Conclusions: Dative & Passive Alternations in Real English

[Linguistics 128/228: Real English: The Syntax of Language Use]

1 The Syntax of Language Use

Two motivations:

—Grammaticality judgments of constructed examples may be systematically biased by the probability of similar descriptions of the event types depicted by the examples.

—The statistical structure of language use may reveal generalizations that lie beneath the threshold of grammaticality judgments.

2 Statistical structure in syntax:

—the same distributions of properties (such as pronominality, person) that are categorical in some languages are gradient in other languages, including English

—the same properties of information structure, animacy, and length that distinguish subjects from nonsubjects also distinguish core datives (NPs) from oblique datives (PPs) in English


3 Example: Categorical Effects of Person on Voice

The effects of the person hierarchy on grammar are categorical in some languages, most famously in languages with inverse systems, but also in languages with person restrictions on passivization.

In Lummi, for example, the person of the subject argument cannot be lower than the person of a nonsubject argument. If this would happen in the active, passivization is obligatory; if it would happen in the passive, the active is obligatory (Jelinek and Demers 1983, 1994).
Lummi is a dialect of Coast Salish spoken in British Columbia.

**Lummi examples:**

* * ‘The man knows me/you’

χ̃i-t-η=son/=sx*' c̄ c̄ sway?qo? 
know-TR-PASS=1/2.SG.NOM by the man
‘I am/you are known by the man’

χ̃i-t=son/=sx* c̄ sway?qo? 
know-TR=1/2.SG.NOM the man
‘If you know the man’

* * ‘The man is known by me/you’

---

When both the agent and patient are third person, the passive is optional. *

**Lummi examples:**

χ̃i-t-s c̄ sway?qo? c̄ swi?qoʔaŋ 
know-TR-3.TR.SUBJ the man the boy
‘The man knows the boy’

χ̃i-t-ŋ c̄ swi?qoʔaŋ c̄ n c̄ sway?qo? 
know-TR-PASS the boy by the man
‘The boy is known by the man’

---

Similar constraints appear in languages unrelated to Salish. An example is Picur’, a dialect of Northern Tewa spoken in New Mexico (Zaharlick 1982: 40–1; Mithun 1999: 226–8).

**Picur’ is examples:**

* * ‘The man saw me.’

Ta-mɔn-miaʔ ʔən 
1SG SUBJ ANIM-see-PASS-PAST man-OBL
‘I was seen by the man.’

Sɔnene tì-mɔn-ʔən. 
man 1SG SUBJ ANIM OBJ-see-PASS 
‘I saw the man.’

* Sɔnene mɔn-miaʔ ʔən n̥n-pə. 
man see-PASS-PAST 1SG-OBL 
‘The man was seen by me.’

---

When both the agent and patient are third person, the passive is optional.

**Picur’ is examples:**

Sɔnene mɔnʔən. 
man see-PAST
‘He saw him.’

Mɔn-miaʔ ʔən sɔnene-pə. 
see-PASS-PAST man-OBL
‘He was seen by the man.’

---

*Third persons are unmarked except as objects of a verb with local person subject.

---

*When both the agent and patient arguments are local persons (first or second), the active is obligatory. According to Jelinek and Demers (1994: 714), Lummi pronouns are clitics restricted to subject and object functions; the oblique function may be filled by third person deictic expressions designating speaker and hearer. The latter have determiners and behave syntactically like full nominals; they are not subject to the person-voice restrictions.
English person/role by voice (full passives):

<table>
<thead>
<tr>
<th>action</th>
<th># Act</th>
<th># Pass</th>
<th>% Act</th>
<th>% Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>179</td>
<td>0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1,2</td>
<td>6246</td>
<td>0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>3110</td>
<td>39</td>
<td>98.8</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>472</td>
<td>14</td>
<td>97.1</td>
<td>2.9</td>
</tr>
</tbody>
</table>

The leftmost column gives the four types of inputs (local person acting on local, local acting on nonlocal, etc.). We estimate the number of times each input was evaluated as the number of actives plus passives with that person/structure association. We then calculate the rate of passivization as the number of times that input was realized as passive.

The person/voice interaction is highly significant (2-sided Fisher’s exact test, p < 0.0001). Similar significance levels result if short passives are included, but we omit them because the person of the agent is not always clear.\(^a\)

\(^a\)See Dingare (2001) for detailed analysis and methodological discussion.

Compared to the rate of passivization for inputs of third persons acting on third persons (1.2%), the rate of passivization for first or second person acting on third is substantially depressed (0%) while that for third acting on first or second (2.9%) is substantially elevated.

The same disharmonic person/argument associations which are avoided categorically in languages like Lummi by making passives either impossible or obligatory, are avoided in the SWITCHBOARD corpus of spoken English by either depressing or elevating the frequency of passives relative to actives.

In sum, the ‘hard’ grammatical constraints on person/voice interactions seen in languages like Lummi and Picurí continue to show up as statistical preferences in English.
Soft ‘Topicality’ Effects on English Active/Passive Choice: the Switchboard Corpus

The parsed Switchboard corpus is not tagged for topicality or givenness, but we can approximate this information-status concept by comparing the distributions of more and less definite nominal expression types, such as pronouns, proper names, definite and indefinite noun phrases.

In the Treebank Switchboard corpus local person pronouns are plentiful, but the distribution of pronouns and lexical (= nonpronominal) NPs is highly skewed (Francis et al. 1999):
- 91% of subjects are pronominal
- 66% of objects are lexical

We found significant ‘topicality’ effects on passivization in Switchboard, following the methods of Dingare (2001: 19–23).

<table>
<thead>
<tr>
<th>Pt → Pronoun</th>
<th>Non-Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag ↓ Pronoun</td>
<td>A: 2457 P: 4 (0.16%)</td>
</tr>
<tr>
<td>Non-Pronoun</td>
<td>A: 179 P: 18 (9.14%)</td>
</tr>
</tbody>
</table>

A: Active count; P: Passive count; (n%): percent passives

* one-sided Fisher exact, $p < 0.05$
* one-sided Fisher exact, $p < 0.0001$

Hard ‘Topicality’ Effects in Lummi

Lummi categorically avoids pronominal objects with non-pronominal subjects (Jelinek and Demers 1983, 1994):

* "The man knows it."
  \( \chi\text{-ti-η} \text{ a co swayʔqəʔ} \) 
  know-TR-PASS by the man
  ‘It is known by the man’
  \( \chi\text{-ti-s} \text{ co swayʔqəʔ} \) 
  know-TR-3,TR.SUBJ the man
  ‘He knows the man.’


Stochastic Optimality-theoretic Grammars

Partial stochastic grammar of English:

Partial stochastic grammar of Lummi:

\*Obl₁₂ \*S₃ₚₛ \*S₃ \*O₃ \*O₁₂ \*S₃ \*S₃ \*O₁₂ \*S₃ \*O₃