Evidence for Two Kinds of OV Word Order

NINJAL International Symposium on Exploiting Parsed Corpora: Applications in Research, Pedagogy, and Processing

Anthony Kroch & Beatrice Santorini
University of Pennsylvania
December 10, 2017
Preliminary Considerations
A few references on the Constant Rate Effect and Statistical Independence

The shift from Tense-final to Tense-medial word order in the history of Yiddish (Santorini 1993)
Unambiguous Tense-final cases:

(1) ven der vatr ivrit un doyts leyen kan
   if the father Hebrew and German read can

(2) ven du mir meyn kop ab shneydst
   if you me my head off cut

Unambiguous Tense-medial cases:

(3) az di nshmh zal nit oys gin
   that the soul shall not out go

(4) ven du shneydst mir meyn kop ab
   if you cut me my head off
Noun Phrase and Prepositional Phrase
Extraposition in Yiddish

(5) ven der vatr ivrit un doyts leyen kan
(6) ven der vatr leyen kan ivrit un doyts
(7) daz ikh reyn fun der ashin verde
(8) daz ikh reyn verde fun der ashin
Ambiguity between Tense-medial and Tense-final structure

(9) daz er hat eyn brudr
    that he had a brother

(10) da ishue kam in arts isral
    when Joshua came into the land of Israel
Frequency of Tense-medial versus Tense-final word order in the history of Yiddish, unambiguous cases

<table>
<thead>
<tr>
<th>Date</th>
<th>Simple verb case</th>
<th></th>
<th>Complex verb case</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tense-medial</td>
<td>Tense-final</td>
<td>freq.</td>
<td>Tense-medial</td>
</tr>
<tr>
<td>1400-1489</td>
<td>0</td>
<td>27</td>
<td>.00</td>
<td>1</td>
</tr>
<tr>
<td>1490-1539</td>
<td>5</td>
<td>37</td>
<td>.12</td>
<td>2</td>
</tr>
<tr>
<td>1540-1589</td>
<td>13</td>
<td>59</td>
<td>.18</td>
<td>4</td>
</tr>
<tr>
<td>1590-1639</td>
<td>5</td>
<td>81</td>
<td>.06</td>
<td>7</td>
</tr>
<tr>
<td>1640-1689</td>
<td>13</td>
<td>33</td>
<td>.28</td>
<td>18</td>
</tr>
<tr>
<td>1690-1739</td>
<td>15</td>
<td>20</td>
<td>.43</td>
<td>25</td>
</tr>
<tr>
<td>1740-1789</td>
<td>1</td>
<td>1</td>
<td>.50</td>
<td>11</td>
</tr>
<tr>
<td>1790-1839</td>
<td>54</td>
<td>3</td>
<td>.95</td>
<td>79</td>
</tr>
<tr>
<td>1840-1950</td>
<td>90</td>
<td>0</td>
<td>1.00</td>
<td>62</td>
</tr>
</tbody>
</table>
Frequency of Tense-medial versus Tense-final word order

Complex

Simple
Frequency of Tense-medial versus Tense-final word order in the history of Yiddish, ambiguous case

<table>
<thead>
<tr>
<th>Date</th>
<th>Possibly Tense-medial</th>
<th>Tense-final</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1400-1489</td>
<td>3</td>
<td>9</td>
<td>.25</td>
</tr>
<tr>
<td>1490-1539</td>
<td>13</td>
<td>13</td>
<td>.50</td>
</tr>
<tr>
<td>1540-1589</td>
<td>58</td>
<td>39</td>
<td>.60</td>
</tr>
<tr>
<td>1590-1639</td>
<td>41</td>
<td>29</td>
<td>.59</td>
</tr>
<tr>
<td>1640-1689</td>
<td>32</td>
<td>21</td>
<td>.60</td>
</tr>
<tr>
<td>1690-1739</td>
<td>21</td>
<td>11</td>
<td>.66</td>
</tr>
<tr>
<td>1740-1789</td>
<td>5</td>
<td>2</td>
<td>.71</td>
</tr>
<tr>
<td>1790-1839</td>
<td>58</td>
<td>3</td>
<td>.95</td>
</tr>
<tr>
<td>1840-1950</td>
<td>69</td>
<td>0</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Rise of T-medial word order, II

- simple
- ambiguous

Graph showing the rise of T-medial word order from 1445 to 1840+. The simple and ambiguous orders are compared over time.
### Frequency of DP and PP postposing in the history of Yiddish (Santorini 1993)

<table>
<thead>
<tr>
<th>Date</th>
<th>DP postposing</th>
<th>PP postposing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Postposed</td>
<td>Not postposed</td>
</tr>
<tr>
<td>1400-1489</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>1490-1539</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>1540-1589</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>1590-1639</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>1640-1689</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>1690-1739</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1740-1789</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1790-1839</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1840-1950</td>
<td>noTense-final data</td>
<td>–</td>
</tr>
</tbody>
</table>
Stability of DP and PP postposing

mean rate of PP postposing = .56
mean rate of DP postposing = .20
Correcting for postposing in ambiguous cases

For each time period, let
\[ A = \# \text{ surface ambiguous T-medial cases} \]
\[ F = \# \text{ surface T-final cases that would be ambiguous if medial} \]
\[ p = \text{rate of postposing} \]

Then the estimated true \# of T-medial cases in A is

\[ \text{Estimated I-medial} = A - F \cdot p \]

This calculation is done separately for sentences with postverbal DPs and PPs, with the results then summed.
Rise of T-medial word order, III

- simple
- ambiguous
- corrected ambiguous
- complex
The conclusion to carry forward

1. The rates of NP & PP extraposition in the Yiddish corpus are statistically independent of the rate of choice between Tense-final and Tense-medial clause structure.

2. This independence result is confirmed in other research, including work on Ancient Greek (Taylor 1994).

3. This sort of statistical independence is also reflected in the Constant Rate Effect (Kroch 1989).
Detecting stages in the transition from OV to VO in four languages: English, French, Yiddish and Icelandic
Data sources: English

Data sources: French


- Anthony Kroch and Beatrice Santorini. 2016. Penn supplement to the MCVF corpus.

- Alexei Lavrentiev, Christiane Marchello-Nizia, Céline Guillot and Serge Heiden. 2014. BFM – Base de Français Médiéval [En ligne].
Data sources: Yiddish

Data sources: Icelandic

• Joel C. Wallenberg, Anton Karl Ingason, Einar Freyr Sigurðsson and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC).
Preliminaries

• Only non–finite VPs are considered to avoid interference from V–to–C and V–to–T.

• Modals are treated as auxiliary verbs in all four languages.

• Sentences in which the direct object moves further left than T are also excluded since the “in situ” position is not recoverable.
English
Evidence for VO word order in Early Middle English

(1) oðet he habe ízettet ou al þet ȝe wulleð
    ‘until he has granted you all that you want'
    (CMANCRIW,I.68.229)

(2) þet he schulde in huden him ȝef he walde libben
    'that he should hide himself if he would live'
    (CMANCRIW,II.132.1744)
More evidence for VO word order in Early Middle English

(1) hwaso mei gan in
'whoever may go in'

(CMANCRIW,II.60.5)

(2) ha wes sone ibroht forð
'she was soon brought forth'

(CMKATHE, B.827)
More evidence for VO word order in Early Middle English

(1) worþy mennes sones þat sche myȝte han be married to 'worthy men’s sons that she might have been married to'
(CMAELR3-M23,33.189)

(2) þe terme, þe which hij ne shul nouȝt passe over 'the limit which he should not pass over'
(CMEARLPS-M2,125.5471)
Possible evidence for OV word order in Early Middle English

(1) þeos ne schulen neaver song singen song

'these should never sing songs'

(CMHALI, I42.222)

(2) þat ne have noht here sinnes forleten here sinnes

'who have not forsaken their sins'

(CMTRINIT, 67.934)
More evidence for OV word order in Early Middle English

(1) al þe blodi sunnen þet ha is wið iwundet
    'all the bloody sins that she is wounded with'
    (CMANCRIW,I.62.202)

(2) sumþing þet god maȝe of arisen
    'something that good may arise from'
    (CMANCRIW,I.74.296)
Two-argument VPs: OOV word order

(1) Ne durste nauere gume nan oðerne ufele igreten
   'Nor did a man ever dare to afflict evil on another'
   (1200-BRUT,564.1322)

(2) þatt icc have ȝuw summ del nu spelledd offe
   'which I have now told you something of'
   (CMORM-M1,1,221.1820)
Two-argument VPs: OVO word order

(1) For all þeo the habbeð any good idon me
    'For all those who have done me any good'
    (CMANCRW,I.64.212)

(2) I sal yu lere þe dute of god
    'I shall teach you the fear of God'
    (CMBENRUL-M3,2.20)
Two-argument VPs: VOO word order

(1) oðet he habe iȝetted ou al þet ȝe wulleð
    'until he has granted you all that you want'  
    (CMANCRIW,I.68.229)

(2) and wile ȝelden eche men his mede efter his werke
    'and will pay each man his reward by his work'  
    (CMLAMBI-MI,143.310)
Distribution of Full DP Objects in Double Object Clauses in Early Middle English (<1420)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>2</td>
<td>31</td>
<td>0.06</td>
</tr>
</tbody>
</table>

rate of DO scrambling = 0.03

N=35

Chi-square: .731 (ns)

Expected rate of OOV based on rates of IO and DO scrambling = .03 x .06 = 0.002

Actual rate of OOV = 1/35 = 0.03
Distribution of All Objects in Double Object Clauses in Early Middle English (<1420)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>20</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>30</td>
<td>260</td>
<td>0.1034</td>
</tr>
</tbody>
</table>

rate of DO scrambling: 0.037  
N=321

Chi-square: 62.498

Expected rate of OOV based on rates of IO and DO scrambling: 0.0038

Actual rate of OOV: 0.06
French
VO & OV word order: modal + infinitive

(1) Je veul avoir mon loier.  
'I want to have my pay.'  
(127X-CASSIDORUS-P,164.1546)

(2) Kar ne poeit le jur choisir le jur.  
'For he cannot choose the day.'  
(116X-MARIE-DE-FRANCE-R,111.2262)
(1) Rollant ad mis l' olifan a sa buche.
    'Roland raised the ivory horn to his mouth.'
    (1100-ROLAND-V,133.1772)

(2) Li reis Marsilie out sun cunseill finet sun cunseill.
    'King Marsilla had adjourned his council.'
    (1100-ROLAND-V,5.53)
Two-argument VPs: OOV word order

(1) Or *ad Deus* _saint Thomas_ cel’ _ampole_ donee.  
'Now God gave Saint Thomas this phial.'  
(1173-BECKET-P-BFM,182.14984)

(2) ainsi *pourroit Grace a Dieu querre*.  
'In this way, he could ask God for grace.'  
(1190-BORON-R-PENN,7.88)
Two-argument VPs: OVO word order

(1) Tu auoiz dous choses amises al creator.
    'You had presented two things to the creator.'
    (1190-SBERNAN-P-BFM,10.325)

(2) Ancor uolt plus grant honor faire a nostre lum.
    'He wished to do our man an even great honor.'
    (1190-SBERNAN-P-BFM,37.1192)
Two-argument VPs: VOO word order

(1) Et Pilates a douné le cors Joseph.
    'and Pilate gave the body to Joseph.'
    (1210-BORON-P-PENN,24.230)

(2) É Deu ad dune le regne a Absalon tun fils.
    'and God has given the kingdom to your son Absalom.'
    (1150-QUATRERELIVRE-P-PENN,88.3317)
Distribution of Objects in Double Object Clauses in Early Old French (<1260)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>11</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>9</td>
<td>55</td>
<td>0.14</td>
</tr>
<tr>
<td>rate of DO scrambling</td>
<td>0.10</td>
<td>N=81</td>
<td></td>
</tr>
</tbody>
</table>

Expected rate of OOV based on rates of IO and DO scrambling: \(0.14 \times 0.10 = 0.014\)

Actual rate of OOV: \(11/81 = 0.14\)

Chi-square: 18.52
Distribution of Objects in Double Object Clauses in late Old French (<1460)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>2</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>31</td>
<td>176</td>
<td>0.15</td>
</tr>
</tbody>
</table>

rate of DO scrambling

Expected rate of OOV based on rates of IO and DO scrambling: \(0.15 \times 0.09 = 0.013\)

Actual rate of OOV: \(\frac{2}{226} = 0.01\)

Chi-square: 0.276

N=226
Yiddish
VO & OV word order: modal+infinitive

(1) da velin mir vermisin di khasene
   'Then we will ruin the wedding'
   (1615E-COURT,108.80)

(2) …ver nur kan zayn gezind farshiken zayn gezind
    'whoever can send away his servants'
    (1619W-LETTERS,.16)
VO & OV word order: \( hobn + \text{participle} \)

(1) \( \ldots \text{vau min hat fergebin unzi zind} \)
\( \text{'where they have forgiven our sins'} \)
\( (1704\text{E-ELLUSH,.16}) \)

(2) \( \text{d}i \ hbn \text{ eyn yudn drmurt eyn yudn} \)
\( \text{‘who have murdered a Jew’} \)
\( (1465\text{W-COURT,16.67}) \)
Two argument VPs: OOV word order

(1) ikh hab den isral eyn tubh gtan
   'I have done the Israelites a good turn'
   (1579E-SHIR, 10.60)

(2) un mustn imrdarn dem mtsraim ir fikh hitn
   'and always had to guard the animals
    for the Egyptians'
   (1589E-ESTER, 7.123)
Two argument VPs: OVO word order

(1) sukhr habn unzri bridr gigebn fil gelt
    'Merchants gave our brothers much money'
    (1692E-VILNA,217.134)

(2) drum hat er dem menshn gebn di turh …
    'therefore he gave the people the Torah'
    (1620E-LEVTOVI,41.47)
Two argument VPs: VOO word order

(1) hat gibrakht meyn oybrstn alirley shpetsirey
    '[who] brought my boss all kinds of spices'
    (1665W-COURT,221.246)

(2) mer haben unzer formuner gegeben meinem stieffater tsvay hundert gulden
    'but our guardians gave my stepfather 200 guilders'
    (1518W-GOETZ,.137)
Distribution of Objects in Double Object Clauses in early East Yiddish (<1800)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>Rate of IO Scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>24</td>
<td>4</td>
<td>0.57</td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>5</td>
<td>3</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Expected rate of OOV based on rates of IO and DO scrambling: \(0.57 \times 0.62 = 0.357\)

Actual rate of OOV: \(24/36 = 0.667\)

Chi-square: 2.14
Distribution of Objects in Double Object Clauses in pre-contemporary East Yiddish (<1900)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>3</td>
<td>8</td>
<td>0.27</td>
</tr>
</tbody>
</table>

rate of DO scrambling: 0.11

N=22

Chi-square: 9.1

Expected rate of OOV based on rates of IO and DO scrambling

\[ 0.11 \times 0.27 = 0.030 \]

Actual rate of OOV

\[ \frac{10}{22} = 0.454 \]
Distribution of Objects in Double Object Clauses in contemporary East Yiddish (>1900)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>Rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>6</td>
<td>25</td>
<td>0.19</td>
</tr>
</tbody>
</table>

 rate of DO scrambling | 0.19 | N=39     |

Expected rate of OOV based on rates of IO and DO scrambling: \(0.19 \times 0.19 = 0.037\)

Actual rate of OOV: \(\frac{2}{39} = 0.051\)

Chi-square: 0.124
Icelandic
Distribution of Objects in Double Object Clauses in pre-contemporary Icelandic (<1900)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>41</td>
<td>27</td>
<td>0.04</td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>8</td>
<td>192</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Expected rate of OOV based on rates of IO and DO scrambling: \(0.12 \times 0.04 = 0.005\)

Actual rate of OOV: \(41/268 = 0.152\)

Chi-square: 107.6

N=268
Distribution of Objects in Double Object Clauses in contemporary Icelandic (>1900)

<table>
<thead>
<tr>
<th></th>
<th>IO&gt;V</th>
<th>V&gt;IO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DO&gt;V</strong></td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>V&gt;DO</strong></td>
<td>2</td>
<td>47</td>
</tr>
<tr>
<td>rate of DO</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>scrambling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rate of IO</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>scrambling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Expected rate of OOV based on rates of IO and DO scrambling: \(0.06 \times 0.04 = 0.002\)

Actual rate of OOV: \(0/52 = 0.000\)

Chi-square: 0.127
Fronted VPs after the subject in OF

(1) car nous rien faire ne devons
 Madagascar: 'for we must do nothing'
(1190-BORON-R-PENN,100.1575)

(2) li roi le cors mener an pu et
 Madagascar: 'the king can take the body from there'
(1170-YVAI N-R,81.2816)
Fronted VPs before the subject in OF

(1) *grant demi pied mesurer i pout hom*
   'one could measure a full half foot there'
   (1100-ROLAND-V,94.1189)

(2) *seignur servir bien deit l’um tel*
   'one should serve such a lord well'
   (1120-BRENDAN-R,55.665)
A roll-up derivation of OV word order


Rollup Example

Underlying order:

   car nous ne devons faire rien
   'because we should not do anything'

Derived order:

   car nous rien faire ne devons
Carnage, nous ne devons rien faire.
car

DPm

nous

Negk

ne devonsj

pres

NegP

tk

vP

v′

TPo

vPn

DPm

t

vi

t

rien

vPn

t

vi

t

faire inf
devons
pres
rien
faire inf
fais
devons
pres
rien
faire inf
fais
The end