Topicalization:
The IO/DO Asymmetry in Icelandic
Elena Callegari and Anton Karl Ingason
University of Iceland

Abstract In this paper, we investigate differences in the frequency of direct-object versus indirect-object topicalization (i.e. fronting with no accompanying pronominal resumption) in Icelandic using the Icelandic Parsed Historical Corpus. We find that the overall incidence of DO topicalization is double that of IO left topicalization. We argue that this follows from the cross-linguistic preference towards having topical information appear before focal one: while DO topicalization can help ensure that this configuration is obtained when the IO is in focus, there is nothing to gain from topicalizing the IO when the DO is in focus, as IO > DO is already the unmarked order in Icelandic.

1 Ditransitives & Topicalization in Icelandic

There are six possible case patterns that can occur with verbs taking two objects in Icelandic (Zaenen and Maling 1990), the most frequent one being an indirect object (IO) in the dative case and a direct object (DO) in the accusative case (DAT + ACC pattern). The DAT + ACC pattern is also the only case pattern that allows for the two objects to appear in either order: in principle, both the order DAT < ACC (illustrated in (1)) and the order ACC<DAT (as seen in (2)) are possible.

(1) Ég gaf Elínu bókina.
   I gave Eileen.DAT book.the.ACC
   ‘I gave Eileen the book.’

(2) Ég gaf bókina Elínu.
   I gave book.the.ACC Eileen.DAT
   ‘I gave Eileen the book.’

*This paper was presented at DGfS in Freiburg 2021 (virtually). Thanks to the DGfS audience for useful discussions.

1 We use the symbol "<" to indicate linear precedence.
In practice, however, the DAT<ACC order is considered unmarked while the ACC<DAT one is restricted to specific contexts. According to Ottósson (1991), for instance, the ACC<DAT order is possible only whenever the dative object is in focus and the accusative object is not. According to Collins and Thráinsson (1996); Falk (1990), there is also an effect of idiomaticity and of the nominal vs. pronominal nature of the internal arguments. Regarding the latter, Falk (1990) noticed that the ACC<DAT is not possible if the IO is pronominal, as illustrated in (3) below, from (Falk 1990:86):

(3) *Ég gaf bókina honum.
I gave book.the ACC him DAT
‘Intended: I gave him the book.’

The ungrammaticality of (3) is likely linked to the same constraint preventing a non-focused IO from appearing after the DO: pronominal objects are generally not in focus. An experimental study by Dehé (2004) showed how the ACC<DAT order is in fact considerably more restricted than previously assumed. Dehé ran an acceptability study testing 18 native speakers of Icelandic. She found out that the order ACC<DAT was consistently rated quite poorly (i.e. participants rated ACC<DAT as “rather odd” or as plain “wrong”, in all but two experimental items), even whenever the dative IO was in focus. Thus, there seems to be a strong tendency to favor the order DAT<ACC over the order ACC<DAT, regardless of the focal or non-focal nature of the constituents in questions. Dehé concluded that Icelandic uses prosody rather than syntactic movement to mark focus. Hence whenever the DO (in the accusative case) is in focus, Icelandic resorts to stress shift rather than to scrambling: the dative IO undergoes destressing and main stress is relocated to the rightmost stressable unit of the DO.

Given the strong preference for the order IO<DO, we wanted to determine what happens when internal arguments are the target of topicalization. Icelandic has two left-dislocation strategies: under topicalization (see (4)), a constituent is
dislocated to the left periphery, leaving a gap in the position from which it is extracted. Under left dislocation, a corresponding pronominal element appears in place of the constituent appearing in the left periphery (see (5)).

**Icelandic Topicalization**

(4) Elínu gaf ég bókina.
Eileen.DAT gave I book.the.ACC
‘I gave Eileen the book.’

**Icelandic Left Dislocation**

(5) [Presturinn], María sá [hann] í bænum í gær.
the.priest Mary saw him in town yesterday
‘The priest, Mary saw him downtown yesterday.’
(Thráinsson 2007: 358)

Concerning the informational-structural properties of constituents which are the target of either topicalization or left-dislocation, it should be noted that the situation in Icelandic is less clear-cut than what we observe in other languages. In a language like Italian, for example, left dislocation accompanied by clitic resumption is strongly associated with topicality, whereas left dislocation not accompanied by clitic resumption is generally associated with focal constituents. In Icelandic, both constructions appear to be compatible with a topic interpretation of the constituent which appears in the left periphery; to quote Thráinsson (2007), in both topicalization and left dislocation, the left-peripheral XP must generally be definite and hence already introduced as a "topic (or theme) of the discussion" (Thráinsson 2007: 342). About left dislocation, Thráinsson then states that the discourse func-

---

2 Following much existing literature on Icelandic (Thráinsson 1975, 1979, Thráinsson et al. 2007, Maling 1980), we refer to structures where a resumptive pronominal element is present as "left dislocation", and to structures where no resumptive element is present as "topicalization". Note that this is a bit of a terminological tangle: in other languages, for instance those in the Romance subgroup, the term "topicalization" is used to describe the opposite type of structure: left dislocation accompanied by a resumptive element. Moreover, the term "dislocation" evokes a movement operation, whereas the fact that the pronominal object occurs together with the left-peripheral constituent "presturinn" in (5) is rather indicative of a base-generation analysis for "presturinn".
tion of this operation is the "reintroduction of a discourse topic or theme" (Thráins-
son 2007: 358).

2 Local vs. Non-local precedence relations

Examples (1) and (2) illustrate instances of local precedence: the two internal ar-
gments are adjacent and appear within an identical clausal domain, the VP. An
internal argument may also non-locally precede the other internal argument; such
is the case, for instance, when either object has been the target of topicalization.
This is illustrated in (6) for IO topicalization, and in (7) for DO topicalization:

**IO topicalization**

(6) Elínu gave ég bókina.
Eileen.DAT gave I book.the.ACC
‘I gave Eileen the book.’

**DO topicalization**

(7) Bókina gave ég Elínu.
book.the.ACC gave I Eileen.DAT
‘I gave Eileen the book.’

Thanks to Dehé (2004), we know that in local configurations the preferred order
is IO<DO: the dative object preferably precedes the accusative object. Is this same
ordering preference maintained when the two internal arguments no longer appear
in the same local domain, as it is the case in (6) and (7)? If the answer to this ques-
tion were to be affirmative, we would expect DO topicalization to be fundamentally
less frequent than IO topicalization; that is because in DO topicalization structures
the DO precedes the IO object. It is precisely this type of question that we set to
investigate with this paper. More specifically, we wanted to determine whether the
preference for the IO<DO order that we observe at the local level is an absolute
type of constraint, i.e. it applies regardless of the relative distance and structure
between DO and IO, or a relative one. Several languages display asymmetries be-
tween local and non-local configurations, hence assuming that the preference for
IO<DO could be tramped non-locally was not a particularly far-fetched hypothe-
sis. An example of a local/non-local asymmetry is the relative order of fronted foci
and the polarity complementizer "if" (Callegari, in press). In several languages,
foci must necessarily follow the polarity complementizer if both elements appear
in the same left periphery (local configuration). If the constituent in focus is fronted
to a higher left periphery, however, the focus can grammatically precede the polar-
ity complementizer. An illustration of this asymmetry is provided below (example
from Italian):

**Locally = *FOC < if**

(8) Mi domando se A GIANNI hai parlato.
    Refl I-wonder if TO GIANNI you-have spoken
    ‘I wonder if you have spoken to GIANNI (not to JOHN).’

(9) *Mi domando A GIANNI se hai parlato.
    Refl I-wonder TO GIANNI if you-have spoken
    Intended: ‘I wonder if you have spoken to GIANNI (not to JOHN).’

**Non-locally = FOC < if**

(10) A GIANNI mi domando se hai parlato.
    TO GIANNI refl I-wonder if you-have spoken
    ‘I wonder if you have spoken to GIANNI (not to JOHN).’

Getting back to the relative order of internal arguments under topicalization in
Icelandic double-object constructions, we expected either of the following three,
logically possible scenarios:

- The preference for the order IO < DO at the local level is an absolute type of
  constraint: IO topicalization is more frequent than DO topicalization.
- There is a local/non-local asymmetry: DO topicalization is more frequent than
IO topicalization.

• There is a local/non-local asymmetry: DO and IO topicalization are equally frequent.

3 IcePaHC

To determine the overall frequency of IO versus DO topicalization in Iceland, we performed a corpus study using the Icelandic Parsed Historical Corpus, IcePaHC (Wallenberg et al. 2011; Rögnvaldsson et al. 2011, 2012; Rögnvaldsson et al. 2012). The IcePaHC is a diachronic corpus of Icelandic written texts, dating from the 12th century all the way to the 21st century. These belong to a variety of different genres: the IcePaHC features scientific, legal, religious, narrative and biographical texts. Searching the IcePaHC thus allows one to investigate the frequency of a given linguistic construction as spanning throughout several different century and across several different genres. The IcePaHC offers another advantage: it can easily be searched using the PaCQL (Parsed Corpus Query Language, Ingason 2016) through the freely available online platform treebankstudio.org, which also gives users the possibility of obtaining visual summaries of their results, broken down for century, genre and coding criteria.

To perform our query, we searched for all instances of matrix-clause double-object constructions, and coded these depending on whether:
• IO topicalization had occurred,
• DO topicalization had occurred,
• no topicalization had occurred.

This was done with an aim to determine:

• The overall frequency of left dislocation of internal arguments in ditransitive constructions;
• The frequency of DO topicalization in ditransitive constructions;
• The frequency of IO topicalization ditransitive constructions

The exact queries we used can be found in the Appendix. Note that we restricted our search to matrix clauses because even though Icelandic is more liberal than other Scandinavian languages in allowing for topicalization in embedded clauses, not all types of subordinates support it (Thráinsson 2007) (see also Angantýsson (2011) for some extremely interesting experimental data on inter-speaker variation on the acceptability of embedded topicalization).

4 Results

We found 1100 instances of matrix-clause double-object constructions; 128 out of these featured the topicalization of either the direct or the indirect object. This means that the overall rate of object topicalization for ditransitive constructions in the IcePaHC is 11%: roughly one in ten instances of ditransitive features a topicalized internal argument. Out of these 128 instances of object left dislocation, 89 were instances of DO topicalization (incidence: 8%), while 39 were instances of IO topicalization (incidence: 3.5%). This means that, in our corpus, DO topicalization is more than twice as frequent as IO topicalization. Below is a breakdown by century for the two types of topicalization. Overall the rate of DO and IO topicalization seem to proceed in parallel, with perhaps the exception of the 12th century,
where we see an unexpected spike in the number of DO topicalizations. We speculate that this might be due to the particular genre involved: most of the texts from the 12th century are religious texts, and these generally tend to be grandiloquent in style.

Several instances of IO topicalization featured syntactically complex fronted objects; an example is provided in (11) below. In (11), we see that the IO contains an embedded relative.

(11) og [öllum ríkismönnum þeim sem þar voru] gaf hann nokkura and [all powerful-men those that there were](DAT) gave he some
góða gjöf og sæmilega
good gift and respectable
’and he gave all the powerful men that were present some good and respectable gift’ (From Finnboga Saga Ramma, 1330-1370)

As such, we decided to also investigate the average length of the constituents which were the target of topicalization. The average length in words of topicalized IO objects was 2.6, whereas the average length of topicalized DO objects was 1.9 words. On average, dislocated IOs are thus longer (and hence syntactically more complex) than dislocated DOs.
5 Analysis

Our results point to the existence of an asymmetry between local and non-local ordering configurations of the two internal arguments: while IOs must preferably precede DOs locally, non-locally we observe no preference for topicalizing the IO rather than the DO. In non-local configurations, precedence relations are thus reversed: DO topicalization is twice as common as IO topicalization. How can we make sense of this asymmetry? We argue that it follows from a combination of factors:

1. IOs being externally merged higher than DOs in Icelandic (Collins and Thráinsson 1996).

2. The well-known, cross-linguistic preference towards having topical constituents precede focal ones (Prince 1981).

3. A penalty for overt syntactic movement, which privileges structures where no internal merge has applied if internal merge can be avoided (Chomsky 1995), i.e. if it is not necessary for convergence.

As we will see in detail, we argue that the interaction of these three factors causes the in situ configuration to be the preferred one whenever the DO is in focus, discouraging speakers from topicalizing the IO even when this is a topic. Whenever the IO is in focus, on the other hand, the in situ configuration is no longer the most optimal configuration, generating optionality wrt how topicality of the DO is marked.

We will be modeling the interaction of these three factors using Prince & Smolensky’s Optimality Theory (Prince and Smolensky 1993/2004). In our models, the constraint $Topic < Focus$ will represent the preference for having focalized constituents appear after topical ones, whereas the constraint $Stay$ will represent the penalty for overt syntactic movement. Note that for simplicity we have grouped together both topicalization (movement to the left periphery) and A-scrambling.
(VP-internal reordering of the the two internal arguments), meaning that these two types of syntactic movements will be equally violating Stay.

Let us first consider what happens in situations where the DO is in focus and the IO is not. Table 1 lists the 4 possible word order configurations that could logically arise in this instance; in options 3 and 4, topicalization has occurred. We see order nr. 1 is the only configuration that does not violate Stay nor Topic<Focus: no scrambling or left dislocation of the objects has occurred, and the focused DO correctly appears after the non-focused IO. Order 2 violates both conditions: scrambling of the objects has occurred, violating Stay, and the focused DO precedes the non-focused IO, violating Topic < Focus. The same applies to order 4, the one difference wrt to order 3 being that in this case Stay is violated because topicalization, and not scrambling, has occurred. Finally, configuration 3 violates Stay, as the IO is the target of topicalization.

Put differently, whenever the DO is focused and the IO is not, the "best" configuration is the one that already obtains naturally through external merge. In these cases, scrambling and topicalization do not result in any improvement of the structure because they either result in a violation of Stay or of Topic<Focus; as such, they are dispreferred. This accounts for why IO topicalization is relatively infrequent: the in situ configuration, which is also compatible with a non-focal interpretation of the IO, does not violate Stay.

Let us now consider what happens whenever the IO is in focus and the DO is not, i.e. the reverse situation. As can be seen from Table 2, it is immediately evident that all ordering configurations now violate some constraint. Order 1, the in situ configuration, violates Topic<Focus: the DO appears right-mostly even though it

<table>
<thead>
<tr>
<th></th>
<th>Order</th>
<th>Topic&lt;Focus</th>
<th>Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>V &lt; IO &lt; DO</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>V &lt; DO &lt; IO</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>IO &lt; V &lt; S &lt; DO</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DO &lt; V &lt; S &lt; IO</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Ordering options with DO in focus
is the IO which is in focus. Order 2, where scrambling has occurred, violates Stay. Order 3 violates both constraints: topicalization of the IO has occurred, violating Stay as well as Topic < Focus, since the IO is in focus. Finally, order 4, where DO topicalization has taken place, violates Stay.

What is important to note here is that DO topicalization violates as many constraints as leaving the DO in situ, and scrambling the DO past the the IO VP-internally. As orders 1,2 and 4 all violate some constraint, we expect optionality wrt to what order is selected (see for instance Bobaljik & Wurmbrand’s (2008) concept of a "¾ signature").

Summing up, in cases where the IO is topical, left dislocation of the IO is dispreferred, as the in situ configuration already obeys Topic<Focus and does so without violating Stay. In cases where the DO is topical, on the other hand, DO topicalization, the in situ configuration and the A-scrambled configuration all violate at least a constraint. We argue that this accounts for the local/non-local asymmetry we see in the IcePahC concerning the relative order of the two internal objects: IOs are rarely topicalized because this operation brings no advantage over simply leaving the IO in situ. When it comes to DOs, on the other hand, DO topicalization violates the same number of constraints as the in situ and the A-scrambled con-

3Note that DO topicalization violating as many constraints as, say, the in situ configuration, and hence being equally preferred to it cannot be the end of the story: if that were the case, we would expect DO topicalization to be just as frequent as the in situ configuration, when in fact the rate of DO topicalization in our corpus is just 8%. Clearly, some ranking of the constraints displayed in Tables 1 and 2 must also be in place; in particular, Stay seems to be ranked higher than Topic<Focus. According to such an analysis, then, topicalizing the DO whenever the IO is in focus is slightly worse than leaving the DO in situ, as Stay is ranked higher than Topic<Focus. This violation is however not as bad as topicalizing the IO when the DO is in focus, since in the former case both ordering configurations violate at least a constraint. Hence the asymmetry we observe between DO and IO topicalization.
figuration, rendering DO topicalization no longer significantly worse than leaving both objects in situ. If IO topicalization is a fundamentally dispreferred strategy, why does it occur at all in our corpus? There are likely to be different factors at play. A first factor could be the desire to overtly mark contrastivity: IOs that are interpreted as contrastive topics would be moved to the left periphery because this operation makes it possible to overtly mark the domain of contrast associated with that specific proposition (see in particular Neeleman and Van De Koot (2012). A second possible trigger behind IO topicalization could be heaviness. Recall that the average length of topocalized IOs was 2.6 words, while the average length of topocalized DOs was 1.9 words. On average, topocalized IOs were thus around a word longer than topocalized DOs. In a recent paper, Indriðadóttir and Ingason (2019) found a distinct effect of heaviness on the likelihood of whether or not a constituent appears in the left periphery rather than in situ. In particular, they found that left-dislocated constituents are on average considerably longer -and hence more syntactically complex- than constituents which appear in situ. They speculated that heaviness draws constituents to the edge of the clause; not just to the right edge, as it is generally assumed (for example to account for Heavy NP Shift), but also to the left edge. Recall that we also found a marginal effect of length distinguishing the targets of IO versus DO topocalization, with IO topics being on average almost a word longer than DO topics. If heaviness does indeed draw phrases to the edges of a clause, IOs would be prime candidates for this type of dislocation: unlike DOs, the in-situ position of IOs in Icelandic is not an edge position since IOs are merged above DOs.

As the editor points out, it would be interesting to obtain a clearer picture of the discourse status of in situ DOs: what percentage of them is focal, and what percentage of them is topical? Estimating this rate however is a non-trivial manual step because such annotation is not included in the corpus. We thus leave this question for future research.
6 Conclusion

In this paper, we investigated the relative order of DOs and IOs in ditransitive constructions in Icelandic. We were particularly interested in determining whether the preference for the order IO < DO, which we observe locally (e.g., whenever the two objects both appear within the VP) is maintained whenever object topicalization dislocates either one of the two internal arguments to the left periphery. To determine whether or not this was the case, we searched the Icelandic Parsed Historical Corpus for instances of matrix-clause ditransitive constructions, annotating the results differently depending on whether IO topicalization, DO topicalization or no topicalization had taken place. We found that DO topicalization is more than twice as frequent as IO topicalization, and hence that there is an asymmetry between local and non-local configurations. We argued that the low incidence of IO topicalization follows from the combination of a series of factors: the cross-linguistic preference for having topical information first (Topic < Focus), the penalty for syntactic movement operations that can be avoided (Stay), and the fact that IOs are externally merged higher than DOs in Icelandic. IO topicalization is always dispreferred as leaving the IO in situ obeys both Stay and Topic < Focus, unlike IO topicalization, which violates Stay. On the other hand, DO topicalization is not significantly worse than simply leaving the DO in situ, as both configurations violate at least one constraint: DO topicalization violates Stay, while leaving the objects in situ violates Topic < Focus. This generates optionality wrt what configuration is selected, rendering DO topicalization more likely to occur than IO topicalization.

7 Appendix

For reproducibility: Below are the queries we used to search for IO and DO topicalization in ditransitive structures.

**Default word order**

Main clause ditransitives where finite verb precedes both objects (982 results):
IO topicalization
Main clause ditransitives where the indirect object precedes the finite verb, which precedes the direct object (39 results):

DO topicalization
Main clause ditransitives where the direct object precedes the finite verb, which
precedes the indirect object (89 results).

define:
finverb ..[PD][IS]

IP-MAT idoms finverb
IP-MAT idoms NP-OB1
IP-MAT idoms NP-OB2
IP-MAT idoms NP-SBJ
NP-OB1 domswords> 0
NP-OB2 domswords> 0
NP-SBJ domswords> 0
NP-OB1 sprec finverb
finverb sprec NP-OB2

References


