Antisocial Syntax
Disentangling the Icelandic VO/OV parameter and its lexical remains

Anton Karl Ingason¹,² Einar Freyr Sigurðsson¹,²
Joel Wallenberg¹,²,³

¹University of Pennsylvania
²University of Iceland
³Newcastle University

DiGS 14, University of Lisbon, July 6th, 2012
**Labov and Harris (1986:21)**

Abstract linguistic structure has little or no social impact on members of the community. The interface of language and society is narrow, and primarily on the surface: the words and sounds of the language.
Goals of the paper

**Empirical**

- A study on Icelandic OV-to-VO in the Icelandic Parsed Historical Corpus (IcePaHC) (expanding Hróarsdóttir 2000)
- A study on variable V2 in the Paston letters (PCEEC corpus) (Taylor et al. 2006)

**Theoretical**

- Push a strong Antisocial Hypothesis for Syntax:

  **The Antisocial Hypothesis**

  Word order is not socially evaluated, unless it can be identified with specific phonological or lexical material.

  - Implications for the architecture of the grammar
Overview

1. Introduction
   - labovharris

2. OV-to-VO in Icelandic
   - Relating style/genre and word order

3. Variable V2 in English
   - Relating audience gender and word order

4. Implications
   - Word order as a coherent empirical category

5. Conclusion
(1) a. She will a man kill
b. She will kill a man

- The evolution of OV/VO has been studied in great detail in English (Kroch and Taylor 2000; Pintzuk and Taylor 2006)
- An ongoing change in Icelandic from earliest records (12th century) to recent centuries (Hróarsdóttir (2000))
- The Icelandic Parsed Historical Corpus (IcePaHC) (Wallenberg et al. 2011) is now available
Data from IcePaHC (Wallenberg et al. 2011)

- We extracted 6378 tokens which contain a finite auxiliary, followed by a non-finite main verb and an object (including both OV/VO)
- Each token coded for:
  - **Word order, VO/OV** (dependent variable)
  - Object type (pronoun, NP=non-pronoun, quantifier)
  - Clause type (matrix/subordinate)
  - Year (exact or philological estimate)
  - Object weight, log(nr. of words)
  - **Genre** (narrative/religious)
- Analyzed using a mixed effects model with text (≈speaker) as a random factor (using Rbrul, cf. Johnson 2008, 2009)
Significant conditioning factors (p<.05)

OV favored by:

- **Object type:**
  Quantifiers > NPs > Pronouns (controlling for heaviness)

- **Clause type:**
  Subordinate > Matrix

- **Style:**
  Religious (formal) > Narratives (informal)

VO favored by:

- **Object heaviness:**
  More words in object, more surface VO

- **Time:**
  VO gradually takes over

Interaction:

- Interaction between style and object type (p=.0000137)
Overall diachronic trend of OV usage (logistic regression)
Trend by genre (religious more OV than narratives)
Next few slides:

We argue that the apparent style shift to a conservative grammar in religious texts is attached to stored lexical items rather than attitudes towards abstract word order.
Style shift and lexical items

- Social evaluation can clearly be attached to lexical items
- The situation is less clear for more abstract structures, like general word order patterns
- What if word order style shift is always tied to a stored item?
- Verb phrases (verb+object) can clearly be stored as units

(2) a. He **kicked the bucket**
   b. They **kicked the bucket**
   c. Someone **kicked the bucket**

Subject+verb relationship looser (Marantz 1984)

(3) a. **Subject verbed** him
   b. **Subject verbed** them
   c. **Subject verbed** someone
Can complements be open spaces in idiomatic phrases?

Svenonius (2005)

I have made two empirical observations about idioms [...] The first observation is that the open positions in idioms are very restricted, namely to certain specifiers and adjuncts. Idiomatic material seems to leave complement positions inaccessible. The second observation is that there is a strict division between the T-domain and the V-domain. There are many idioms in each domain, but very few which span the domain.
Any systematic exceptions?

Verb-PP idioms with an open (ACC) object position:

(4) a. ... verbed him.ACC to ...
    b. ... verbed them.ACC to ...
    c. ... verbed someone.ACC to ...

- We coded 561 entries from an Icelandic idiom collection (Sveinsson 1995) for the syntactic status of the open space
- Even under a very broad definition, complements — mostly of type (4) — are only \( \approx 10\% \) (56/561) of open spaces
Prediction

If social evaluation can only be attached to stored/storable items:

- **verb-NP** can be associated with exceptional probabilities and style-shift
- **verb-pronoun** shows no exceptional probabilities, only the true usage probabilities of the parameter

Difficult to evaluate individual verb-NP phrases for sensitivity to probabilistic idiomaticness:

(5) og ungmenni yðar skulu sjónir sjá, og yðrir öldungar skulu drauma dreyma
and young men your shall **visions** see, and your elders shall **dreams** dream
(IcePaHC; Acts; 1540.NTACTS.REL-BIB,238.61-62)

However, we can reliably separate non-pronouns from pronouns ...
The style shift comes from the non-pronouns (NPs)

Time
order
will kill a man will a man kill
np
1200 1400 1600 1800 2000
pro
1200 1400 1600 1800 2000
quant
1200 1400 1600 1800 2000
The data support the Antisocial Hypothesis (AH) for Syntax: Unambiguously productive usage of the competing grammars shows no style shift

- Pronouns and quantified show no sig. interaction of genre:year (p = 0.687) ; a CRE for style.

The AH interpretation also makes the data consistent with the Constant Rate Effect (Kroch 1989) – a truly different development of NPs and pronouns would contradict previous findings in historical syntax

Balanced cross-genre corpora like IcePaHC allow us to study style phenomena in ways that would otherwise be impossible
Variable V2 in English

Second case study, goals:

- Look at a different parameter, V2
- Better controlled social context

Experiment design:

- V2 in the Paston family letters; PCEEC corpus (Taylor et al. 2006)
- Measure usage of V2 for letters to women vs. men
- In general, usage for different audiences gives a very direct measure of style shift (cf. Bell 1984)
- Female/male-targeted speech is already sharply differentiated at the mother-to-infant level (Taylor et al. 2006)
V2 in the Paston letters, overall results

<table>
<thead>
<tr>
<th>Variant</th>
<th>Diagnostic</th>
<th>Male to Female</th>
<th>Male to Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-V2</td>
<td>XP-Subject-Verb</td>
<td>51</td>
<td>375</td>
</tr>
<tr>
<td>V2</td>
<td>XP-Verb-Subject</td>
<td>24</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td><strong>Percent V2:</strong></td>
<td><strong>32.0%</strong></td>
<td><strong>24.8%</strong></td>
</tr>
</tbody>
</table>

Table: V2 in the Paston letters, including everything (N=550)

- Not a significant difference at the $p < 0.05$ level ($\chi^2$-squared = 2.26, df = 1, $p = 0.133$)
- But apparently suggestive
Address phrases

- A disproportionate amount of V2 is triggered by idiomatic address phrases

(6) To myn rith wurchipfull broder Jon Paston be þis delyueryd. (PCEEC corpus: PASTON,I,155.043.1077)

- V2 in address phrases: 94%
- V2 elsewhere: 16%
- And those are not evenly distributed
V2 study repeated without address phrases

<table>
<thead>
<tr>
<th>Variant</th>
<th>Diagnostic</th>
<th>Male to Female</th>
<th>Male to Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-V2</td>
<td>XP-Subject-Verb</td>
<td>50</td>
<td>354</td>
</tr>
<tr>
<td>V2</td>
<td>XP-Verb-Subject</td>
<td>9</td>
<td>66</td>
</tr>
<tr>
<td><strong>Percent V2:</strong></td>
<td><strong>15.3%</strong></td>
<td><strong>15.7%</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** V2 in the Paston letters, excluding address phrases (N=479)

- The difference disappears
  \( \chi^2 \text{-squared}=0.008, \text{df}=1, \ p=0.927 \)
- The Antisocial Hypothesis is supported
A Postpositional Construction in the Paston Letters

- Two variants with demonstratives:

  (7) prey hym to gyf yow his help in this;
      (John Paston I to James Gresham; PASTON,I,58.019.409)

  (8) and I was ther-in of þe same Prioures conseill,
      (William Paston I to John Urry; PASTON,I,5.003.39)

- Two variants with wh-words (inanimate):

  (9) and was bounde for hym in cc li., of which he reherseth for
to resseyve at the begynnyng of thys terme cxx li.
      (Elizabeth Poynings/Browne (née Paston) to Agnes
      Paston; PASTON,I,207.058.1548)

  (10) honorable lettres...þe tenour wher-of I send vnto yow
      herin inclossid. (Elizabeth Poynings/Browne to John
      Paston II; PASTON,I,208.059.1564)
### $[PP\ P\ this/that\ ]\ vs.\ [PP\ there\text{-}P\ ]$

<table>
<thead>
<tr>
<th>Variant (N=905)</th>
<th>Female to Female</th>
<th>Female to Male</th>
<th>Male to Female</th>
<th>Male to Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>this/that</td>
<td>0</td>
<td>24</td>
<td>20</td>
<td>76</td>
</tr>
<tr>
<td>there-P</td>
<td>1</td>
<td>282</td>
<td>111</td>
<td>391</td>
</tr>
<tr>
<td><strong>Percent there-P:</strong></td>
<td><strong>100%</strong></td>
<td><strong>92.2%</strong></td>
<td><strong>84.7%</strong></td>
<td><strong>83.7%</strong></td>
</tr>
</tbody>
</table>

- Looking at only Male Authors: $\chi^2=0.0204$ on 1df, $p=0.886$.
- However, loglinear models can look at all three variables simultaneously (Author, Recipient, Variant):
  1. **Conditional Independence of Recipient and Variant** (given Author): $G^2 = 0.241$ on 2df, $p=0.887$
  2. **Conditional Independence of Author and Variant** (given Recipient): $G^2 = 12.7$, df=2, $p=1.75 \times 10^{-3}$

**Conclusion:** Women authors use *there*-P more than males, but there is no significant effect of Recipient.
### [PP P which/what] vs. [PP where-P]

<table>
<thead>
<tr>
<th>Variant</th>
<th>Female to Female</th>
<th>Female to Male</th>
<th>Male to Female</th>
<th>Male to Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>P which/what</td>
<td>1</td>
<td>12</td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td>where-P</td>
<td>5</td>
<td>80</td>
<td>100</td>
<td>297</td>
</tr>
<tr>
<td><strong>Percent where-P:</strong></td>
<td>83.3%</td>
<td>86.9%</td>
<td>97.1%</td>
<td>87.6%</td>
</tr>
</tbody>
</table>

- Loglinear models relating Author, Recipient, Variant:
  1. **Conditional Independence of Recipient and Variant** (given Author): $G^2 = 9.82$ on 2df, $p=0.007$
  2. **Conditional Independence of Author and Variant** (given Recipient): $G^2 = 1.79$, df=2, $p=0.409$

- **Conclusion**: No significant effect of Author sex, but authors use *where*-P more when writing to female Recipients.
  - According to Model 2, all authors are **3.76 times more likely to use where-P to a female** (model odds ratio).
On real empirical categories:

- Falling apple / Orbit of a planet (SAME)
- Whale / Fish (DIFFERENT)
- Movement / Headedness / Agreement / Phonology (???)

When we say that syntax is not socially evaluated, what do we mean by “syntax”?

- Movement / Directionality of Headedness (SAME?)
- Movement / Agreement (DIFFERENT?)

Does the Antisocial Hypothesis apply to everything that happens in “syntax” (narrow, LF, PF)?
Maybe the AH applies to narrow syntax movement

- Negative Concord shows that “syntax” is too vague a notion, because it is clearly socially evaluated.

(11) We don’t need no education (Waters 1979)

- NC involves some overt morphology (depending on the language)
- NC also certainly involves some syntax, at least in the sense of Agree (or other scope-marking device) and LF movement (cf. Haegeman and Zanuttini 1996)
- Note that in a sense the LF-movement is socially evaluated (along with morphology): you need to get the NC reading in order to get the sociolinguistic effect.
Maybe the AH applies to narrow syntax movement

Now empirical sociolinguistic observations have consequences for how we divide up the syntactic domain:

- Antisymmetry (Kayne 1994) treats left/right-headedness as XP movement (narrow syntax), rather than purely an issue of linearization (PF).
- If movement is the relevant notion, then the AH indirectly supports an antisymmetric account of head-finality and the OV-to-VO change, as proposed in Biberauer and Roberts (2005); Wallenberg (2009).
- If the V2 results are on the right track, then the AH could imply that head movement takes place in the narrow syntax (see Roberts 2010).
Is social evaluation always attached to surface forms?

- Warner (2005) discussed style shift in variable *do*-support, and hypothesized that the exceptional probabilities were attached to the surface form of the negative clitic.

- The Icelandic New Impersonal Passive Construction (e.g., Maling and Sigurjónsdóttir 2002) is plausibly socially evaluated, but so is the expletive subject *það*, which is an important part of it and a specific word.

(12) \[ Það \text{ var lamið} \text{ strákinn (NewC)} \]

\[ \text{there Expletive was beaten Passive boy.the ACC} \]

‘The boy was beaten’
Empirically, we carried out a large scale study on the OV-to-VO change in Icelandic, expanding the work of Hróarsdóttir (2000).

- The size of the data set and careful sampling for genre in IcePaHC allowed a more detailed and controlled study of the change.

- Our results support a strong version of the Antisocial Hypothesis (AH) – that word order is not socially evaluated unless it can be attached to lexical material.
Two studies on the Paston letters in late Middle English provided independent test cases:

- Variable V2 confirmed the AH, and showed an effect of lexical exceptions.
- Postpositional constructions confirmed the AH, and showed an effect of lexical exceptions.

Pushing the AH further as a working hypothesis may allow us to uncover new evidence for real empirical categories in syntax.

Hypotheses like the AH show that not only does comparative diachronic quantitative syntax have consequences for syntactic theory, but that sociolinguistics does as well.


<table>
<thead>
<tr>
<th>References III</th>
</tr>
</thead>
</table>