Detecting Grammatical Properties in Usage Data

Anthony Kroch & Beatrice Santorini
University of Pennsylvania
March 2017
The shift from Infl-final to Infl-medial word order in the history of Yiddish (Santorini 1993)
Unambiguous I-final cases:

(1) ven der vatr  nurt doyts  leyen kan
   if  the father only German read  can

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

Unambiguous I-medial cases:

(3) ven der vatr  kan  nurt doyts  leyen

(4) ven du  shneydst  mir meyn kop  ab
Extraposition in Yiddish

(5) ven der vatr **nurt doyts** leyen kan

(6) ven der vatr leyen kan **nurt doyts**
Ambiguity between I-medial and I-final structure

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

(8) da ishue kam in arts isral
when Joshua came into the land of Israel
daz er hat i eyn bruder ti
daz er hat_{i} eyn bruder
Frequency of Infl-medial versus Infl-final word order in the history of Yiddish, unambiguous cases

<table>
<thead>
<tr>
<th>Date</th>
<th>Simple verb case</th>
<th></th>
<th></th>
<th>Complex verb case</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INFL-medial</td>
<td>INFL-final</td>
<td>freq.</td>
<td>INFL-medial</td>
<td>INFL-final</td>
</tr>
<tr>
<td>1400-1489</td>
<td>0</td>
<td>27</td>
<td>.00</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>1490-1539</td>
<td>5</td>
<td>37</td>
<td>.12</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>1540-1589</td>
<td>13</td>
<td>59</td>
<td>.18</td>
<td>4</td>
<td>51</td>
</tr>
<tr>
<td>1590-1639</td>
<td>5</td>
<td>81</td>
<td>.06</td>
<td>7</td>
<td>57</td>
</tr>
<tr>
<td>1640-1689</td>
<td>13</td>
<td>33</td>
<td>.28</td>
<td>18</td>
<td>44</td>
</tr>
<tr>
<td>1690-1739</td>
<td>15</td>
<td>20</td>
<td>.43</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>1740-1789</td>
<td>1</td>
<td>1</td>
<td>.50</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>1790-1839</td>
<td>54</td>
<td>3</td>
<td>.95</td>
<td>79</td>
<td>0</td>
</tr>
<tr>
<td>1840-1950</td>
<td>90</td>
<td>0</td>
<td>1.00</td>
<td>62</td>
<td>0</td>
</tr>
</tbody>
</table>
Frequency of Infl-medial versus Infl-final word order

1.00

0.75

0.50

0.25

0.00

1445 1515 1565 1615 1665 1715 1765 1815 1840+

Simple

Complex
Frequency of Infl-medial versus Infl-final word order in the history of Yiddish, ambiguous case

<table>
<thead>
<tr>
<th>Date</th>
<th>Possibly INFL-medial</th>
<th>INFL-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 I-medial word order, II

- Simple
- Ambiguous
- Complex

Year:
- 1445
- 1515
- 1565
- 1615
- 1665
- 1715
- 1765
- 1815
- 1840+
## Frequency of DP and PP postposing in the history of Yiddish (Santorini 1993)

<table>
<thead>
<tr>
<th>Date</th>
<th>DP postposing</th>
<th></th>
<th>PP postposing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Postposed</td>
<td>Not postposed</td>
<td>freq.</td>
<td>Postposed</td>
</tr>
<tr>
<td>1400-1489</td>
<td>1</td>
<td>12</td>
<td>.08</td>
<td>9</td>
</tr>
<tr>
<td>1490-1539</td>
<td>7</td>
<td>19</td>
<td>.27</td>
<td>13</td>
</tr>
<tr>
<td>1540-1589</td>
<td>7</td>
<td>24</td>
<td>.23</td>
<td>52</td>
</tr>
<tr>
<td>1590-1639</td>
<td>10</td>
<td>40</td>
<td>.20</td>
<td>39</td>
</tr>
<tr>
<td>1640-1689</td>
<td>4</td>
<td>19</td>
<td>.17</td>
<td>17</td>
</tr>
<tr>
<td>1690-1739</td>
<td>1</td>
<td>5</td>
<td>.17</td>
<td>6</td>
</tr>
<tr>
<td>1740-1789</td>
<td>1</td>
<td>2</td>
<td>.33</td>
<td>8</td>
</tr>
<tr>
<td>1790-1839</td>
<td>0</td>
<td>1</td>
<td>.00</td>
<td>1</td>
</tr>
<tr>
<td>1840-1950</td>
<td>no INFL-final data</td>
<td>–</td>
<td>no INFL-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

\[ A = \# \text{ surface ambiguous I-medial cases} \]
\[ F = \# \text{ surface I-final cases that would be ambiguous if medial} \]
\[ p = \text{rate of postposing} \]

Then the estimated true \# of I-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.
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).
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 įȝettet 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'
   (CMANCR IW, 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) wopȝy mennes sones þat sche myȝte han be maried 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,142.222)

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

'who have not their sinnes forsaken.'

(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 mæþe 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'
   (CMANCRIVW,1.64.212)

(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 iȝettet ou al þet ȝe wulleð
   'until he has granted you all that you want'
   (CMANCRİW, l.68.229)

(2) and wile ȝelden eche men his mede efter his werke
   'and will pay each man his reward by his work'
   (CMLAMBI-M1, l.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)

<table>
<thead>
<tr>
<th></th>
<th>Expected rate of OOV based on rates of IO and DO scrambling</th>
<th>Actual rate of OOV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.03 \times 0.06 = 0.002)</td>
<td>(1/35 = 0.03)</td>
</tr>
</tbody>
</table>
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

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

\[
0.037 \times 0.1034 = 0.0038
\]

Actual rate of OOV

\[
\frac{20}{321} = 0.06
\]

Chi-square: 62.498

N = 321
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)
VO & OV word order: *avoir* + participle

(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: \(\frac{11}{81} = 0.14\)

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 | 0.09 |

<table>
<thead>
<tr>
<th>Expected rate of OOV based on rates of IO and DO scrambling</th>
<th>(0.15 \times 0.09 = 0.013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual rate of OOV</td>
<td>(2/226 = 0.01)</td>
</tr>
</tbody>
</table>

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
   'who ever can send away his servants'
   (1619W-LETTERS, .16)
VO & OV word order: *avoir*+participle

(1) …vau min *hat fergebin unzi zind*  
'where they have forgiven our sins'  

(1704E-ELLUSH,.16)

(2) *di hbn eyn yudn drmurt eyn yudn*  
'They murdered a Jew.'  

(1465W-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) sukhritahunzribridrgigebnfilgelt
'Merchants gave our brothers much money'
(1692E-VILNA,217.134)

(2) drumhaterdemmenshngebnditurh...
'therefore has he the people given the Torah'
(1620E-LEVTOV1,41.47)
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></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>5</td>
<td>3</td>
<td>0.62</td>
</tr>
</tbody>
</table>

| rate of DO scrambling | 0.57 | N=36 |

**Expected rate of OOV based on rates of IO and DO scrambling**  
\[0.57 \times 0.62 = 0.357\]

**Actual rate of OOV**  
\[\frac{24}{36} = 0.667\]

**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: \(2/39 = 0.051\)

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>
<th>Rate of DO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>41</td>
<td>27</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>8</td>
<td>192</td>
<td></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: \(\frac{41}{268} = 0.152\)

Chi-square: 107.6

Chi-square:

\(N=268\)
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>
<th>rate of IO scrambling</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO&gt;V</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>V&gt;DO</td>
<td>2</td>
<td>47</td>
<td>0.04</td>
</tr>
<tr>
<td>rate of DO scrambling</td>
<td>0.06</td>
<td>N=52</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 riens feire ne devons
    'for we must do nothing'
    (1190-BORON-R-PENN,100.1575)

(2) li rois le cors mener an puet
    'the king can take the body from there'
    (1170-YVAIN-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


der Artikel

C
  \text{dass}

TP
  \text{hat}
    \text{Tense}
    vP_{n}
      T'
        \text{Tense}
          t
        vP_{n}
          T'
            \text{Tense}
              t

vP_{n}
  \text{gestern}
  \text{Part}_{m}
    \text{Part}_{l}
      vP_{l}
        \text{nicht}
          DP
            Johann
              \text{t}_{j}
                v'
                  Part
                    Part
                      vP_{l}
                        v'
                          \text{gekauft}_{i}
                            \text{t}_{i}
                              \text{das Buch}
                            \text{DP}
                              \text{t}_{i}
Car nous allons faire cela.

Diagramme de CTDP:

- **CP**
  - **C**
  - **TP**
    - **DP-p**
      - **car**
    - **nous**
      - **vP-n**
        - **InfP-m**
          - **vP-l**
            - **DP-p**
              - **t**
            - **t-j**
              - **RootP**
                - **faire-i**
                  - **v**
            - **t-i**
              - **DP**
                - **rien**
          - **Inf’**
            - **Inf**
              - **v’**
                - **3**
                  - **vP-n**
                    - **T’**
                      - **Tense**
                        - **Neg**
                          - **ne**
                            - **devons-k**
                              - **t-j**
                                - **vP-n**
                                  - **t**
                      - **Tense**
                        - **t-j**
              - **InfP-m**
                - **t**
            - **Inf’**
              - **Inf**
                - **v’**
                  - **2**
                    - **InfP-m**
                      - **vP-l**
                        - **t**
        - **Inf’**
          - **Inf**
            - **v’**
              - **1**
                - **InfP-m**
                  - **vP-l**
                    - **t**
            - **Inf**
              - **v’**
                - **InfP-m**
                  - **vP-l**
                    - **t**

The end