

# PUTTING ASPECTUAL COERCION IN CONTEXT – EVIDENCE FROM EYETRACKING DURING READING

OLIVER BOTT, PROJECT *COMPOSITION IN CONTEXT*  
ABSTRACT FOR THE ANNUAL XPRAG.DE WORKSHOP 2015 IN GÖTTINGEN

## 1. INTRODUCTION

The nature of contextual influences on compositional processes is central for long lasting debates about the architecture of the semantics-pragmatics interface (see, e.g., Recanati 2010). The following two hypotheses express two alternative views on the nature of the interaction between sentence interpretation and the information contained in the larger discourse context: a) Modularity: in the absence of context dependent lexical items (presupposition triggers, indexicals, etc.) sentence meaning is processed in isolation from pragmatic information; contextual information is only considered for the repair of an otherwise uninterpretable configuration. b) Interactivity: semantic and discourse pragmatic information has the same status, that is, contextual information percolates into the composition of complex meaning right from the start. The present study employed the phenomenon of aspectual coercion as a tool to study the online processing of compositional processes. We focussed on the interplay of sentential and contextual information during the online resolution of temporary and global aspectual mismatch. Our aim was to find out whether the compositional operations in the sentence have immediate access to the preceding discourse context ( $\triangleright$  interactivity), or, alternatively, whether temporary aspectual mismatch triggers aspectual coercion as a cognitively costly repair process ( $\triangleright$  modularity).

- (1) # Peter joggte in fünf Minuten.  
# Peter jogged in five minutes.

When uttered out of the blue, sentence (1) is uninterpretable. The *in-adverbial* requires a telic event predicate of the accomplishment type (Vendler 1957, Dowty 1979), but *Peter jogged* expresses an atelic activity. However, if the sentence makes reference to a spatially bounded path argument (2), the sentence is perfectly well formed.

- (2) Peter joggte in fünf Minuten durch den ganzen Park.  
Peter jogged in five minutes through the whole park.

To address the modularity and interactivity hypotheses, our currently running study investigates the processing of sentences such as (1) in a discourse context which introduced a bounded path argument in the preceding discourse. A translated sample item is provided in (3).

- (3) Half a year ago Peter started to jog a distance of four kilometers every day. When he started he was quite slow but he is becoming faster and faster.  
 Als er es heute schaffte, in nur einer halben Stunde zu joggen, freute er sich sehr.  
 When he it today managed in only half an hour to jog, he was very happy.

The interactivity hypothesis predicts immediate availability of the path argument from the context and hence no difficulty when composing the verb of the target sentence and the adverbial. The modularity hypothesis, however, led us to expect a temporary semantic mismatch at the underlined critical region *to jog*, followed by contextually driven repair by aspectual enrichment of the activity into the accomplishment *jog a distance of four kilometers*. We thus expected to observe a slow-down relative to an aspectual control condition. As aspectual controls, we included *for-modification* of the activity in combination with an atelic context, illustrated in (4).

- (4) Half a year ago Peter started to jog every day. When he started he could barely jog for ten minutes long but he is becoming better and better.  
 Als er es heute schaffte, eine ganze halbe Stunde zu joggen, freute er sich sehr.  
 When he today managed for half an hour to jog, he was very happy.

To be able to compare the effects of the aspectual enrichment condition with the effects of plain aspectual mismatch we also included a mismatch condition. Here, we combined the atelic context of the control condition (4) with the *in-modified* target sentence (3).

While the modularity and the interactivity hypothesis clearly differ with respect to the predicted time course of context effects in aspectual enrichment, both hypotheses predict that (in the case of modularity: after successful repair) the aspectual enrichment condition should receive a sensible interpretation eventually. We therefore ran two offline acceptability rating experiments showing that the sentences (Exp. 2) and discourses (Exp. 1) in fact have the intended interpretations.

## 2. INTERPRETATION DATA

After constructing 24 items in the aspectual enrichment, control and mismatch conditions, we ran two acceptability rating experiments with 30 and 20 participants, respectively. The first experiment tested the discourses in the three conditions. The task was to provide a discourse sensibility judgment on a scale from 1 (= complete nonsense) to 7 (= perfectly sensible). The experiment also included 33 clearly nonsensical filler discourses and 33 clearly sensible ones. The materials were identical to those used in the eyetracking study. The statistical analysis of the ratings revealed that the aspectual enrichment condition was fully acceptable: on average, enrichment was rated 5.1 relative to 5.3 for control (no sign. difference between conditions  $t_1(29) = 1.65, p = .11$ ;  $t_2(23) = .76, p = .46$ ) and 5.3 for the sensible fillers. Mismatch, by contrast, was perceived as nonsensical and received a mean rating of 2.6 (ANOVAs comparing all three conditions  $F_1(1, 29) = 89.7, p < .001$ ;  $F_2(1, 23) = 122.6, p < .001$ ), which was roughly at the same level as the nonsensical fillers (mean rating 2.4). The interpretation data thus confirm that the contextual support in the aspectual enrichment condition is as strong as outlined in the introduction and that the telic target sentences do not fit an atelic context.

The second offline experiment confirmed that without supporting context the telic target sentences were not fully well-formed, but required further contextual support. We presented the target sentences in isolation (together with 50 nonsensical and 50 sensible filler sentences) with both kinds of adverbials and again gathered acceptability ratings on a seven point scale. As expected, target sentences with *for-adverbials* were rated significantly better than sentences with *in-adverbials* (mean ratings: 5.9 vs. 4.2; paired t-tests  $p_{1/2} < .001$ ). However, the target sentences with *in-adverbials* were still rated significantly better than the nonsensical fillers (mean rating: 2.3) suggesting that participants were well aware of the fact that the sentences with *in-adverbials* might turn out to be well-formed given the right kind of context whereas the semantically odd fillers contained aspectual (e.g., *win for five minutes*) and tense violations (e.g., *tomorrow came*) beyond any chance of contextual support.

### 3. EYETRACKING DATA

So far, we have conducted preliminary statistical analyses on the eyetracking data of the first 25 participants from the running experiment<sup>1</sup>. Participants read discourses and provided yes-no sensibility judgments after each trial while we monitored their eye movements during reading. In addition, we assessed participants' reading span (Daneman & Carpenter 1980) to investigate whether verbal working memory capacity interacts with modular/interactive processing of the aspectual enrichment cases. We hypothesized that fully interactive interpretation will demand more working memory capacity than more local modular processing because for the former the entire discourse representation has to be held active in working memory. It is thus plausible that processing costs due to aspectual enrichment and reading span are negatively correlated with each other. The analysis of eyetracking data contingent on reading span may also offer a better understanding of the nature of any potential effects in the enrichment condition. If processing costs are due to a resolvable semantic mismatch as predicted by the modularity hypothesis we expect to find processing difficulty independently of working memory capacity across high and low span readers. If, on the other hand, the contextual information has to be (re-)activated due to working memory limitations we expect to find rather strong correlations between working memory capacity and any potential effects in the enrichment condition. At the moment, the data set is still too small for any correlational analyses and we have to postpone this discussion to the workshop. Nevertheless, the analysis of the first 25 participants already revealed some interesting findings:

Aspectual enrichment was judged sensible 80% of the time, somewhat lower than control, which received 89% 'yes, sensible' judgments (difference not significant in GLMER analysis  $p = .15$ ). Aspectual mismatch was rejected as uninterpretable 69% of the time, similar to the 75% rejection rate of the nonsensical fillers. Thus, participants perceived an aspectual mismatch in the mismatch condition and computed aspectually enriched interpretations of the target sentences in the aspectual enrichment condition.

The preliminary analysis of first pass times revealed a reliable mismatch effect already at the first part of the adverbial region *in nur* relative to *ganze* (t-tests mismatch vs.

---

<sup>1</sup>At the workshop we will present the full set of data with at least 36 participants. We would like to point out that this is preliminary data, and therefore only highlight some effects already reliable in this relatively small data set.

control  $t_1(24) = 2.2, p < .05$ ;  $t_2(23) = 2.2, p < .05$ ). Mismatch took on average more than 50 ms more to read than control. The enrichment condition showed the same numerical trend, which did not reach significance ( $t_1(24) = 1.9, p = .07$ ;  $t_2(23) = 1.4, p = .17$ ). These effects could of course also be driven by the lexical differences between the two kinds of adverbials. However, since the verb is highly predictable in the discourse conditions tested here we may expect to see effects related to aspectual composition already at the adverbial. The preliminary analysis of total times of the critical verb region revealed processing costs of the aspectual mismatch condition (control: mean RT 399ms; mismatch: mean RT 547ms;  $t_1(24) = 6.1, p < .01$ ;  $t_2(23) = 5.1, p < .01$ ). A mismatch effect was also present in the proportion of regressions out of the verb region during first pass reading (mean first pass regression ratios: control 16%, mismatch 31%; GLME analysis:  $p < .05$ ). Interestingly, also the aspectual enrichment condition had marginally longer total times than control (enrichment: mean RT 472ms;  $t_1(24) = 2.1, p < .05$ ;  $t_2(23) = 1.7, p = .10$ ). If this effect should turn out to be reliable in the final data set, too, this would provide evidence for encapsulated processing of the target sentence in line with the modularity hypothesis. However, at the current stage of the project it is still too early to draw any conclusions from the eyetracking data.

#### REFERENCES

- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- David Dowty (1979). *Word meaning and Montague Grammar*. Dordrecht: Reidel
- Recanati, F. (2010). *Truth-conditional pragmatics*. Oxford: Oxford University Press.
- Vendler, Z. (1957). Verbs and times. *The Philosophical Review* LXVI: 143160.