Incremental syntactic prediction in the comprehension of Swedish

Thomas Hörberg
Department of Linguistics,
Stockholm University

Assumptions – Grammatical Relations

- express argument functions (actor/undergoer, topic/focus)
- grammatical encoding conditioned on prominence (e.g. Silverstein 1976)

- Animacy: human < animate < inanimate
- Person: first, second < third
- Referentiality: pronoun < proper name < common noun
- Definiteness: definite < specific indefinite < unspecific indefinite

- Actor > Undergoer in prominence
- Exceptions (i.e. Undergoer > Actor) typologically marked and infrequent in discourse
Assumptions – Argument interpretation

- Assignment of argument FUNCTIONS (Actor / Undergoer)

- Highly incremental process that is probabilistic and frequency-driven, i.e. draws upon statistical regularities in the input

- Morphosyntactic and prominence information serve as **Argument Interpretation Cues (AICs)**

(MacWhinney & Bates 1989)  (Bornkessel-Schlesewsky & Schlesewsky 2009)

Objectives

- Corpus-based modeling of incremental argument interpretation in Swedish transitive sentences
- Experimentally test model predictions using SPR
Corpus properties

Svensk Trädbank: balanced written Swedish texts

<table>
<thead>
<tr>
<th>Corpus Genre</th>
<th>N texts</th>
<th>N sentences</th>
<th>N words</th>
<th>N hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press: reportage</td>
<td>44</td>
<td>7278</td>
<td>106079</td>
<td>1492</td>
</tr>
<tr>
<td>Press: Editorial</td>
<td>17</td>
<td>2385</td>
<td>40887</td>
<td>473</td>
</tr>
<tr>
<td>Press: Reviews</td>
<td>27</td>
<td>3961</td>
<td>66002</td>
<td>712</td>
</tr>
<tr>
<td>Skills, Trades and Hobbies</td>
<td>58</td>
<td>8933</td>
<td>134947</td>
<td>1840</td>
</tr>
<tr>
<td>Popular Lore</td>
<td>48</td>
<td>5625</td>
<td>109665</td>
<td>1503</td>
</tr>
<tr>
<td>Belles Letters, Biography, Memoirs</td>
<td>26</td>
<td>3598</td>
<td>61297</td>
<td>805</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>70</td>
<td>10847</td>
<td>163333</td>
<td>1540</td>
</tr>
<tr>
<td>Learned and Scientific Writing</td>
<td>83</td>
<td>9833</td>
<td>192827</td>
<td>1809</td>
</tr>
<tr>
<td>General fiction</td>
<td>82</td>
<td>13028</td>
<td>191507</td>
<td>3110</td>
</tr>
<tr>
<td>Mysteries and Science fiction</td>
<td>19</td>
<td>4070</td>
<td>45321</td>
<td>826</td>
</tr>
<tr>
<td>Light reading</td>
<td>20</td>
<td>2908</td>
<td>46126</td>
<td>749</td>
</tr>
<tr>
<td>Humor</td>
<td>6</td>
<td>1071</td>
<td>14428</td>
<td>248</td>
</tr>
<tr>
<td>Brochure texts</td>
<td>25</td>
<td>1733</td>
<td>23122</td>
<td>390</td>
</tr>
<tr>
<td>Newspaper texts</td>
<td>28</td>
<td>1669</td>
<td>24125</td>
<td>361</td>
</tr>
<tr>
<td>Educational texts</td>
<td>14</td>
<td>1624</td>
<td>25623</td>
<td>374</td>
</tr>
<tr>
<td>Debate articles</td>
<td>18</td>
<td>1134</td>
<td>23476</td>
<td>316</td>
</tr>
</tbody>
</table>

SVO sentences: 15679
OVS sentences: 872

Sentence properties

<table>
<thead>
<tr>
<th>Adverbial position 1</th>
<th>Adverbial position 2</th>
<th>Word order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnen får inte äta upp all glass innan middan</td>
<td>barnen inte får åta upp all glass innan middan</td>
<td>SVO</td>
</tr>
<tr>
<td>Innan middan får barnen inte äta upp all glass</td>
<td>Innan middan får inte barnen åta upp all glass</td>
<td>VSO</td>
</tr>
<tr>
<td>All glass får barnen inte åta upp innan middan</td>
<td>All glass får inte barnen åta upp innan middan</td>
<td>OVS</td>
</tr>
</tbody>
</table>

- NP:s of any length
- Up to 4 verbs
- Adverbials + verb particles optional
NP1 + NP2 properties

- Animacy: animate vs. inanimate
- Givenness: given vs. new
- Definiteness: definite vs. indefinite
- Number: singular vs. plural
- Egophoricity: 1st / 2nd vs. 3rd person
- Pronominality: pronoun vs. noun
- Case: unmarked vs. nominative vs. accusative

Verb semantic properties (annotation)

- Verbs assign role semantic properties to NP arguments, i.e. "degree of Actorhood" (e.g. Dowty 1991, Primus 2006)
- General verb semantic "entailments" therefore included

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Volitionality</td>
<td>Vollitional involvement</td>
<td>Control – volitionality and intentionality</td>
</tr>
<tr>
<td>Experience</td>
<td>Undergoer</td>
<td>Possession</td>
</tr>
<tr>
<td>Movement</td>
<td>Change of state</td>
<td>Causally affected</td>
</tr>
<tr>
<td>Physical involvement</td>
<td>Causally affected</td>
<td>Physical involvement</td>
</tr>
<tr>
<td>Possession</td>
<td>-</td>
<td>Possession</td>
</tr>
</tbody>
</table>
Assumptions – Argument interpretation

• assignment of argument functions <-> determining word order

• In most cases unproblematic:

  Under hot fick [han NP1] hjälpa [rånarna NP2] att ta sig in
  "Under threat he got to help the robbers to get in"

• ... but not always:

  [Vakthunden NP1] släppte in [tjuven NP2]
  "The watchdog let the burglar in"

  [Byggherren NP1] kände [hon NP2] litet grann
  "The constructor she knew somewhat"

Corpus distributions of GRs in written Swedish

• Investigates the distribution of GRs with respect to AICs and word order
  - distributional differences between subjects & objects
  - distributional differences between subject- and object-initial sentences
    → the functional motivation(s) of the object initial word order

• Logistic Mixed Effects modeling
  - Quantify strengths of individual AICs
  - Predict probabilities of sentence word order conditional on sentence properties
Quantifying AIC strengths

- Logistic mixed effects modeling
  \[
  \log\left(\frac{P(Y)}{1-P(Y)}\right) = \beta_0 + \beta_1(X_1) + \beta_2(X_2) \ldots + \beta_n(X_n) + Z_{1,m} \ldots Z_{m,m} \sim N(0, \sigma^2\Sigma)
  \]
  
  \(Y\): outcome – word order
  
  \(X_n\): predictors – pronominality, definiteness, animacy etc…
  
  \(Z_i\): random effect(s) – subgenre

- Estimates cue strengths in terms of (log) odds ratios
- Predicts the probability for OVS on the basis of AICs
- Object experiencer verb class excluded
- Verb interactions selected on basis of backward elimination

---

Model estimates of AIC strengths
Probability of OVS given AIC cues

[De NP1] har [vapen som väl matchar alla övriga nationers i regionen NP2] p < 0.001
"[The NP1] have [weapons that match those of other nations in the region NP2]"

[Islänningarna från turridningsfirman Eldhestar på Island NP1] hjälpte [mig NP2] mycket p = 0.20
"[The Icelanders from the riding firm Eldhestar on Iceland NP1] helped [me NP2] alot"

[Den utvalda grenen NP1] sågade [han NP2] av en bit från änden p = 0.34
"[The selected branch NP1] [he NP2] sawed off at some distance from the end"

[Det NP1] förklarade [den sydkoreanske USA-ambassadören Hyun Hong Choo NP2] p = 0.68
"[That NP1] [the Southkorean ambassador Hyun Hong Choo NP2] explained"

*[Några planer på att stänga polishuset NP1] har [han NP2] inte* p = 0.99
"[Any plans of shutting down the police department NP1] [he NP2] does not have"

Modeling incremental argument interpretation

- Models the on-line change in the expectation of an object-initial word order (i.e. surprisal, Levy 2008) given AICs provided by constituents over time
- surprisal of OVS modeled in terms of relative entropy / the Kullback-Leibler divergence: \( D_{KL}(P || Q) = \Sigma i \log(P_i/Q_i)P_i \)
- Based upon the probability of OVS at different time points
  - Baseline model: \( p(\text{OVS} | \text{Subgenres}) \)
  - NP1 model: \( p(\text{OVS} | \text{NP1}, \text{Subgenre}) \)
  - NP1 + verb model: \( p(\text{OVS} | \text{NP1}, \text{Verb}, \text{Subgenre}) \)
  - Full model: \( p(\text{OVS} | \text{NP1}, \text{Verb}, \text{NP2}, \text{Subgenre}) \)
Modeling incremental argument interpretation

Probabilities

- baseline model: \( p(OVS) : \sim 0.05 \)
- NP1 model: \( p(OVS \mid NP1) \)
- NP1 + verb model: \( p(OVS \mid NP1 + verb) \)
- full model: \( p(OVS \mid NP1 + verb + NP2) \)

surprisals

- surprisal NP1: \( D_{KL}(p(OVS \mid NP1) \mid p(OVS)) \)
- surprisal verb: \( D_{KL}(p(OVS \mid NP1 + verb) \mid p(OVS \mid NP1)) \)
- surprisal NP2: \( D_{KL}(p(OVS \mid NP1 + verb + NP2) \mid p(OVS \mid NP1 + verb)) \)

surprisal in original data

- OVS sentences with initial lexical NP and final case marked NP

[Gallant Flowering_\text{NP1}] [varnade_\text{verb}] [vi_\text{NP2}] för kraftigt
\( D_{KL} = 0.03 \) \( D_{KL} = 0.00 \) \( D_{KL} = 7.61 \)

"[Gallant Flowering_\text{NP1}] [we_\text{NP2}] [warned_\text{verb}] about heavily"

[En härligt örtkryddad soppa_\text{NP1}] [käkade_\text{verb}] [vi_\text{NP2}] bland annat
\( D_{KL} = 0.32 \) \( D_{KL} = 0.63 \) \( D_{KL} = 0.55 \)

"[A wonderful soup spiced with herbs_\text{NP1}] [we_\text{NP2}] [ate_\text{verb}] among other things"
surprisal across constituents
• lexical NP1, 3rd person plural pronoun NP2
• Cumulative evidence for OVS at NP1 and verb is surprising
• Low surprisal at NP2 when confirmed (OVS), high surprisal when disconfirmed (SVO)

surprisal at NP1
• Unmarked & allophoric NP1
**surprisal at verb**

- Lexical singular one word NP1
- Moderate surprisal for Inanimate & Volitional / Causative + Experiencer

**surprisal at NP2 (disambiguation towards OVS)**

- Unmarked & allophoric NP1
- case marked, egophoric & short NP2 that disambiguates sentence towards OVS
Testing the model

- Testing the strongest model predictions
- Self-paced reading
- Reading times assumed to reflect processing

Self paced reading

- Dependent variable: time latency between button presses
- Analyses done on region RTs rather than word RTs
- Task: Comprehension question following each sentence

Sparkar jag bollen mitt upp i kryset?
Predictions – sentence differences

<table>
<thead>
<tr>
<th>WO</th>
<th>Verb class</th>
<th>Animacy</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volitional</td>
<td>Inanimate</td>
<td>Bollen sparkar jag mitt upp i krysset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Animate</td>
<td>Killen sparkar jag mitt på småbadet</td>
</tr>
<tr>
<td>OVS</td>
<td></td>
<td>Inanimate</td>
<td>Bollen glömmer jag mitt på fotbollsplanen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Animate</td>
<td>Killen glömmer jag sent på kvällen</td>
</tr>
<tr>
<td></td>
<td>Volitional</td>
<td>Inanimate</td>
<td>Jag sparkar bollen mitt upp i krysset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Animate</td>
<td>Jag sparkar killen mitt på småbadet</td>
</tr>
<tr>
<td>SVO</td>
<td></td>
<td>Inanimate</td>
<td>Jag glömmer bollen mitt på fotbollsplanen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Animate</td>
<td>Jag glömmer killen sent på kvällen</td>
</tr>
</tbody>
</table>

Regions
- e.g. /Bollen sparkar reg.1 / jag mitt reg.2 / upp i reg.3 / krysset FW /
- Region 1 RTs correspond to verb surprisal, region 2 RTs to NP2 surprisal

Results – RT differences

Word order
- Faster RTs in SVO sentences vs. OVS sentences at reg. 1 & reg. 2
- No significant RT differences in SVO sentences
Conclusions from SPR experiment

• Ambiguous sentences read slower than unambiguous
• AICs to some extent redundant following disambiguation
• Animacy functions as a cue in argument interpretation
• Animacy interacts with verb class – stronger effect w. volitional verbs
• Unsuprising results(?)
• more fine-grained predictions should be tested (e.g. neurophysiological correlates)

Additional properties (annotation)

NP1 + NP2 length: [continuous]
NP1 + NP2 text deixis: deictic vs. non-deictic
Sentence type: initial NP vs. initial adverbial
Embedded clause: main vs. embedded clause
Auxiliary verb(s): single verb vs. auxiliaries
## Final corpus - example

<table>
<thead>
<tr>
<th>HeadV</th>
<th>Particle</th>
<th>Int. adv.</th>
<th>Embedded</th>
<th>VVol</th>
<th>VEsp</th>
<th>VOExp</th>
<th>VExp</th>
<th>VCaus</th>
<th>VPoss</th>
<th>VPass</th>
<th>Aux.</th>
</tr>
</thead>
<tbody>
<tr>
<td>har</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>har</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>berätter</td>
<td>N</td>
<td>M</td>
<td>V</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s'ingat</td>
<td>hmm</td>
<td>Y</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lyft</td>
<td>Y</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gär</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>C</td>
<td>x</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>har</td>
<td>N</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>viärter</td>
<td>N</td>
<td>E</td>
<td>V</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>har</td>
<td>N</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glina</td>
<td>N</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>C</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fråga</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fråt</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>utvecklat</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avlossa</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>C</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>har</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hörta</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysera</td>
<td>N</td>
<td>M</td>
<td>V</td>
<td>E</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>har</td>
<td>Y</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hugga</td>
<td>N</td>
<td>M</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>infia</td>
<td>N</td>
<td>M</td>
<td>V</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

## Subjects outrank objects in prominence

<table>
<thead>
<tr>
<th>Subject</th>
<th>Initial</th>
<th>Final</th>
<th>OVERALL</th>
<th>Final</th>
<th>Initial</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animacy:Animate</td>
<td>77%</td>
<td>90%</td>
<td>78%</td>
<td>19%</td>
<td>8%</td>
<td>18%</td>
</tr>
<tr>
<td>Animacy:Inanimate</td>
<td>23%</td>
<td>10%</td>
<td>22%</td>
<td>81%</td>
<td>92%</td>
<td>82%</td>
</tr>
<tr>
<td>Givenness:Given</td>
<td>71%</td>
<td>74%</td>
<td>71%</td>
<td>30%</td>
<td>63%</td>
<td>32%</td>
</tr>
<tr>
<td>Givenness:New</td>
<td>29%</td>
<td>26%</td>
<td>25%</td>
<td>70%</td>
<td>37%</td>
<td>68%</td>
</tr>
<tr>
<td>Definiteness:Definite</td>
<td>86%</td>
<td>82%</td>
<td>80%</td>
<td>49%</td>
<td>76%</td>
<td>50%</td>
</tr>
<tr>
<td>Definiteness:Indefinite</td>
<td>20%</td>
<td>18%</td>
<td>20%</td>
<td>51%</td>
<td>24%</td>
<td>50%</td>
</tr>
<tr>
<td>Number:Singular</td>
<td>73%</td>
<td>77%</td>
<td>74%</td>
<td>68%</td>
<td>83%</td>
<td>69%</td>
</tr>
<tr>
<td>Number:Plural</td>
<td>27%</td>
<td>23%</td>
<td>26%</td>
<td>32%</td>
<td>17%</td>
<td>31%</td>
</tr>
<tr>
<td>Egophoricity:Epiphenic</td>
<td>27%</td>
<td>46%</td>
<td>36%</td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>Egophoricity:Allophonic</td>
<td>73%</td>
<td>55%</td>
<td>72%</td>
<td>97%</td>
<td>99%</td>
<td>97%</td>
</tr>
<tr>
<td>Pronoun:Pronominal</td>
<td>54%</td>
<td>71%</td>
<td>55%</td>
<td>16%</td>
<td>46%</td>
<td>18%</td>
</tr>
<tr>
<td>Pronoun:Lexical</td>
<td>46%</td>
<td>29%</td>
<td>45%</td>
<td>84%</td>
<td>54%</td>
<td>82%</td>
</tr>
<tr>
<td>Case:Marked</td>
<td>48%</td>
<td>61%</td>
<td>49%</td>
<td>10%</td>
<td>2%</td>
<td>10%</td>
</tr>
<tr>
<td>Case:Unmarked</td>
<td>52%</td>
<td>33%</td>
<td>51%</td>
<td>90%</td>
<td>98%</td>
<td>90%</td>
</tr>
<tr>
<td>Text deixis:Direct</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>39%</td>
<td>4%</td>
</tr>
<tr>
<td>Text deixis:Not</td>
<td>98%</td>
<td>100%</td>
<td>98%</td>
<td>98%</td>
<td>61%</td>
<td>96%</td>
</tr>
</tbody>
</table>
Discourse topicalization

- objects more frequently given, definite, pronominal and text deictic when fronted

<table>
<thead>
<tr>
<th></th>
<th>Final object</th>
<th>Fronted object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given</td>
<td>30%</td>
<td>63%</td>
</tr>
<tr>
<td>Definite</td>
<td>49%</td>
<td>76%</td>
</tr>
<tr>
<td>Pronominal</td>
<td>16%</td>
<td>46%</td>
</tr>
<tr>
<td>Text deictic</td>
<td>2%</td>
<td>39%</td>
</tr>
</tbody>
</table>

- object fronting is used when the object is a discourse topic e.g.
  “Det kan man visa med ett enkelt experiment”
  “Denna form av blodshämnd kunde kungamakten stoppa”

Sentence topic and contrastive focus

- Not the full story however. Fronted objects more commonly new, indefinite and lexical than subjects

<table>
<thead>
<tr>
<th></th>
<th>Initial subject</th>
<th>Fronted object</th>
<th>Final object</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>29%</td>
<td>37%</td>
<td>70%</td>
</tr>
<tr>
<td>Indefinite</td>
<td>20%</td>
<td>24%</td>
<td>51%</td>
</tr>
<tr>
<td>Lexical</td>
<td>46%</td>
<td>54%</td>
<td>84%</td>
</tr>
</tbody>
</table>

- object fronting also used when
  a) the object is the sentence topic express sentence topic:
     “En sammanfattning på hela arbetet gör jag i kapitel 10”
  b) the object expresses contrastive focus:
     “Oregano kan man också tänka sig”
Predictable subjects

- subjects more frequently animate, egophoric, pronominal &
  person in sentence final position

<table>
<thead>
<tr>
<th>Subject</th>
<th>Initial subject</th>
<th>Final subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>77%</td>
<td>90%</td>
</tr>
<tr>
<td>Egophoric</td>
<td>27%</td>
<td>45%</td>
</tr>
<tr>
<td>Pronominal</td>
<td>54%</td>
<td>71%</td>
</tr>
<tr>
<td>Marked / Person</td>
<td>48%</td>
<td>67%</td>
</tr>
</tbody>
</table>

- Object fronting more frequent when the subject is a "highly predictable discourse topic" (Teleman et al. 1999:4:341-343)

Prominence & verb class predicting OVS

<table>
<thead>
<tr>
<th>Volitionality</th>
<th>Experience</th>
<th>D.experience</th>
<th>Causation</th>
<th>Possession</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Given</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-1.7%</td>
<td>0.2%</td>
<td>-1.7%</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Given</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-1.7%</td>
<td>0.2%</td>
<td>-1.7%</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Marked / Person</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-1.7%</td>
<td>0.2%</td>
<td>-1.7%</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total</td>
<td>0.0%</td>
<td>0.0%</td>
<td>5.4%</td>
<td>-0.7%</td>
<td>-3.2%</td>
</tr>
</tbody>
</table>

MP1

- Given - 0.6% 0.1% 0.8% -0.2% 1.2% 0.0% -0.1% 0.0% -0.8% 0.1% -0.6%
- New 1.7% -0.3% -1.7% 0.3% -2.7% 0.0% 0.9% 0.0% 1.9% -0.3% 1.3%
- Given -0.1% -0.5% -0.1% -0.1% -0.1% 0.0% 0.1% -0.1% 0.0% -0.1% 0.0%
- New 1.7% -0.3% -1.7% 0.3% -2.7% 0.0% 0.9% 0.0% 1.9% -0.3% 1.3%
- Marked / Person -0.1% 0.0% 0.3% -0.1% 1.7% 0.0% -0.5% 0.1% -0.2% 0.0% -4.7%
- New 23.1% -2.0% 42.6% -1.1% -5.4% 0.1% 13.2% -0.7% -2.6% 1.1% 12.9%
- Total -0.8% 0.0% 0.4% -0.1% 0.7% 0.0% 0.0% 0.0% -0.9% 0.2% -0.3%
- New 0.1% 0.0% -1.7% 0.2% -2.6% 0.0% 0.1% 0.0% 3.9% -0.7% 1.1%

MP2

<table>
<thead>
<tr>
<th>Volitionality</th>
<th>Experience</th>
<th>D.experience</th>
<th>Causation</th>
<th>Possession</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Given</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-0.5%</td>
<td>0.1%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Given</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-0.5%</td>
<td>0.1%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Marked / Person</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>New</td>
<td>-0.5%</td>
<td>0.1%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total</td>
<td>0.0%</td>
<td>0.0%</td>
<td>5.4%</td>
<td>-0.7%</td>
<td>-3.2%</td>
</tr>
</tbody>
</table>

2016-09-20 / Thomas Hörberg, Department of Linguistics

Stockholm University
Quantifying AIC strengths

• Many prominence features are highly correlated e.g.,
  - case × egophoricity
  - definiteness × givenness

• Percentages does not account for this

• Need to quantify AIC strengths while controlling for influence of other AICs + other properties

surprisal in fakedata

• Fakedata set with all attested co-occurrences of predictor values

• surprisal in the full population of possible sentence types (N = 833504)
Fakedata – vectors of attested combinations predictor values

<table>
<thead>
<tr>
<th>NP1Given</th>
<th>NP1Anim</th>
<th>NP1Def</th>
<th>NP1Num</th>
<th>NP1Ego</th>
<th>NP1Pro</th>
<th>NP1Deixis</th>
<th>NP1Case</th>
<th>NP2Given</th>
<th>NP2Anim</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

surprisal at NP1

- Low surprisal overall, some surprisal for inanimate / text deicitic NP1
Conclusions from corpus studies

• Strongest NP1 AICs:
  - animacy, text deixis, markedness and definiteness
• Strongest NP2 AICs:
  - animacy, egophoricity, pronominality and definiteness
• But it is the interactions between AIC:s and verb semantics that really matter
  - Inanimate NP1/animate NP2 with volitional or causative verb predicts OVS
  - Unmarked NP1 with experiencer verb predicts OVS
  - Indefinite NP1/given NP2 with possessive verb predicts OVS
• Sentences w. object experiencer verbs
  - reversed pattern: OVS more frequent w. high prominent subjects and low prominent objects

Conclusions from corpus studies

• Animacy × volitionality / causation
  - Physical actions often caused by animate beings
  - A volitional / intentional action requires an animate being
• Case × experiencer
  - SAP / egophoric subjects highly frequent w. experiencer verbs (Dahl 2000)
  - Experiencer verbs frequently express private knowledge & subjective experiences (e.g. "know", "think", "see", "feel") from the perspective of the speaker (Dahl 2000)
  - Case marked pronouns most frequently 1st or 2nd person
• Definiteness / givenness × possession
  - possessums more frequently discourse new and sentence topic?
    "Ingen skam i kroppen har de, ingen skam!"
    "No shame they have, no shame!"
Limitations

- Context limitation
  - results only valid in the context of transitive sentences
- Representativeness
  - written corpus preferable
  - genre limitations
- Disambiguating information (e.g. case)
  - several models used but none is optimal

Psycholinguistic experiment
Self paced reading

Bollen ******* *** **** *** # *******
Self paced reading

####### sparkar ### #### ### # #######

Self paced reading

####### ####### jag #### ### # #######
Self paced reading

######## ######## ### mitt ### # #######

Self paced reading

######## ######## ### ### upp # #######
Self paced reading

######## ######## ### #### ### i ########

Self paced reading

######## ######## ### #### ### # krysset
Results – RT differences

Animacy
- Region 1 RTs in OVS significantly slower w. inanimate NP1 vs. animate NP1
- Region 2 RTs in OVS w. VOLITIONAL VERBS ONLY significantly faster w. inanimate NP1 vs. animate

Experimental procedure
References


