Eye-tracking data analyzed using Generalized Additive Mixed Modeling

Atty Schouwenaars¹, Esther Ruigendijk¹ & Petra Hendriks²
(¹University of Oldenburg, ²University of Groningen)
Introduction

- Adults and children’s processing of subject, object and passive *which*-questions.

- Object questions, in which the object precedes the subject, are difficult to comprehend for children (among others; Friedmann & Novogrodsky, 2011; Biran & Ruigendijk, 2015; Roesch & Chondrogianni, 2015).
Examples wh-questions disambiguated by Case

- **Subject question**
  1) Welcher Esel wäscht den Pinguin?  
  Which-NOM donkey-SG wash-SG the-ACC Penguin-SG  
  ´Which donkey washes the penguin?´

- **Object question**
  2) Welchen Esel wäscht der Pinguin?  
  Which-ACC donkey-SG wash-SG the-NOM Penguin-SG  
  ´Which donkey does the penguin wash?´

23.03.2016
Eye-tracking data analyzed using Generalized Additive Mixed Modeling - Atty Schouwenaars
Examples wh-questions disambiguated by Agreement

- **Subject question**
  3) Welche *Giraffe* wäscht die *Kühe*?
  
  Which-∅ giraffe-SG wash-SG the-∅ cow-PL
  
  ´Which giraffe washes the cows?´

- **Object question**
  4) Welche *Giraffe* waschen die *Kühe*?
  
  Which-∅ giraffe-SG wash-PL the-∅ cow-PL
  
  ´Which giraffe does the cows wash?´
Examples wh-questions disambiguated by Agreement and Case

- **Subject question**
  5) Welche Füchse waschen den Schwan?
  Which-Ø fox-PL wash-PL the-ACC swan-SG
  ‘Which foxes wash the swan?’

- **Object question**
  6) Welche Füchse wäscht der Schwan?
  Which-Ø fox-PL wash-SG the-NOM swan-SG
  ‘Which foxes does the swan wash?’

Eye-tracking data analyzed using Generalized Additive Mixed Modeling - Atty Schouwenaars
Passive questions

- **Object Question**
  - *Welchen* Esel wäscht *der* Pinguin?

- **Passive Question**
  - *Welcher* Esel *wird von dem* Pinguin *gewaschen?*
    - ‘Which Dunkey is washed by the penguin?’
Research questions

1. How do children and adults process *which*-questions?

2. Do different cues lead to differences in processing in terms of eye-gaze patterns?
### Incremental optimization

- * constraint violation
- ! fatal violations
- ✨ optimal output

#### Stage I

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe … which-AMB giraffe.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Subject-question interpretation</td>
</tr>
<tr>
<td>b</td>
<td>Object-question interpretation</td>
</tr>
</tbody>
</table>

**AGREEMENT:**
The verb agrees with the subject

**SUBJECT-FIRST:**
The subject comes first

#### Stage II

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe wäscht … which-AMB giraffe.SG wash.SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Subject-question interpretation</td>
</tr>
<tr>
<td>b</td>
<td>Object-question interpretation</td>
</tr>
</tbody>
</table>

**AGREEMENT**

**SUBJECT-FIRST**

#### Stage III

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe wäscht die Kühe which-AMB giraffe.SG wash.SG the-AMB cow.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Subject-question interpretation</td>
</tr>
<tr>
<td>b</td>
<td>Object-question interpretation</td>
</tr>
</tbody>
</table>

**AGREEMENT**

**SUBJECT-FIRST**

* constraint violation

! fatal violations

✨ optimal output
### Incremental optimization

- * constraint violation
- ! fatal violations
- ☑ optimal output

#### Stage I

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe … which-AMB giraffe.PL</th>
<th>AGREEMENT: The verb agrees with the subject</th>
<th>SUBJECT-FIRST: The subject comes first</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑</td>
<td>a Subject-question interpretation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b Object-question interpretation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Stage II

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe waschen … which-AMB giraffe.SG wash.PL</th>
<th>AGREEMENT</th>
<th>SUBJECT-FIRST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a Subject-question interpretation</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>☑</td>
<td>b Object-question interpretation</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

#### Stage III

<table>
<thead>
<tr>
<th></th>
<th>Welche Giraffe waschen die Kühe which-AMB giraffe.SG wash.PL the-AMB cow.PL</th>
<th>AGREEMENT</th>
<th>SUBJECT-FIRST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a Subject-question interpretation</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>☑</td>
<td>b Object-question interpretation</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Predictions

- Differences between initial and final interpretations based on incremental optimization for object questions and passives.
  - Different patterns for subject questions vs. object questions and passive questions.
  - Differences in cues: object questions with ambiguous case morphology vs. object questions disambiguated with case in the beginning of the sentence.

23.03.2016
Eye-tracking data analyzed using Generalized Additive Mixed Modeling - Atty Schouwenaars
Participants

- 30 Adults, 14 male (19-30, Mean:24)
- 36 Children, 22 male (7;05-10;09, Mean: 9;01)
Experiment

- Picture selection task with eye-tracking

- 4 lists, 54 test items
- Tobii TX300, two computer setup

- Case
- Agr
- Agr+Ca

- Subject qs
- Object qs
- Passive qs
Experiment

1. Fixation cross

2. Preview (2500ms)

3. Fixation cross

4. Picture + sentence
results left out
for more information please write me

atty.schouwenaars@uni-oldenburg.de
Gaze data predictions

- Incremental process
  - interpretations may change over time

■ Like Linear Mixed Effects Models GAMMs also allow for an inclusion of different types of effects.
  ▪ By including random factors, the experimental noise, such as the effect of preceding items or participant-related preferences can be brought under statistical control (Baayen et al., 2008).

■ The crucial difference is that GAMMs manage **nonlinear datasets**
Smooth functions

- GAMMs model the relation between a dependent variable and a predictor in terms of a **smooth function**.
- Smooth functions can be seen as a set of base functions that have different shapes and different weights. Together they can fit patterns (linear as well as nonlinear) of the data.
- In order to prevent that the data becomes overgeneralized and overfitted, the smooth functions and parameters are designed in terms of estimation processes (Wood 2006).
Preprocessing of data

- **Areas of interest**
  - Target (red)
  - Distractor (yellow=blue)
  - Not an AOI
  - Unvalid
Research questions

1. How do children and adults process *which*-questions?

2. Do different cues lead to differences in processing in terms of eye-gaze patterns?
Preprocessing for Analysis

- Timebins of 200ms
  - Advantages:
    - makes the data frame smaller (reduces runtime model)
    - Reduces the correlation between subsequent measures (in comparison with raw binomial data).
Dependent variable

- The input was **TDDiff** (the difference between the sum of looks towards the target minus the sum of looks towards the distractor picture) for Timebins of 200 ms.

```r
> pr_AOI1 = with(data,aggregate(AOI1,by=list(Subject=Subject, AgeMonths=AgeMonths, Sentence=Sentence, TypeOfCue=TypeOfCue, nrCue=nrCue, EarlyAvailabilityCue=EarlyAvailabilityCue, Condition=Condition, TypeOfQuestion=TypeOfQuestion, Timebin200=Timebin200, Item.ACC=Item.ACC, Group=Group),sum))

> names(pr_AOI1)[length(names(pr_AOI1))] <- "nrRed"

> data$TDDiff <- (data$nrRed-data$nrYellow)
```
<table>
<thead>
<tr>
<th>Subject</th>
<th>Age Months</th>
<th>Sentence</th>
<th>TypeOfCue</th>
<th>TypeOfQues</th>
<th>Timebin</th>
<th>Group</th>
<th>nrRed</th>
<th>nrYellow</th>
<th>TDDiff</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>115</td>
<td>Sent1</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>60</td>
<td>-60</td>
</tr>
<tr>
<td>101</td>
<td>115</td>
<td>Sent2</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>60</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>101</td>
<td>115</td>
<td>Sent3</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>60</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>101</td>
<td>115</td>
<td>Sent4</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>60</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>101</td>
<td>115</td>
<td>Sent5</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>60</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>101</td>
<td>115</td>
<td>Sent6</td>
<td>AgCa</td>
<td>object</td>
<td>100</td>
<td>1</td>
<td>17</td>
<td>41</td>
<td>-24</td>
</tr>
</tbody>
</table>

Eye-tracking data analyzed using Generalized Additive Mixed Modeling - Atty Schouwenaars
Predictors & Random factors effects

- **Predictors:** All interactions between Group (Adults or Children) and Type of Question (subject, object, passive) were combined into one predictor.

```r
> data$QuestionGroup = interaction(data$TypeOfQuestion, data$Group)
```

- **Random effects:** An interaction of Participant and Type of Question was used as a random effect factor.

```r
> data$SubjectQuestion = interaction(data$Subject, data$TypeOfQuestion)
```
Best model

\[
\text{\texttt{m5c = bam(TDDiff ~ s(Timebin200,by=QuestionGroup) + QuestionGroup + s(Timebin200,SubjectQuestion, bs='fs',m=1), data=data1, gc.level=2, method='ML')}}
\]

<table>
<thead>
<tr>
<th>A. parametric coefficients</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>20.929</td>
<td>11.727</td>
<td>17.848</td>
<td>0.0743</td>
</tr>
<tr>
<td>QuestionGrouppassive.1</td>
<td>58.016</td>
<td>16.689</td>
<td>34.764</td>
<td>0.0005</td>
</tr>
<tr>
<td>QuestionGroupsubject.1</td>
<td>103.545</td>
<td>16.644</td>
<td>62.211</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>QuestionGroupobject.2</td>
<td>69.235</td>
<td>17.443</td>
<td>39.691</td>
<td>0.0001</td>
</tr>
<tr>
<td>QuestionGrouppassive.2</td>
<td>120.631</td>
<td>17.567</td>
<td>68.667</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>QuestionGroupsubject.2</td>
<td>117.532</td>
<td>17.567</td>
<td>66.904</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(Timebin200):QuestionGroupobject.1</td>
<td>70.336</td>
<td>80.114</td>
<td>160.984</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200):QuestionGrouppassive.1</td>
<td>76.508</td>
<td>84.797</td>
<td>235.884</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200):QuestionGroupsubject.1</td>
<td>56.784</td>
<td>67.175</td>
<td>176.074</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200):QuestionGroupobject.2</td>
<td>72.170</td>
<td>81.596</td>
<td>214.662</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200):QuestionGrouppassive.2</td>
<td>72.569</td>
<td>81.985</td>
<td>229.765</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200):QuestionGroupsubject.2</td>
<td>65.208</td>
<td>75.637</td>
<td>158.119</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>s(Timebin200,SubjectQuestion)</td>
<td>5.485.142</td>
<td>17.760.000</td>
<td>0.9975</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 1: Model of gaze data per different type of question (subject, object, passive) per group (1=children, 2=adults). m5c = bam(TDDiff~s(Timebin200,by=QuestionGroup) + QuestionGroup + s(Timebin200,SubjectQuestion, bs='fs',m=1), data=data1, gc.level=2, method='ML')
Smooth plot

> plotSmooths99(m5c,"Timebin200","QuestionGroup", eegAxis=F,colors=c('blue', 'grey'), dropRanef="SubjectQuestion", main='m5c smooths per QuestionType per Group, ci=99%')
functions: `get_difference` and `find_difference`

Adults vs. Children

Subject question (SVO)  Object question (OVS)  Passive question

Figure 1: Difference plots per Type of Question (subject, object and passive questions). An area indicated with red means a significant difference between children’s and adults’ gaze.
results left out
for more information please write me
1. How do children and adults process *which*-questions?

2. Do different cues lead to differences in processing in terms of eye-gaze patterns?
Differences between cues

- Dependent variable: TDDiff for Timebins of 200 ms.
- Only focused on the object questions
- All interactions between Group (Adults or Children), TypeOfCue (AgCa, Agre, Case) were combined into one predictor.
- Participant was used as a random effect factor.
results left out
for more information please write me
conclusions left out
for more information please write me
GAMMs and eye-tracking: Jacolien van Rij
e-mail: vanrij.jacolien “at” gmail.com

useful package itsadug

useful article (supplements available upon request)