    | 1               | 3    | 3                |      |
| 4  | 23  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 3                | 1    | 2               | 3    | 1               | 1    | 3                | 3    |
| 5  | 23  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 1                | 1    | 3               | 1    | 1               | 1    | 2                | 3    |
| 6  | 38  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 1                | 1    | 3               | 2    | 1               | 1    | 3                | 2    |
| 7  | 59  | F      | 2               | 1    | 3               | 1    | 3                | 3    | 1                | 1    | 2               | 1    | 3               | 1    | 3                | 3    |
| 8  | 55  | F      | 2               | 1    | 3               | 1    | 3                | 3    | 1                | 1    | 1               | 1    | 3               | 2    | 3                | 3    |
| 9  | 64  | M      | 3               | 1    | 3               | 1    | 3                | 1    | 1                | 1    | 3               | 1    | 1               | 1    | 3                | 3    |
| 10 | 66  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 1                | 1    | 3               | 1    | 1               | 1    | 3                | 3    |
| 11 | 50  | F      | 3               | 1    | 3               | 1    | 3                | 1    | 3                | 1    | 3               | 2    | 1               | 1    | 3                | 3    |
| 12 | 39  | M      | 3               | 1    | 3               | 1    | 3                | 1    | 3                | 1    | 3               | 2    | 1               | 1    | 3                | 3    |
| 13 | 73  | M      | 3               | 1    | 3               | 1    | 3                | 1    | 3                | 1    | 1               | 1    | 3               | 1    | 3                | 3    |
| 14 | 72  | F      | 1               | 1    | 3               | 1    | 1                | 1    | 3                | 1    | 1               | 1    | 3               | 1    | 3                | 3    |
| 15 | 58  | M      | 1               | 1    | 3               | 1    | 1                | 1    | 3                | 3    | 1               | 1    | 3               | 1    | 2                | 3    |
| 16 | 64  | M      | 1               | 1    | 3               | 3    | 1                | 1    | 3                | 1    | 1               | 1    | 2               | 3    | 3                | 3    |
| 17 | 62  | F      | 1               | 1    | 1               | 1    | 1                | 1    | 3                | 1    | 1               | 1    | 2               | 3    | 3                | 3    |
|    |     |        | 2.29            | 1.00 | 2.88            | 1.12 | 2.29             | 1.35 | 2.06             | 1.12 | 1.94            | 1.53 | 1.71            | 1.29 | 2.88             | 2.94 |

Table B.12: Judgments from speakers on a scale from 1 to 3 on indirect objects - with extraposition

The results are summarized in the following two tables. In the first one, there is no distinction made between extraposed and non-extraposed variants, whereas the second one shows this distinction:

| DATIVE     |                                                                                                                                                                                                                                           |      |              |              |       |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------------|--------------|-------|
| Type       | Sentence                                                                                                                                                                                                                                  | Type | -það         | +það         | p     |
| Active     | <i>Jón veitti því að Sara væri að gráta enga athygli</i><br><i>Jón veitti því enga athygli að Sara væri að gráta</i><br>'Jón did not pay attention to the fact that Sara was crying.'                                                     | tht  | 1.12<br>(17) | 2.88<br>(17) | <.001 |
| Passive    | <i>Því að Sara væri að gráta var engin athygli veitt</i><br><i>Því/Það var engin athygli veitt að Sara væri að gráta</i><br>'No attention was paid to the fact that Sara was crying.'                                                     | tht  | 1.47<br>(17) | 2.88<br>(17) | <.001 |
| ACCUSATIVE |                                                                                                                                                                                                                                           |      |              |              |       |
| Type       | Sentence                                                                                                                                                                                                                                  | Type | -það         | +það         | p     |
| Active     | <i>Vopnahléið svipti það að hermenn fórnðu lífi sínu öllum hetjuljóma</i><br><i>Vopnahléið svipti það öllum hetjuljóma að hermenn fórnðu lífi sínu</i><br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.' | tht  | 1.82<br>(17) | 2.59<br>(17) | .010  |
| Passive    | <i>Það að hermenn fórnðu lífi sínu var svipt öllum hetjuljóma</i><br>'The fact that soldiers sacrificed their lives was deprived of all heroism.'                                                                                         | tht  | 2.94<br>(17) | 2.88<br>(17) | .779  |

Table B.13: Summary of oral interviews - Clausal indirect objects

| DATIVE     |                                                                                                                                                                |      |              |              |      |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------------|--------------|------|
| Type       | Sentence                                                                                                                                                       | Type | -það         | +það         | p    |
| Active     | <i>Jón veitti (því) að Sara væri að gráta enga athygli</i><br>'Jón did not pay attention to the fact that Sara was crying.'                                    | tht  | 1.00<br>(17) | 2.29<br>(17) | .000 |
| -Extr.     |                                                                                                                                                                |      |              |              |      |
| Active     | <i>Jón veitti (því) enga athygli að Sara væri að gráta</i><br>'Jón did not pay attention to the fact that Sara was crying.'                                    | tht  | 1.12<br>(17) | 2.88<br>(17) | .000 |
| +Extr.     |                                                                                                                                                                |      |              |              |      |
| Passive    | <i>(Því) að Sara væri að gráta var engin athygli veitt</i><br>'No attention was paid to the fact that Sara was crying.'                                        | tht  | 1.35<br>(17) | 2.29<br>(17) | .020 |
| -Extr.     |                                                                                                                                                                |      |              |              |      |
| Passive    | <i>Því/Það var engin athygli veitt að Sara væri að gráta</i><br>'No attention was paid to the fact that Sara was crying.'                                      | tht  | 1.12<br>(17) | 2.06<br>(17) | .020 |
| +Extr.     |                                                                                                                                                                |      |              |              |      |
| ACCUSATIVE |                                                                                                                                                                |      |              |              |      |
| Type       | Sentence                                                                                                                                                       | Type | -það         | +það         | p    |
| Active     | <i>Vopnahléið svipti (það) að hermenn fórnðu lífi sínu öllum hetjuljóma</i><br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.' | tht  | 1.53<br>(17) | 1.94<br>(17) | .262 |
| -Extr.     |                                                                                                                                                                |      |              |              |      |
| Active     | <i>Vopnahléið svipti (það) öllum hetjuljóma að hermenn fórnðu lífi sínu</i><br>'The armistice deprived the soldiers' sacrifice of their lives of all heroism.' | tht  | 1.29<br>(17) | 1.71<br>(17) | .242 |
| +Extr.     |                                                                                                                                                                |      |              |              |      |
| Passive    | <i>(Það) að hermenn fórnðu lífi sínu var svipt öllum hetjuljóma</i><br>'The fact that soldiers sacrificed their lives was deprived of all heroism.'            | tht  | 2.94<br>(17) | 2.88<br>(17) | .779 |
| -Extr.     |                                                                                                                                                                |      |              |              |      |

Table B.14: Summary of oral interviews - Clausal indirect objects - detailed

## B.9 Data on age and gender groups

The observations on age and gender groups were not based on single tokens. Rather, I merged the results of tokens of the same category in order to have more judgments grouped together. I chose this option due to the fact that, for certain questionnaires, not enough participants of the three age groups involved (21–30, 41–50, 61–70) answered. I applied the same principle to gender groups for the sake of consistency.

I classified the categories of tokens on the basis of case and syntactic position, and observed the data in two ways. On the one hand, I checked whether age groups differed in judging a specific non-NomCl variant or a specific NomCl variant in a certain syntactic position (see *Age group difference*). On the other hand, I observed whether a specific age group would judge both variants differently in the same syntactic position (see *Difference between variants based on age group*). Regarding the first observation, I applied a Kruskal-Wallis test (as we are dealing with three different sets of judgments). As for the second observation, I applied a Mann-Whitney U test (as there are only two sets of judgments):

| Position                     | Judgments |       |       |       |       |       |       |       |        | Age group difference |        |             |         |        |                 | Difference between variants based on age group |        |                 |        |        |                 |  |  |  |
|------------------------------|-----------|-------|-------|-------|-------|-------|-------|-------|--------|----------------------|--------|-------------|---------|--------|-----------------|------------------------------------------------|--------|-----------------|--------|--------|-----------------|--|--|--|
|                              | -það      |       |       | það   |       |       | -það  |       |        | það                  |        |             | 20-30   |        |                 | 40-50                                          |        |                 | 60-70  |        |                 |  |  |  |
|                              | 20-30     | 40-50 | 60-70 | Total | 20-30 | 40-50 | 60-70 | Total | H      | p                    | H      | p           | z       | U      | p               | z                                              | U      | p               | z      | U      | p               |  |  |  |
| Nom. subjects 1st            | 37        | 54    | 39    | 130   | 9     | 24    | 12    | 45    | 2.91   | .232                 | 0.497  | .779        | -0.359  | 153    | .718            | 0.573                                          | 595    | .568            | 0.61   | 206    | .541            |  |  |  |
| Nom. subjects in Spec,T      | 46        | 45    | 40    | 131   | 11    | 42    | 16    | 69    | 3.324  | .189                 | 6.067  | <b>.048</b> | -0.98   | 204    | .327            | -1.694                                         | 745    | .091            | -2.421 | 186    | <b>.015</b>     |  |  |  |
| Gen. subjects 1st            | 17        | 24    | 11    | 52    | 7     | 27    | 7     | 41    | 4.47   | .106                 | 3.858  | .145        | -1.206  | 40     | .226            | 2.774                                          | 177    | <b>.005</b>     | -2.762 | 8      | <b>.005</b>     |  |  |  |
| Gen. subjects in Spec,T      | 15        | 29    | 7     | 51    | 22    | 29    | 12    | 63    | 1.589  | .451                 | 1.692  | .429        | 2.335   | 89     | <b>.019</b>     | -4.276                                         | 145    | <b>&lt;.001</b> | 1.352  | 26     | .177            |  |  |  |
| Nom. objects                 | 37        | 60    | 16    | 113   | 36    | 31    | 41    | 108   | 0.862  | .649                 | 0.689  | .708        | -0.948  | 580    | .342            | -0.816                                         | 832    | .412            | -0.55  | 297    | .582            |  |  |  |
| Acc. objects                 | 31        | 50    | 67    | 148   | 48    | 67    | 70    | 185   | 4.773  | .091                 | 0.650  | .722        | 1.937   | 550.5  | .052            | 0.404                                          | 1601   | .689            | -0.581 | 2209.5 | .561            |  |  |  |
| Dat. objects                 | 19        | 36    | 49    | 104   | 27    | 91    | 28    | 146   | 1.618  | .445                 | 4.227  | .892        | 3.491   | 99.5   | <b>&lt;.001</b> | 4.854                                          | 730    | <b>&lt;.001</b> | -3.616 | 344    | <b>&lt;.001</b> |  |  |  |
| Gen. objects                 | 29        | 39    | 37    | 105   | 30    | 47    | 34    | 111   | 1.14   | .565                 | 0.163  | .921        | 2.478   | 271    | <b>.013</b>     | 5.248                                          | 311    | <b>&lt;.001</b> | -3.55  | 320    | <b>.003</b>     |  |  |  |
| Dat. indirect objects        | 19        | 7     | 1     | 27    | 12    | 5     | 2     | 22    | —      | —                    | 0.504  | .777        | -1.601  | 74     | .110            | -1.46                                          | 8      | .144            | —      | —      | —               |  |  |  |
| Acc. indirect objects        | 17        | 7     | 5     | 29    | 27    | 12    | 2     | 41    | 3.203  | .201                 | —      | —           | 0.012   | 230    | .992            | 0.549                                          | 35     | .582            | —      | —      | —               |  |  |  |
| Only-Acc. PP compl.          | 51        | 64    | 63    | 178   | 48    | 88    | 64    | 200   | 1.832  | .399                 | 5.411  | .066        | -1.2287 | 1048   | .218            | 0.2183                                         | 2757   | .825            | 0.3303 | 1947   | .741            |  |  |  |
| Only-Dat. PP compl.          | 14        | 37    | 22    | 73    | 28    | 47    | 52    | 127   | 6.615  | <b>.036</b>          | 1.0396 | .594        | 1.9877  | 121    | <b>.046</b>     | 5.271                                          | 284    | <b>&lt;.001</b> | 3.4650 | 278.5  | <b>&lt;.001</b> |  |  |  |
| Gen. PP compl.               | 40        | 89    | 76    | 205   | 38    | 58    | 62    | 158   | 3.073  | .215                 | 0.533  | .765        | -1.754  | 584    | .080            | -1.186                                         | 2281   | .234            | -1.926 | 1905   | .053            |  |  |  |
| Two-case P, Acc.             | 38        | 49    | 51    | 138   | 28    | 88    | 59    | 175   | 4.732  | .093                 | 6.611  | <b>.036</b> | 0.674   | 479.5  | .502            | -1.868                                         | 1739.5 | .061            | -1.255 | 1294.5 | .207            |  |  |  |
| Two-case P, Dat.             | 60        | 118   | 49    | 227   | 61    | 78    | 38    | 177   | 2.337  | .310                 | 2.22   | .329        | 3.4577  | 1162.5 | <b>&lt;.001</b> | -1.003                                         | 4211.5 | .317            | -1.155 | 795.5  | .246            |  |  |  |
| Acc. lexical PP              | 7         | 27    | 7     | 41    | 17    | 24    | 11    | 52    | 1.016  | .601                 | 5.208  | .073        | 1.619   | 34     | .105            | -1.453                                         | 247    | .147            | -0.905 | 28     | .362            |  |  |  |
| Dat. lexical PP              | 3         | 23    | 17    | 43    | 8     | 11    | 11    | 30    | —      | —                    | 2.367  | .306        | —       | —      | —               | -1.822                                         | 76     | .068            | -2.657 | 36     | <b>.007</b>     |  |  |  |
| Gen. lexical PP              | 5         | 5     | 1     | 11    | 8     | 2     | 0     | 10    | —      | —                    | —      | —           | 2.122   | 5      | <b>.034</b>     | —                                              | —      | —               | —      | —      | —               |  |  |  |
| Gen. compl. of noun          | 12        | 19    | 28    | 59    | 23    | 32    | 36    | 91    | 1.1883 | .552                 | 3.4079 | .181        | 3.5621  | 35     | <b>&lt;.001</b> | 4.8802                                         | 53     | <b>&lt;.001</b> | 6.2253 | 43.5   | <b>&lt;.001</b> |  |  |  |
| Dat. compl. of noun          | 8         | 2     | 0     | 10    | 5     | 5     | 1     | 11    | —      | —                    | —      | —           | -2.488  | 3      | <b>.012</b>     | —                                              | —      | —               | —      | —      | —               |  |  |  |
| Acc. compl. of adj.          | 7         | 27    | 7     | 41    | 17    | 24    | 11    | 52    | 0.001  | .999                 | 5.636  | .059        | 2.032   | 27     | <b>.042</b>     | -2.509                                         | 191    | <b>.012</b>     | 0.271  | 35     | .787            |  |  |  |
| Dat. compl. of <i>feginn</i> | 9         | 2     | 1     | 12    | 8     | 4     | 0     | 12    | —      | —                    | —      | —           | -0.962  | 26     | .337            | —                                              | —      | —               | —      | —      | —               |  |  |  |
| Dat. compl. of <i>háður</i>  | 11        | 23    | 29    | 63    | 32    | 45    | 59    | 136   | 1.83   | .400                 | 1.62   | .444        | 4.328   | 20     | <b>&lt;.001</b> | 5.612                                          | 84     | <b>&lt;.001</b> | 5.787  | 203    | <b>&lt;.001</b> |  |  |  |
| Gen. compl. of adj.          | 19        | 7     | 1     | 27    | 12    | 5     | 2     | 22    | —      | —                    | 3.251  | .196        | -2.859  | 43     | <b>.004</b>     | -1.136                                         | 10     | .254            | —      | —      | —               |  |  |  |
| Neut.Sg. post-copular        | 9         | 11    | 7     | 27    | 22    | 37    | 10    | 69    | 9.524  | <b>.008</b>          | 0.68   | .711        | 0.913   | 77     | .362            | -0.907                                         | 166    | .362            | -2.342 | 11     | <b>.019</b>     |  |  |  |
| Non-default post-copular     | 27        | 65    | 44    | 136   | 5     | 22    | 35    | 62    | 1.545  | .461                 | 2.009  | .366        | -1.115  | 45.5   | .262            | 0.502                                          | 663    | .617            | 0.301  | 739    | .764            |  |  |  |

Table B.15: Comparison between age groups

Below are the data based on gender. I applied the same types of observations, but this time I only adopted the Mann-Whitney U test, as one always has to observe two sets of judgments:

| Position                     | Judgments |       |     |       |       |       | Gender group difference |         |             |        |         |             | Diff. between variants based on gender |        |                 |        |         |                 |
|------------------------------|-----------|-------|-----|-------|-------|-------|-------------------------|---------|-------------|--------|---------|-------------|----------------------------------------|--------|-----------------|--------|---------|-----------------|
|                              | -það      |       | það |       | Total |       | -það                    |         | það         |        | Total   |             | Men                                    |        |                 | Women  |         |                 |
|                              | Men       | Women | Men | Women | Men   | Women | Men                     | Women   | Men         | Women  | Men     | Women       | Men                                    | Women  | Men             | Women  | Men     |                 |
| Nom. subjects 1st            | 69        | 180   | 249 | 19    | 63    | 82    | 0.1                     | 6159    | .920        | 1.088  | 499     | .275        | 0.425                                  | 613    | .667            | -0.930 | 5223    | .352            |
| Nom. subjects in Spec,T      | 65        | 227   | 292 | 31    | 101   | 132   | -2.027                  | 6160    | <b>.042</b> | -0.789 | 1418    | .429        | -1.116                                 | 865    | .262            | -3.557 | 8643    | <b>&lt;.001</b> |
| Gen. subjects 1st            | 24        | 108   | 132 | 10    | 66    | 76    | -2.197                  | 923     | <b>.028</b> | 2.105  | 193     | <b>.034</b> | -0.037                                 | 119    | .968            | -6.649 | 1420    | <b>&lt;.001</b> |
| Gen. subjects in Spec,T      | 12        | 87    | 99  | 32    | 126   | 158   | -0.246                  | 499     | .802        | -1.066 | 1769    | .284        | 2.977                                  | 79     | <b>.002</b>     | 5.779  | 2925    | <b>&lt;.001</b> |
| Nom. objects                 | 70        | 154   | 224 | 58    | 156   | 214   | -0.848                  | 5008    | .395        | 0.547  | 4303    | .582        | 0.370                                  | 1952   | .711            | 1.273  | 11007   | .204            |
| Acc. objects                 | 89        | 214   | 303 | 98    | 274   | 372   | 0.801                   | 8966    | .423        | 0.664  | 12818.5 | .509        | 0.722                                  | 4093.5 | .471            | 0.910  | 27909.5 | .362            |
| Dat. objects                 | 53        | 151   | 204 | 80    | 201   | 281   | -1.153                  | 3574.5  | .250        | 0.597  | 7672.5  | .548        | 3.835                                  | 1285   | <b>&lt;.001</b> | 7.699  | 7899.5  | <b>&lt;.001</b> |
| Gen. objects                 | 43        | 151   | 194 | 71    | 143   | 214   | 0.001                   | 3246    | 1.000       | -0.011 | 5071    | .992        | 5.489                                  | 587    | <b>&lt;.001</b> | -8.859 | 4342    | <b>&lt;.001</b> |
| Dat. indirect objects        | 18        | 39    | 57  | 17    | 25    | 42    | -1.948                  | 237     | .051        | -0.435 | 195     | .659        | -1.683                                 | 102    | .092            | -2.889 | 277     | <b>.003</b>     |
| Acc. indirect objects        | 22        | 43    | 65  | 23    | 60    | 83    | -0.02                   | 471     | .984        | -0.676 | 623     | .496        | 0.522                                  | 230    | .603            | -0.066 | 1280    | .944            |
| Only-Acc. PP compl.          | 84        | 272   | 356 | 111   | 307   | 418   | -0.171                  | 11282   | .865        | 0.750  | 16219   | .453        | -0.828                                 | 4338   | .406            | -0.220 | 41309   | .825            |
| Only-Dat. PP compl.          | 40        | 120   | 160 | 65    | 213   | 278   | 0.857                   | 2182    | .389        | 1.650  | 5985.5  | .098        | 4.859                                  | 563    | <b>&lt;.001</b> | 8.770  | 5382    | <b>&lt;.001</b> |
| Gen. PP complements          | 101       | 337   | 438 | 87    | 257   | 344   | -1.472                  | 15375   | .141        | -0.475 | 10798   | .631        | -1.513                                 | 3830   | .131            | -4.801 | 33353   | <b>&lt;.001</b> |
| Two-case P, Acc.             | 77        | 241   | 318 | 71    | 292   | 363   | -2.095                  | 7806.5  | <b>.035</b> | -0.307 | 10121.5 | .756        | 2.665                                  | 2038.5 | <b>.007</b>     | -2.212 | 31270.5 | <b>.027</b>     |
| Two case P, Dat.             | 108       | 371   | 479 | 79    | 295   | 374   | 0.878                   | 18921.5 | .378        | -0.244 | 11443.5 | .810        | -2.415                                 | 3382.5 | <b>.015</b>     | -2.939 | 47472   | <b>.003</b>     |
| Acc. Lex. PP                 | 10        | 66    | 76  | 24    | 108   | 132   | -1.252                  | 248     | .211        | 0.660  | 1184    | .509        | -0.453                                 | 108    | .652            | 2.577  | 2733    | <b>.009</b>     |
| Dat. Lex. PP                 | 22        | 79    | 101 | 18    | 47    | 65    | -0.686                  | 785     | .490        | 0.623  | 380     | .535        | -2.555                                 | 103    | <b>.010</b>     | -5.586 | 748     | <b>&lt;.001</b> |
| Gen. Lex. PP                 | 8         | 18    | 26  | 2     | 21    | 23    | -0.194                  | 68      | .849        | —      | —       | —           | —                                      | —      | —               | 2.704  | 93      | <b>.006</b>     |
| Gen. compl. of noun          | 30        | 94    | 124 | 39    | 136   | 175   | -0.309                  | 1356.5  | .756        | -0.457 | 2524    | .645        | 5.810                                  | 104.5  | <b>.001</b>     | 10.973 | 948     | <b>&lt;.001</b> |
| Dat. compl. of noun          | 2         | 21    | 23  | 8     | 18    | 26    | —                       | —       | —           | -1.750 | 40      | .080        | —                                      | —      | —               | -4.099 | 43      | <b>&lt;.001</b> |
| Acc. compl. of adj.          | 10        | 66    | 76  | 24    | 108   | 132   | -1.175                  | 253     | .238        | -2.088 | 942     | <b>.036</b> | 2.475                                  | 54     | <b>.013</b>     | 3.703  | 2370    | <b>&lt;.001</b> |
| Dat. compl. of <i>feginn</i> | 3         | 16    | 19  | 2     | 20    | 22    | —                       | —       | —           | —      | —       | —           | —                                      | —      | —               | 0.859  | 133     | .389            |
| Dat. compl. of <i>háður</i>  | 43        | 81    | 124 | 71    | 200   | 271   | -1.716                  | 1414    | .085        | 1.031  | 6515    | .303        | 6.843                                  | 356    | <b>&lt;.001</b> | 11.253 | 1156    | <b>&lt;.001</b> |
| Neut.Sg. post-copular        | 8         | 39    | 47  | 36    | 89    | 125   | 0.75                    | 129     | .453        | -0.730 | 1468    | .465        | 0.654                                  | 122    | .515            | -0.893 | 1563    | .373            |
| Non-default post-copular     | 69        | 193   | 262 | 34    | 93    | 127   | -0.682                  | 6289.5  | .496        | 1.314  | 1339    | .190        | 1.364                                  | 978    | .173            | -0.622 | 8566    | .535            |

Table B.16: Comparison between gender groups

# Bibliography

- Aaron, J. E. (2010). Pushing the envelope: Looking beyond the variable context. *Language variation and change*, 22(1):1–36.
- Abramovitz, R. (2020). *Successive-cyclic wh-movement feeds case competition in Koryak*. Massachusetts Institute of Technology. <https://ling.auf.net/lingbuzz/005213>.
- Alrenga, P. (2005). A sentential subject asymmetry in English and its implications for complement selection. *Syntax*, 8(3):175–207.
- Angelopoulos, N. (2021). *Nominalization of clauses: the clausal prolepsis strategy*. Handout presented at NELS 2021.
- Angelopoulos, N. (2022). Clausal prolepsis: an overview. In Özge Bakay, Pratley, B., and Neu, E., editors, *Proceedings of NELS*, volume 52, pages 29–42.
- Audring, J. (2009). *Reinventing pronoun gender*. Netherlands Graduate School of Linguistics.
- Axel-Tober, K., Holler, A., and Krause, H. (2016). Correlative es vs. das in German. In Frey, W., Meinunger, A., and Schwabe, K., editors, *Innersentential propositional proforms: Syntactic properties and interpretative effect*, pages 49–71.
- Bailyn, J. F. (2012). *The syntax of Russian*. Cambridge University Press.
- Baker, M. (2015). *Case: Its Principles and Its Parameters*. Cambridge University Press.
- Baker, M. C. (2014). On dependent ergative case (in Shipibo) and its derivation by phase. *Linguistic inquiry*, 45(3):341–379.
- Barðdal, J. (2011). Lexical vs. structural case: A false dichotomy. *Morphology*, 21(3-4):619–654.
- Bayer, J., Bader, M., and Meng, M. (2001). Morphological underspecification meets oblique case: Syntactic and processing effects in German. *Lingua*, 111(4-7):465–514.
- Benediktsdóttir, Á. (2023). *Polmynd á fallanda fæti: Fallglötun í þolmynd*. MA thesis in Icelandic. University of Iceland.
- Bennis, H. (2010). *Gaps and dummies*. Amsterdam University Press.
- Bermel, N. and Knittl, L. (2012). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*, 8(2):241–275.

- Bermel, N., Knittl, L., and Russell, J. (2018). Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. *Corpus Linguistics and Linguistic Theory*, 14(2):197–231.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In Hayes, J. R., editor, *Cognition and language development*, pages 279–362. Wiley.
- Blake, B. J. (2001). *Case*. Cambridge University Press.
- Bondaruk, A. (2012). Predicational, specificational, equative and identificational copular clauses in English and Polish. *Roczniki Humanistyczne*, 60(5):7–26.
- Borsley, R. D. and Kornfilt, J. (2000). Mixed extended projections. *Syntax and Semantics*, 32:101–131.
- Bošković, Ž. (2004). PF merger in stylistic fronting and object shift. In Stepanov, A., Fanselow, G., and Vogel, R., editors, *Minimality Effects in Syntax*, pages 37–72. De Gruyter.
- Bošković, Ž. (2005). On the locality of left branch extraction and the structure of NP. *Studia linguistica*, 59(1):1–45.
- Bošković, Ž. (2012). On NPs and clauses. In Grewendorf, G. and Zimmermann, T. E., editors, *Discourse and grammar: From sentence types to lexical categories*, pages 179–245.
- Bowers, J. (2002). Transitivity. *Linguistic inquiry*, 33(2):183–224.
- Brandtler, J. (2013). Evaluability: an alternative approach to polarity sensitivity. In Csipak, E., Eckhart, R., Liu, M., and Sailer, M., editors, *Beyond 'any' and 'ever': New explorations in negative polarity sensitivity*, pages 103–125. Mouton de Gruyter.
- Bresnan, J. (1972). *Theory of complementation in English syntax*. PhD thesis, Massachusetts Institute of Technology.
- Bresnan, J. (1997). Mixed categories as head sharing constructions. In Butt, M. and King, T. H., editors, *Proceedings of the LFG97 Conference*, pages 1–17. CSLI Publications.
- Brugé, L. (2002). The positions of demonstratives in the extended nominal projection. In Cinque, G., editor, *Functional structure in DP and IP: The cartography of syntactic structures*, volume 1, pages 15–53. New York: Oxford University Press.
- Burton, S. and Grimshaw, J. (1992). Coordination and VP-internal subjects. *Linguistic Inquiry*, 23:305–313.
- Butt, M. (2006). *Theories of case*. Cambridge University Press.
- Caha, P. (2009). *The nanosyntax of case*. PhD thesis, University of Tromsø.

- Chambers, J. K. and Trudgill, P. (1980). *Dialectology*. Cambridge University Press.
- Chomsky, N. (1957). *Syntactic structures*. The Hague/Paris: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, N. (1993). A minimalist program for linguistic theory. In Hale, K. and Keyser, S. J., editors, *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*, pages 1–52.
- Chomsky, N. (1995). *The minimalist program*. MIT Press.
- Chomsky, N. (1999). Derivation by phase. In Kayne, R., Zanuttini, R., and Leu, T., editors, *An Annotated Syntax Reader. Lasting Insights and Questions*, pages 482–496. Wiley-Blackwell.
- Chomsky, N. (2000). Minimalist inquiries: The framework (MITOPL 15). *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, pages 89–155.
- Chomsky, N. (2008). On phases. In Freidin, R., Otero, C. P., and Zubizarreta, M. L., editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pages 132–166. MIT Press.
- Christensen, K. K. (1991). AGR, adjunction, and the structure of Scandinavian existential sentences. *Lingua*, 84(2-3):137–158.
- Cornips, L. and Gregersen, F. (2016). The impact of Labov's contribution to general linguistic theory. *Journal of Sociolinguistics*, 20(4):498–524.
- Culbertson, J. and Gross, S. (2009). Are linguists better subjects? *The British journal for the philosophy of science*.
- Culicover, P. W. and Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry*, 21(1):23–47.
- Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The linguistic review*, 27(1):1–23.
- Davies, W. D. and Dubinsky, S. (2003). On extraction from NPs. *Natural language & linguistic theory*, 21(1):1–37.
- De Cuba, C. and Ürögdi, B. (2010). Clearing up the facts on complementation. *University of Pennsylvania Working Papers in Linguistics*, 16(1):41–50.

- Delicado Cantero, M. (2013). Clausal substantivization in Spanish: Syntax and constraints. *Australian Journal of Linguistics*, 33(2):106–120.
- Dinkin, A. J. (2016). Variant-centered variation and the like conspiracy. *Linguistic Variation*, 16(2):221–246.
- Dixon, R. M. (1994). *Ergativity*. Cambridge University Press.
- Dyakonova, M. (2009). *A phase-based approach to Russian free word order*. Netherlands Graduate School of Linguistics.
- Faarlund, J. T. (2019). *The syntax of mainland Scandinavian*. Oxford University Press.
- Farudi, A. (2007). *An antisymmetric approach to Persian clausal complements*. Amherst, University of Massachusetts.
- Frey, W. (2016). On properties differentiating constructions with inner-sentential pro-forms for clauses. In Frey, W., Meinunger, A., and Schwabe, K., editors, *Inner-sentential propositional proforms: Syntactic properties and interpretative effects*, pages 73–104. John Benjamins Publishing.
- Friðjónsson, J. G. (1989). *Samsettar myndir sagna*. Málvísindastofnun Háskóla Íslands.
- Garofalo, M. (2015). *Fyrir að eða fyrir því að? Um fornöfn í forsetningarliðum sem innleiða setningar*. Bachelor's Thesis. University of Iceland.
- Garofalo, M. (2020). The case of Icelandic clausal arguments. *Working Papers in Scandinavian Syntax*, 104:1–28.
- Garofalo, M. (2023). The role and distribution of nominalized clauses: A view from Icelandic. In Jónsson, J. G. and Árnadóttir, T., editors, *Frændafundur*, volume 11, pages 104–116.
- Gentens, C. (2016). The discursive status of extraposed object clauses. *Journal of Pragmatics*, 96:15–31.
- Georgala, E. and Whitman, J. (2009). Ditransitives and applicative structure in Greek. In Elliott, M., Kirby, J., Sawada, O., Staraki, E., and Yoon, S., editors, *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 43, number 1, pages 77–90. Chicago Linguistic Society.
- Gorrie, C. (2008). The split-PP hypothesis and its implications. In Jones, S., editor, *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*, pages 1–13.
- Halliday, M. A. (1991). Towards probabilistic interpretations. *Functional and systemic linguistics: Approaches and uses*, pages 39–61.

- Han, H. J. (2005). A DP/NP-shell for subject CPs. In Cover, R. T. and Kim, Y., editors, *Proceedings of the Annual Meeting of the Berkeley Linguistics Society: General session and parasession on Prosodic variation and change*, volume 31, pages 133–144.
- Han, H. J. (2008). A nominal-shell for CPs: Beyond subject CPs. In Edwards, R. L., Midthlyng, P. J., Sprague, C. L., and Stensrud, K. G., editors, *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 41, number 1, pages 95–110. Chicago Linguistic Society.
- Hara, Y., Kim, Y., Sakai, H., and Tamura, S. (2013). Projections of events and propositions in Japanese: A case study of koto-nominalized clauses in causal relations. *Lingua*, 133:262–288.
- Hartman, J. (2012). *Varieties of clausal complementation*. PhD thesis, MIT.
- Heck, F. and Müller, G. (2007). Extremely local optimization. In Bainbridge, E. and Agbayani, B., editors, *WECOL 2006*, volume 17, pages 170–182.
- Higgins, F. R. (2015). *The pseudo-cleft construction in English*. London: Routledge.
- Hoffmann, T. (2006). Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus linguistics and linguistic theory*, 2(2):165–195.
- Holmberg, A. (2000). Scandinavian stylistic fronting: How any category can become an expletive. *Linguistic inquiry*, 31(3):445–483.
- Hornstein, N. (1999). Movement and control. *Linguistic inquiry*, 30(1):69–96.
- Hornstein, N. and Polinsky, M. (2010). Control as movement: Across languages and constructions. In Hornstein, N. and Polinsky, M., editors, *Movement theory of control*, pages 1–41. John Benjamins Publishing.
- Horsch, J. (2023). Corroborating corpus data with elicited introspection data: A case study. *Jazykovedný časopis*, 74(1):60–69.
- Hrafnbjargarson, G. H. (2003). On stylistic fronting once more. *Working papers in Scandinavian syntax*, 72:153–205.
- Hrafnbjargarson, G. H. (2004). Stylistic fronting. *Studia linguistica*, 58(2):88–134.
- Ingason, A. K. (2016). Suffixation under structural adjacency: The case of Icelandic the-support. In Hammerly, C. and Prickett, B., editors, *Proceedings of NELS*, volume 46, pages 147–160.
- Ingason, A. K. (2018). Icelandic case-marked CP. *Canadian Journal of Linguistics*, 63(3):444–453.
- Jahromi, A. M. (2011). Expletives in Modern Persian. *Acta Linguistica Asiatica*, 1(3):45–56.

- Johansson, T. (2014). *Internprissättning med särskilt fokus på immaterialrätter [Transfer pricing with a particular focus on intellectual property rights]*. Lund University.
- Jónsson, J. G. (1991). Stylistic fronting in Icelandic. *Working Papers in Scandinavian Syntax*, 48:1–43.
- Josefsson, G. (2006). Semantic and grammatical genders in Swedish: Independent but interacting dimensions. *Lingua*, 116(9):1346–1368.
- Jónsson, J. G. (2003). Not so quirky: On subject case in Icelandic. *New perspectives on case theory*, 156:127–163.
- Jónsson, J. G. (2005a). Merkingarflokkar nafnliða og setningagerð. In *Íslensk tunga III: Setningar. Handbók um setningafræði*, chapter 10, pages 435–458. Reykjavík: Almenna bókafélagið.
- Jónsson, J. G. (2005b). Merkingarhlutverk, rökliðir og fallmörkun. In *Íslensk tunga III: Setningar. Handbók um setningafræði*, chapter 8, pages 350–408. Reykjavík: Almenna bókafélagið.
- Jónsson, J. G. (2013). Dative versus accusative and the nature of inherent case. In Fernández, B. and Etxepare, R., editors, *Variation in datives: A microcomparative perspective*, pages 144–160. Oxford University Press.
- Jónsson, J. G. and Eythórsson, T. (2011). Structured exceptions and case selection in Insular Scandinavian. In Simon, H. J. and Wiese, H., editors, *Expecting the unexpected: Exceptions in grammar*, volume 216 of *Trends in Linguistics. Studies and Monographs*, pages 213–242. De Gruyter.
- Kallulli, D. (2006). Triggering factivity: Prosodic evidence for syntactic structure. In *Proceedings of WCCFL*, volume 25, pages 211–219.
- Kaltenböck, G. (2003). On the syntactic and semantic status of anticipatory it. *English Language & Linguistics*, 7(2):235–255.
- Kaplan, R. M. and Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In Bresnan, J., editor, *The mental representation of grammatical relations*, pages 173–218. MIT Press.
- Kastner, I. (2015). Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua*, 164:156–188.
- Kayne, R. S. (1994). *The antisymmetry of syntax*. MIT Press.
- Kim, J.-B. and Sag, I. (2005). Variations in English object extraposition. In Edwards, R. L., Midtlyng, P. J., Sprague, C. L., and Stensrud, K. G., editors, *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 41, pages 251–266. Chicago Linguistic Society.

- Kiparsky, P. and Kiparsky, C. (1971). Fact. In Steinberg, D. and Jakobovits, L., editors, *Semantics*, pages 345–369. Cambridge University Press.
- Knyazev, M. (2016). *Licensing clausal complements: The case of Russian čto-clauses*. Netherlands Graduate School of Linguistics (LOT).
- Kornfilt, J. and Whitman, J. (2011a). Afterword: Nominalizations in syntactic theory. *Lingua*, 121(7):1297–1313.
- Kornfilt, J. and Whitman, J. (2011b). Introduction: Nominalizations in syntactic theory. *Lingua*, 121(7):1160–1163.
- Korsah, S. and Murphy, A. (2019). Removing clausal determiners in Kwa. In Murphy, A., editor, *Structure Removal*, volume 94 of *Linguistische Arbeits Berichte*, pages 247–294. Institut für Linguistik. Universität Leipzig.
- Koster, J. (1978). Why subject sentences don't exist. In Keyser, S. J., editor, *Recent transformational studies in European languages*, pages 53–64. MIT Press.
- Kratzer, A. (1996). Severing the external argument from its verb. In Rooryck, J. and Zaring, L., editors, *Phrase structure and the lexicon. Studies in natural language and linguistic theory*, volume 33, pages 109–137. Dordrecht: Springer.
- Kress, B. (1982). *Isländische Grammatik*. Hueber.
- Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3):199–244.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.
- Labov, W. (1966). The linguistic variable as a structural unit. *Washington Linguistics Review*, 3:4–22.
- Landau, I. (2007). EPP extensions. *Linguistic Inquiry*, 38(3):485–523.
- Lau, J. H., Clark, A., and Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive science*, 41(5):1202–1241.
- Levin, T. and Preminger, O. (2015). Case in Sakha: Are two modalities really necessary? *Natural Language & Linguistic Theory*, 33:231–250.
- Lohndal, T. (2017). Sentential subjects in English and Norwegian. In *Formal Grammar*, pages 175–202. Routledge.

- Maling, J. (1990). Inversion in embedded clauses in Modern Icelandic. In Maling, J. and Zaenen, A., editors, *Modern Icelandic Syntax*, volume 24 of *Syntax and Semantics*, pages 71–91. Brill.
- Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 18:50–60.
- Marantz, A. (1981). *On the nature of grammatical relations*. PhD thesis, MIT.
- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics*, 4(2):201–225.
- Marantz, A. (2000). Case and licensing. In Reuland, E. J., editor, *Arguments and case: Explaining Burzio's generalization*, pages 11–30. John Benjamins Publishing.
- Mardale, A. (2011). Prepositions as a semilexical category. *Bucharest Working Papers in Linguistics*, 13(2):57–73.
- Masuoka, T. (2007). *Nihongo Modariti Tankyuu [Investigations of Japanese Modality]*. Tokyo: Kuroshio Shuppan.
- McCloskey, J. (1991). There, it, and agreement. *Linguistic Inquiry*, 22(3):563–567.
- McNally, L. (1992). VP coordination and the VP-internal subject hypothesis. *Linguistic Inquiry*, 23:336–341.
- Mikkelsen, L. (2005). *Subject choice in copular clauses*. University of California Berkeley.
- Moro, A. (1997). *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge University Press.
- Moulton, K., Bogal-Allbritten, E., and Shimoyama, J. (2020). *Things we embed*. Talk at Brussels Conference on Generative Linguistics 13. <https://www.crissp.be/wp-content/uploads/2020/12/H-moulton2.pdf>.
- Müller, G. (2017). Structure removal: An argument for feature-driven Merge. *Glossa: a journal of general linguistics*, 2(1):1–35.
- Müller, G. (2018). Structure removal in complex prefields. *Natural Language & Linguistic Theory*, 36:219–264.
- Myers, J. (2009). Syntactic judgment experiments. *Language and Linguistics Compass*, 3(1):406–423.
- Myler, N. (2018). Complex copula systems as suppletive allomorphy. *Glossa: a journal of general linguistics*, 3(1).

- Nowenstein, I. (2014). Intra-speaker variation in subject case: Icelandic. *University of Pennsylvania Working Papers in Linguistics*, 20(1):1–10.
- Ott, D. (2009). Stylistic fronting as remnant movement. *Working papers in Scandinavian syntax*, 83:141–178.
- Ott, D. (2014). Controlling for movement: Reply to Wood 2012. *Working Papers in Scandinavian Syntax*, 92:58–65.
- Ott, D. (2018). Stylistic fronting as remnant movement. *Studia Linguistica*, 72(1):1–38.
- Ott, D. and de Vries, M. (2014). A biclausal analysis of right-dislocation. In Huang, H.-L., Poole, E., and Rysling, A., editors, *Proceedings of NELS*, volume 43, pages 41–54.
- Panagiotidis, E. P. and Grohmann, K. K. (2009). Mixed projections: Categorical switches and prolific domains. *Linguistic Analysis*, 35(1):141–162.
- Pereltsvaig, A. (2007). The universality of DP: A view from Russian. *Studia linguistica*, 61(1):59–94.
- Pereltsvaig, A. (2013). Noun phrase structure in article-less Slavic languages: DP or not DP? *Language and linguistics compass*, 7(3):201–219.
- Pesetsky, D. (2016). *Exfoliation: towards a derivational theory of clause size*. Paper presented at Shrinking Trees Workshop. University of Leipzig.
- Pesetsky, D. (2021). *Exfoliation: Towards a derivational theory of clause size*. <https://lingbuzz.net/lingbuzz/0044440/current.pdf>.
- Pesetsky, D. and Torrego, E. (2001). T-to-C movement: Causes and consequences. *Current Studies in Linguistics Series*, 36:355–426.
- Pesetsky, D. and Torrego, E. (2004). Tense, case, and the nature of syntactic categories. In Guéron, J. and Lecarme, J., editors, *The syntax of time*, volume 40 of *Current studies in linguistics*, pages 495–538. MIT.
- Pesetsky, D. and Torrego, E. (2011). Case. In Boeckx, C., editor, *Oxford Handbook of Linguistic Minimalism*, pages 52–72. Oxford University Press.
- Picallo, M. C. (2002). Abstract agreement and clausal arguments. *Syntax*, 5(2):116–147.
- Pietraszko, A. (2019). Obligatory CP nominalization in Ndebele. *Syntax*, 22(1):66–111.
- Pimentel, J. L. (2019). Some biases in Likert scaling usage and its correction. *International Journal of Science: Basic and Applied Research (IJSBAR)*, 45(1):183–191.

- Platzack, C. (1987). The Scandinavian languages and the null-subject parameter. *Natural Language & Linguistic Theory*, 5:377–401.
- Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry*, 20(3):365–424.
- Preminger, O. (2011). *Agreement as a fallible operation*. PhD thesis, MIT.
- Progovac, L. (1998). Determiner phrase in a language without determiners (with apologies to Jim Huang 1982). *Journal of linguistics*, 34(1):165–179.
- Pütz, H. (1975). *Über die Syntax der Pronominalform "es" im modernen Deutsch*, volume 3 of *Studien zur deutschen Grammatik*. Tübingen: Narr.
- Pütz, M., Robinson, J. A., and Reif, M. (2014). The emergence of cognitive sociolinguistics: An introduction. In Pütz, M., Robinson, J. A., and Reif, M., editors, *Cognitive sociolinguistics: Social and cultural variation in cognition and language use*, pages 1–22. John Benjamins Publishing.
- Pylkkänen, L. (2000). What applicative heads apply to. *University of Pennsylvania Working Papers in linguistics*, 7(1):197–210.
- Quirk, R. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- Rauh, G., Porth-Keromnès, Y., and François, J. (1994). Prépositions et rôles: points de vue syntaxique et sémantique. *Langages*, pages 45–78.
- Rizzi, L. (1997). The fine structure of the left periphery. In Haegeman, L., editor, *Elements of grammar*, pages 281–337. Springer.
- Rochmont, M. S. and Culicover, P. W. (1997). Deriving dependent right adjuncts in English. *Rightward movement*, pages 279–300.
- Rögnvaldsson, E. (1990). *Íslensk orðhlutafræði. Kennslukver handa nemendum á háskólastigi*. Málvísindastofnun Háskóla Íslands.
- Rögnvaldsson, E. and Thráinsson, H. (1990). On Icelandic word order once more. In Maling, J. and Zaenen, A., editors, *Modern Icelandic Syntax. Syntax and Semantics*, volume 24, pages 3–40. San Diego: Academic Press.
- Rosenbaum, P. S. (1967). *The grammar of English predicate complement structures*. PhD thesis, MIT.
- Ross, J. R. (1967). *Constraints on variables in syntax*. ERIC.
- Roussou, A. (1991). Nominalized clauses in the syntax of Modern Greek. *UCL working papers in linguistics*, 3:77–100.

- Saah, K. K. and Goodluck, H. (1995). Island effects in parsing and grammar: Evidence from Akan. *The Linguistic Review*, 12:381–410.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press.
- Schwabe, K., Frey, W., and Meinunger, A. (2016). Sentential proforms. an overview. In Frey, W., Meinunger, A., and Schwabe, K., editors, *Inner-sentential propositional proforms: Syntactic properties and interpretative effects*, pages 1–22. John Benjamins Publishing.
- Sedrins, A. P. (2015). On extraction from definite DPs. *Poznan Studies in Contemporary Linguistics*, 51(2):291–313.
- Shahar, J. (2008). *What some its are: Non-referential it, extraposition, and copies*. PhD thesis, City University of New York.
- Sigurðsson, H. Á. (1989). *Verbal syntax and case in Icelandic: In a comparative GB approach*. PhD thesis, Lund University.
- Sigurðsson, H. Á. (2003). Case: abstract vs. morphological. In Brandner, E. and Zinsmeister, H., editors, *New perspectives on case theory*, pages 223–268.
- Sigurðsson, H. Á. (2006). The Icelandic noun phrase: Central traits. *Arkiv för nordisk filologi*, 121:193–236.
- Sigurðsson, H. Á. (2011). On the new passive. *Syntax*, 14(2):148–178.
- Sigurðsson, H. Á. (2012). Minimalist C/case. *Linguistic Inquiry*, 43(1):191–227.
- Soheili, A. (2024). *A Parametric Approach to Persian Syntax*. Cambridge Scholars Publishing.
- Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic inquiry*, 19(3):425–449.
- Stefánsdóttir, L. B. and Ingason, A. K. (2018). A high definition study of syntactic lifespan change. *University of Pennsylvania Working Papers in Linguistics*, 24(1):169–178.
- Steingrímsson, S. (2019). Risamálheildin. *Orð og tunga*, 21:159–168.
- Stowell, T. A. (1981). *Origins of phrase structure*. PhD thesis, MIT.
- Stroik, T. S. (1996). Extraposition and expletive-movement: A minimalist account. *Lingua*, 99(4):237–251.
- Subbarao, K. V. (1984). *Complementation in Hindi syntax*. Academic Publications.

- Sudhoff, S. (2003). *Argumentsätze und es-Korrelate: zur syntaktischen Struktur von Nebensatzeinbettungen im Deutschen*. Berlin: Wissenschaftlicher Verlag.
- Sudhoff, S. (2016). Correlates of object clauses in German and Dutch. In Frey, W., Meinunger, A., and Schwabe, K., editors, *Inner-sentential propositional proforms: Syntactic properties and interpretative effects*, pages 23–48. John Benjamins Publishing.
- Sybesma, R. (2021). Voice and little *v* and VO–OV Word-Order variation in Chinese languages. *Syntax*, 24(1):44–77.
- Takahashi, S. (2010). The hidden side of clausal complements. *Natural Language & Linguistic Theory*, 28(2):343–380.
- Thorvaldsdóttir, T. (2019). Agreement with conjoined singular noun phrases in Icelandic. *Glossa: a journal of general linguistics*, 4(1):1–33.
- Thráinsson, H. (2009). *Íslensk setningafræði*. 7th edition. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Thráinsson, H., Petersen, H. P., Hansen, Z. S., et al. (2012). *Faroese: An overview and reference grammar*. Fróðskapur Books.
- Thráinsson, H. (1979). *On Complementation in Icelandic*. New York: Garland.
- Thráinsson, H. (2007). *The syntax of Icelandic*. Cambridge University Press.
- Tómasdóttir, Á. B. (2021). *Hann leyndi henni sannleikanum: Tilbrigði í fallmörkun tveggja andlaga sagna*. BA thesis in General Linguistics. University of Iceland.
- Toosarvandani, M. (2008). Wh-movement and the syntax of sluicing. *Journal of Linguistics*, 44(3):677–722.
- Van Riemsdijk, H. (2007). Case in spatial adpositional phrases: The dative-accusative alternation in German. In Alboiu, G., Avram, A., Avram, L., and Dana, I., editors, *Pitar Moș: A building with a view. Festschrift for Alexandra Cornilescu*, pages 1–23. Bucharest: Bucharest University Press.
- Vangsnes, Ø. A. (2002). Icelandic expletive constructions and the distribution of subject types. *Subjects, expletives, and the EPP*, pages 43–70.
- Waters, C. (2008). The internal structure of place class prepositional phrases in English. In Jones, S., editor, *Proceedings of the 2008 annual conference of the Canadian linguistic association*, pages 1–11. University of British Columbia, Vancouver.
- Wood, J. (2012). Against the movement theory of control: Another argument from Icelandic. *Linguistic Inquiry*, 43(1):322–330.

- Wood, J. (2015a). *Applicatives and Applied Datives*, pages 207–250. Springer International Publishing.
- Wood, J. (2015b). *Icelandic morphosyntax and argument structure*. Dordrecht: Springer.
- Wood, J. (2017a). The accusative-subject generalization. *Syntax*, 20(3):249–291.
- Wood, J. (2017b). Icelandic object extraposition is still a problem for the movement theory of control: A reply to Drummond and Hornstein. *Linguistic Inquiry*, 48(3):513–527.
- Woolford, E. (2006). Lexical case, inherent case, and argument structure. *Linguistic inquiry*, 37(1):111–130.
- Wurmbrand, S. (2012). The syntax of valuation in auxiliary-participle constructions. *Coyote Papers*, pages 154–162.
- Yang, C. D. (2000). Internal and external forces in language change. *Language variation and change*, 12(3):231–250.
- Yip, M., Maling, J., and Jackendoff, R. (1987). Case in tiers. *Language*, 63(2):217–250.
- Yoon, H.-J. (2001). Expletive it in English. *Studies in Generative Grammar*, 11:543–562.
- Yousef, S. (2018). *Persian: A comprehensive grammar*. Routledge.
- Zaenen, A., Maling, J., and Thráinsson, H. (1985). Case and grammatical functions: The Icelandic passive. *Natural Language & Linguistic Theory*, 3(4):441–483.
- Zeijlstra, H. (2012). There is only one way to agree. *The Linguistic Review*, 29(3):491–539.
- Zimmermann, I. (2016). Phonological, morphosyntactic and semantic properties of es. In Frey, W., Meinunger, A., and Schwabe, K., editors, *Inner-sentential propositional proforms: Syntactic properties and interpretative effects*, pages 147–170. John Benjamins Publishing.
- Zuckermann, G. (2006). Complement clause types in Israeli. In Dixon, R. M. W. and Aikhenvald, A. Y., editors, *Complementation: A Cross-Linguistic Typology*, pages 72–92. Oxford University Press.
- Práinsson, H., Jónsson, J., Rögnvaldsson, E., Sigurjónsdóttir, S., Magnúsdóttir, S., and Blöndal, P. (2005). *Íslensk tunga III: Setningar: Handbók um setningafræði*. Reykjavík: Almenna bókafélagið.
- Práinsson, H., Ásgrímur Angantýsson, and Sigurðsson, E. F., editors (2013). *Tilbrigði í íslenskri setningagerð: I. Markmið, aðferðir og efniviður*. Málvísindastofnun Háskóla Íslands.