Implicit Events
for Event/Entity-Ambiguous Nouns

Alessandra Zarcone, Sebastian Padó

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
SFB 732 - D6

20/07/2010
Outline

1. Covert events
   - Traditional account
   - Our questions
   - Our agenda
   - Our experiments

2. Self-paced reading study
   - Description

3. Web study
   - Web experiment 1
   - Web experiment 2
Some verb-object pairs require the recovery of covert events (CE):

<table>
<thead>
<tr>
<th>Entity-denoting objects (EN) vs. event-denoting objects (EV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV: begin the <strong>afternoon</strong></td>
</tr>
<tr>
<td>→ ✓ begin(afternoon)</td>
</tr>
<tr>
<td>→ × begin(CE(afternoon))</td>
</tr>
<tr>
<td>EN: begin the <strong>newspaper</strong></td>
</tr>
<tr>
<td>→ × begin(newspaper)</td>
</tr>
<tr>
<td>→ ✓ begin(CE(newspaper))</td>
</tr>
</tbody>
</table>

1. **the trigger problem**
   - sortal trigger hypothesis: CEs are triggered by the ontological type of the object (EN vs. EV) (cfr. Pustejovsky 1995, Jackendoff 1997);

2. **the range problem**
   - qualia structure hypothesis: CEs are retrieved from the *qualia structure* of the noun’s lexical entry (Pustejovesky 1995)
   - range: 1-2 CEs, e.g. **newspaper** → **writing** or **reading**
the trigger problem

type clash only factor responsible for evoking CEs?

Joe is a famous wrestler. He really enjoyed the fight → enjoy(fight)
Joe is a wrestling fan. He really enjoyed the fight → enjoy(\textit{watch}(fight))

Joe began the translation → begin(\textit{translation}), begin(\textbf{CE}(\textit{translation}))

the range problem

the qualia structure hypothesis undergenerates the CE range
(Lascarides and Copestake 1998, Egg 2005)

I need some help with my food → buying, finding, eating, cooking

My goat eats anything. He really enjoyed your book.
Covert events

Our agenda

1 the trigger problem
Traxler et al. (2002), Self-paced reading (SPR) study
- begin-verbs vs. spot-verbs and EN objects vs. EV objects
- longer reading times at Obj. +1 position for CE condition:
  interaction Verb * Object:
  \[ RT_{\text{begin,EN}} > RT_{\text{begin,EV}} \approx RT_{\text{spot,EV}} \approx RT_{\text{spot,EN}} \]

1 the Obj + 1 problem: contexts are not uniform: after school / the reporter told us / David said to Joan / and returned to her room...

2 ... and what about EN/EV type objects?

2 the range problem
how to study recovered CEs in a behavioral experiment?

<table>
<thead>
<tr>
<th>SPR</th>
<th>Web studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ on-line processing</td>
<td>− off-line measures</td>
</tr>
<tr>
<td>+ more natural tasks</td>
<td>− metalinguistic analysis</td>
</tr>
<tr>
<td>− CEs in absentia</td>
<td>+ CE elicitation</td>
</tr>
</tbody>
</table>
Covert events

Our experiments

1. **the trigger problem:**
   - a SPR study to verify the sortal trigger hypothesis (designed, to do)
     - 1. EN/EV objects
     - 2. the Obj + 1 problem (now even more crucial)
   - Web experiments (done)
     - 1. evaluate materials for SPR
     - 2. verify the sortal trigger hypothesis

2. **the range problem:**
   Web experiment 2 to study recovered CEs (done)

Web experiments delivered using the crowdsourcing paradigm (Snow et al. 2008), for fast and affordable collection of judgments over the web.
Self-paced reading study
Description and materials

**aim:** the trigger problem:
verify the sortal trigger hypothesis

**materials:** 10 noun triplets x 6 conditions = 60 sentences:
EN: Jim began/spotted the magazine from the camp on the hill.
EV: Al began/spotted the ceremony from the camp on the hill.
EN/EV Nick began/spotted the conquest from the camp on the hill.

**design:** 2 (begin-verbs and spot-verbs) x 3 (EN, EV, EN/EV)

**triplet selection:** triplets balanced for
- object length and frequency (Francis and Kucera 1967)
- co-occurrence frequency with begin-verbs and spot-verbs (ukWaC corpus, Ferraresi et al. 2008)
- Web experiment 0: threefold expert annotation check
  - Krippendorff’s $\alpha = .71$ (very good agreement)
  - weighted $\alpha = .79$
**Web experiment 1**

**Description**

**aim:** evaluation of materials for SPR
- check for non-expert annotation of objects (EN, EV, EN/EV)
- effect of Obj + 1 position on sortal categorization

**participants:** 14 participants from the US

**procedure:** web-based annotation experiment
Jan enjoyed **the automobile**
(possible answers: EN, EV; they could check either or both)

**materials:** 60 sentences for the SPR study, in 3 “Obj + 1” contexts:
- **no Obj + 1:**
  Jan enjoyed the automobile **short Obj + 1:**
  Jan enjoyed the automobile **on the premises full Obj + 1:**
  Jan enjoyed the automobile **on the premises of the company**
reasonably good agreement (weighted $\alpha = .52$)
very good agreement with the Gold Standard ($\alpha = .70$, weighted $\alpha = .79$)
rulled out effect of the Obj + 1 context and of the verb on the sortal type assigned to the object

- Binomial logistic regression model 1: $entity \sim context + verb$
  Context: binomial $p = .3621 \rightarrow \text{no effect}$
  Verb: $z = 1.491, p = .1359 \rightarrow \text{no effect}$

- Binomial logistic regression model 2: $event \sim context + verb$
  Context: binomial $p = .6138 \rightarrow \text{no effect}$
  Verb: $z = -0.504, p = .614 \rightarrow \text{no effect}$
Web experiment 2

Description

**aim:**
1. **the trigger problem** evaluate correlation between EN/EV and CE/noCE
2. **the range problem** elicit CEs and explore their range

**participant:** 15 participants from the US

**procedure:** Jan enjoyed **the automobile**

*does the sentence involve an additional activity that is not mentioned in the sentence?*  
(answers: *additional activity* or *no additional activity*)  
when they answered *additional activity*, participants were asked to provide examples

**materials:** same of Web experiment 1 and SPR
Web experiment 2
Results

Agreement and accuracy:

- rather low agreement
  \[(\alpha = .35, \text{ruling out EN/EV ambiguous objects, } \alpha = .36)\]
- good agreement with the Gold Standard \[(\alpha = .60)\]

Binary decision *additional activity/no additional activity* (CE/noCE):

- significant effect of object_type and verb_type with interaction
  - Binomial logistic regression model: \[\text{answer} \sim \text{obj}_\text{type} \ast \text{verb}_\text{type};\]
  - Obj_type: binomial \( p < .001 \) \( \rightarrow \) significant effect
  - Verb_type: \( z = -8.322, p < .001 \) \( \rightarrow \) significant effect
  - Interaction: binomial \( p < .001 \) \( \rightarrow \) significant effect

... does this confirm the sortal trigger hypothesis?
→ item-wise analysis
Web experiment 2
Results

<table>
<thead>
<tr>
<th>% Answers</th>
<th>begin,EN/EV</th>
<th>spot,EN/EV</th>
<th>begin,EN</th>
<th>spot,EN</th>
<th>begin,EV</th>
<th>spot,EV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no_CE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Alessandra Zarcone, Sebastian Padó
Implicit Events for Event/Entity-Ambiguous Nouns
# Web experiment 2

A closer look: verbs

<table>
<thead>
<tr>
<th>Verb Type</th>
<th>Verb</th>
<th>CE</th>
<th>no CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>begin-verb</td>
<td>continued</td>
<td>0.71</td>
<td>0.29</td>
</tr>
<tr>
<td>begin-verb</td>
<td>began</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>begin-verb</td>
<td>finished</td>
<td>0.56</td>
<td>0.44</td>
</tr>
<tr>
<td>begin-verb</td>
<td>started</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>begin-verb</td>
<td>tried</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>begin-verb</td>
<td>enjoyed</td>
<td>0.38</td>
<td>0.62</td>
</tr>
<tr>
<td>begin-verb</td>
<td>preferred</td>
<td>0.32</td>
<td>0.68</td>
</tr>
<tr>
<td>spot-verb</td>
<td>considered</td>
<td>0.29</td>
<td>0.71</td>
</tr>
<tr>
<td>begin-verb</td>
<td>ended</td>
<td>0.22</td>
<td>0.78</td>
</tr>
<tr>
<td>spot-verb</td>
<td>remembered</td>
<td>0.18</td>
<td>0.82</td>
</tr>
<tr>
<td>begin-verb</td>
<td>endured</td>
<td>0.16</td>
<td>0.84</td>
</tr>
<tr>
<td>spot-verb</td>
<td>recalled</td>
<td>0.15</td>
<td>0.85</td>
</tr>
<tr>
<td>spot-verb</td>
<td>disdained</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>begin-verb</td>
<td>savored</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>spot-verb</td>
<td>discussed</td>
<td>0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>spot-verb</td>
<td>approved</td>
<td>0.02</td>
<td>0.98</td>
</tr>
<tr>
<td>spot-verb</td>
<td>reviewed</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>organized</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>spotted</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>prepared</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Web experiment 2

**A closer look: VPs**

<table>
<thead>
<tr>
<th>type V</th>
<th>type obj</th>
<th>v</th>
<th>obj</th>
<th>CE</th>
<th>no CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>begin-verb</td>
<td>EN</td>
<td>began</td>
<td>the newspaper</td>
<td>0.89</td>
<td>0.11</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN/EV</td>
<td>began</td>
<td>the breakfast</td>
<td>0.81</td>
<td>0.19</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN</td>
<td>tried</td>
<td>the tent</td>
<td>0.73</td>
<td>0.27</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN</td>
<td>enjoyed</td>
<td>the automobile</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EV</td>
<td>continued</td>
<td>the season</td>
<td>0.46</td>
<td>0.54</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN/EV</td>
<td>enjoyed</td>
<td>the translation</td>
<td>0.39</td>
<td>0.61</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN</td>
<td>remembered</td>
<td>the brandy</td>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EV</td>
<td>enjoyed</td>
<td>the conference</td>
<td>0.24</td>
<td>0.76</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EV</td>
<td>considered</td>
<td>the debate</td>
<td>0.22</td>
<td>0.78</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN/EV</td>
<td>disdained</td>
<td>the blessing</td>
<td>0.21</td>
<td>0.79</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EV</td>
<td>remembered</td>
<td>the revolt</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN/EV</td>
<td>remembered</td>
<td>the shower</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN</td>
<td>savored</td>
<td>the butter</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN</td>
<td>disdained</td>
<td>the portrait</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EN/EV</td>
<td>endured</td>
<td>the shower</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EV</td>
<td>endured</td>
<td>the revolt</td>
<td>0.03</td>
<td>0.97</td>
</tr>
<tr>
<td>begin-verb</td>
<td>EV</td>
<td>ended</td>
<td>the ceremony</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EV</td>
<td>discussed</td>
<td>the expedition</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN/EV</td>
<td>prepared</td>
<td>the harvest</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EV</td>
<td>prepared</td>
<td>the holiday</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN</td>
<td>spotted</td>
<td>the magazine</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>spot-verb</td>
<td>EN</td>
<td>prepared</td>
<td>the package</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Web experiment 2
Range of elicited CEs

CEs elicited per VP item: range 1-15, mean 5

EN: start the portrait → 10 CEs: paint (x20), draw (x4), critique (x3), hang (x2), model (x2), sketch (x2), admire, fix, pose for, review

EN/EV: finish the harvest → 15 CEs: gather (x5), collect (x4), plan (x3), reap (x3), sell (x3), load (x2), store (x2), cook, eat, enjoy, jar, package, pick, pull, ship

EV: enjoy the conference → 4 CEs: attend (x3), hold (x2), participate in, watch
Web experiment 2

Range of elicited CEs

CEs elicited per VP item per participant: range 1-6, mean 1.4

EN: begin the newspaper → 6 CEs: deliver, edit, print, read, sell, write

EN/EV: begin the breakfast → 3 CEs: cook, eat, serve

EV: continue the season → 3 CEs: coach, play, watch
Web experiment 2

Range of elicited CEs

CEs elicited per VP when participants gave only one answer:
range 1-7, mean 3.2;

**EN:** consider the butter → 6 CEs: eat (x4), add, buy, churn, cook with, eat, make, melt

**EN/EV:** prefer the collection → 6 CEs: view (x3), buy, discuss, polish, study, watch

**EV:** start the semester → 3 CEs: spend, teach, join
Conclusions

The sortal trigger hypothesis and the qualia structure hypothesis do not seem enough to explain CE recovery:

1 sortal trigger hypothesis:
   - CEs for begin-verbs with EV obj
   - also for EN/EV obj, the EV reading does not block CE recovery
   - no clear-cut distinction between begin-verbs and spot-verbs

2 qualia structure hypothesis:
   - wide range of CEs elicited per context
   - the range is also fairly wide when participants only give one answer
   - elicited CEs > qualia structure events

Plausibility hypothesis: plausibility-driven CEs retrieval
Thank you!


