

Anaphora and Local Coherences

Lars Konieczny (lars@cognition.uni-freiburg.de)
Helmut Weldle (helmut@cognition.uni-freiburg.de)
Sascha Wolfer (sascha@cognition.uni-freiburg.de)
Daniel Müller (daniel@cognition.uni-freiburg.de)
Peter Baumann (peter@cognition.uni-freiburg.de)

Center for Cognitive Science, University of Freiburg, Friedrichstr. 50
D-79098 Freiburg i. Br., Germany

Abstract

We present two visual world studies indicating that local syntactic coherences interact with binding constraints (Chomsky, 1981) of both reflexives and pronouns. Gazes to depicted referents or events suggest that when sentences containing a local coherence with a pronoun or reflexive are presented, locally coherent antecedents become activated. Our results strengthen the assumption that local syntactic coherences are interpreted and extend the effect to online anaphora resolution and complementary binding constraints.

Keywords: anaphora, anaphora resolution, local syntactic coherences

Introduction

To arrive at a coherent interpretation of a sentence, we need to bind referring expressions to their correct referent. Binding theory (Chomsky, 1981) provides a syntax-driven structural account for the dependencies of co-reference within sentences. Principles based on c-command are assumed to constrain the possible co-referents of anaphoric expressions on a global level. Reflexives and pronouns have complementary structural binding domains, i.e. within sentence boundaries, the accessible antecedents for both anaphora types are mutually exclusive. In sentences like (1-a) and (1-b) determining the referent of the anaphoric expression *himself* or *him* is straightforward.¹

- (1) a. Ken_i who likes John_j saw himself_{i/*j} in the mirror.
b. Ken_j who likes John_i saw him_{i/*j} in the mirror.

Recent findings question strictly structure-driven accounts of anaphora resolution. Runner, Sussman, and Tanenhaus (2006), for instance, report violations of the binding domain complementarity assumption. They examined preferences for pronoun and reflexive binding in picture noun phrases in a series of *visual-world* studies. Fixation probabilities on depicted referents revealed violations of the binding theory assumption. They concluded that reflexives should rather be explained in terms of logophors, depositing reflexives beyond the scope of Binding theory explanations.

Converging evidence was found by Kaiser, Runner, Sussman, and Tanenhaus (2009), suggesting that the interpretation of reflexives is not only sensitive to structural but also semantic information. Moreover, they found differing degrees of

¹Subscripted indices mark coreference. A star * indicates that coreference with the indexed referent is not acceptable considering the global parse but suggested considering the local parse.

sensitivity towards different sources of information for pronouns and reflexives.

Sturt (2003) on the other hand showed that the constraining principle for reflexives operates at the very earliest stages of processing. In eye-tracking-while-reading experiments, he found early effects of binding preferences. He concludes that the responsible binding principle is an early filter for the processing of referring expressions.

Interpretation of locally syntactic coherences

Local syntactic coherences (LSCs) have been shown to interfere with the global sentence interpretation. Tabor, Galantucci, and Richardson (2004) found increased reading times on the spill-over of *tossed* in sentences like *The coach chided the player tossed a frisbee by the opposing team*. Moreover, Konieczny, Müller, Baumann, Hachmann, and Wolfer (2009) have shown that LSCs temporarily affect the interpretation of globally unambiguous sentences.

Interestingly, there are locally coherent substrings in (1), leading to the opposite binding of the reflexive or the pronoun: In Sentence (1-a), *himself* is restricted to be bound to *John* if only the local subparse *John saw himself* is taken into account. In the global parse however, *himself* is bound to *Ken*. In Sentence (1-b), *him* is bound to *Ken* or any other (unmentioned) referent if the LSC *John saw him* is interpreted.

It is still an open question though, whether or not LSCs can affect anaphora resolution, and if pronouns and reflexives are affected equally.

In the remainder of the paper, two experiments will be reported providing insight into the time-course of binding and its interaction with local syntactic coherences. We chose the *visual-world* paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) over reading times, as fixations on visual objects can indicate binding preferences in a much more direct way, without inferencing over processing difficulty. The results suggest that LSCs indeed have an effect on the binding of reflexives and pronouns in a way that strong constraints like *Principle A* are also applied to the local parse, temporarily overriding the globally correct binding.

Experiment 1: Depicted referents in the visual world

In the first experiment, the visual stimuli showed the depictions of three persons, two of which were depictions of the referents introduced in the spoken sentence. When the

antecedents get reactivated when a pronoun or reflexive is processed, we should see increased gaze proportions on the corresponding referent. Following this logic, the gaze pattern should indicate how local syntactic coherences affect anaphora resolution.

Materials and design

We tracked participant's gaze on depictions of three referents (cf. Figure 1) while they listened to sentences like (2). Materials were constructed according to a 2x2-design crossing the factors anaphor type (reflexive vs. pronoun) and presence of LSC (present vs. prevented), leading to four experimental conditions: reflexive with LSC (Sentence (2-a)), reflexive without LSC (Sentence (2-b)), pronoun with LSC (Sentence (2-c)), and pronoun without LSC (Sentence (2-d)). We prevented the LSC by inserting an adverb before the verb of the relative clause. Each participant was presented with 48 experimental sentences and an equal amount of sentence-picture pairs of comparable complexity. The task was to look at the pictures while listening to the sentences. Immediately after the sentence, participants had to click on the agent they considered to be most important in the scene. 25 participants took part in Experiment 1.

Auditory stimuli

We recorded 48 sentences with normal speech tempo. Locally coherent sequences were recorded separately – as main clauses (*Der Sohn kämmt sich im Wohnzimmer ...*) – and spliced into surrounding sentences (without the starting determiner of the LSC to prevent sentence-initial prosody). By doing so, we wanted to minimize prosodic cues induced by the relative clause boundary and thereby destroying the local coherence. To minimize prosodic differences between conditions, control conditions were produced by splicing the adverbs (*gründlich/thoroughly*) into the first two conditions. This method was necessary since earlier findings indicated strong sensitivity to prosodic cues (Konieczny et al., 2009). These effects are outside the scope of this study. The resulting experimental stimuli still sounded natural, as was established in a pre-test with native speakers who were naïve with respect to the research questions at hand.

(2) a. Reflexive, LSC

Während der Vater_i, den der Sohn_j kämmt,
While the father_i, who the son_j combed,
sich_{i/*j} im Wohnzimmer anzog, ...
himself_{i/*j} in the living room dressed, ...
While the father who the son combed dressed himself in the living room, ...

b. Reflexive, no LSC

Während der Vater_i, den der Sohn gründlich
While the father_i, who the son combed,
kämmt, sich_i im Wohnzimmer anzog, ...
thoroughly himself_i in the living room dressed, ...
While the father who the son combed thoroughly dressed himself in the living room, ...

c. Pronoun, LSC

Während der Vater_i, den der Sohn_j kämmt,
While the father_i, who the son_j combed,
ihn_{j/*i} im Wohnzimmer anzog, ...
him_{i/*j} in the living room dressed, ...
While the father who the son combed dressed him in the living room, ...

d. Pronoun, no LSC

Während der Vater, den der Sohn_j gründlich
While the father, who the son_j thoroughly
kämmt, ihn_j im Wohnzimmer anzieht, ...
combs, him_j in the living room dresses, ...
While the father who the son combed thoroughly dresses him in the living room, ...

Visual stimuli The visual stimuli of Experiment 1 consisted of three referents. The globally suggested agent, the agent contained in the LSC and a third, non-mentioned referent (see Figure 1). The positions of the referents was cross-balanced over all trials and participants.

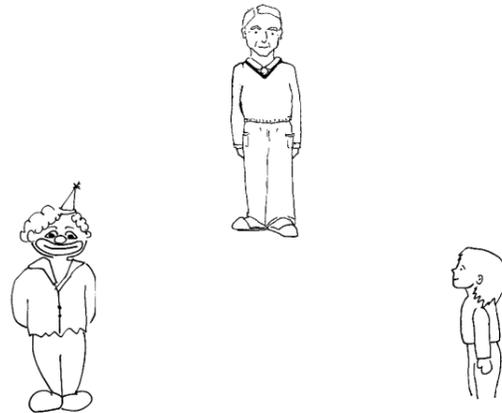


Figure 1: A picture of three referents from Experiment 1. The father (top), the son (right), and a clown (left).

Hypotheses

We expect the LSC to have an influence on fixation probabilities when the anaphor is heard or shortly after. If the LSC has an effect, we expect fixation probabilities on the referent denoted by the LSC to be increased, compared to sentences where no LSC is present. Therefore, when sentences like (2-a) are heard, fixation probabilities on the *son* should increase compared to sentences like (2-b) because the son is the only possible referent of the local interpretation of *sich/himself*. Accordingly when sentences like (2-c) are heard, we expect fixation probabilities on the *father* to increase compared to sentences like (2-d) because the father is one possible referent of the local interpretation of *ihn/him*. We expect the effect for the reflexive condition to be stronger than for the pronoun, because the reflexive

sich/himself corefers in its local interpretation with the agent of the LSC (the *son*) whereas the pronoun *ihn/him* corefers in its local interpretation with the first-mentioned agent (the *father*) but also with every other referent except the *son*.

Results

For reflexives (Figure 2), there are significantly more fixations on the referent denoted by the LSC, when the LSC was present than when it was absent. Fixation probabilities differ significantly in the range from 500 ms to about 1600 ms after the onset of the reflexive in the spoken stimulus. The referent denoted by the global meaning of the sentence is fixated the most after about 1200 ms after the synchronization point, indicating that the participants understood the spoken stimuli. The non-overlapping standard errors indicate significant differences of mean fixation probabilities in the different conditions. This was further validated by fitting a linear mixed effects model using the statistical software R (R Development Core Team, 2010) with the package lme4 (Bates, 2007). Analyses of fixation patterns for the relevant sections revealed significant differences, as tested with MCMC-sampling (all $ps < .05$). However, for the pronoun condition, we found no reliable difference in fixations on the referent denoted by the LSC when a LSC was present compared to when it was absent (Figure 3).

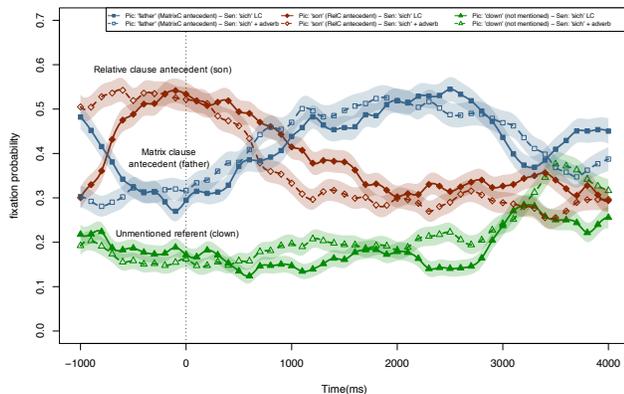


Figure 2: Results for *reflexives*: Proportion of fixations on the *globally correct referent* (blue lines), the *local referent* (red), and on an *unmentioned person* (green). Zero marks the onset of the reflexive in the spoken stimulus. There are significantly more fixations to the local referent (red lines), when a LSC was present (solid) than in the control condition (dashed), in the range of about 500 ms to 1600 ms.

Discussion

The results clearly indicate that binding can be disturbed by a local syntactic coherence, at least for reflexives. The lack of a local coherence effect for pronouns in this experiment might have several reasons: Due to their complementary binding domains, pronouns would have their antecedent outside the

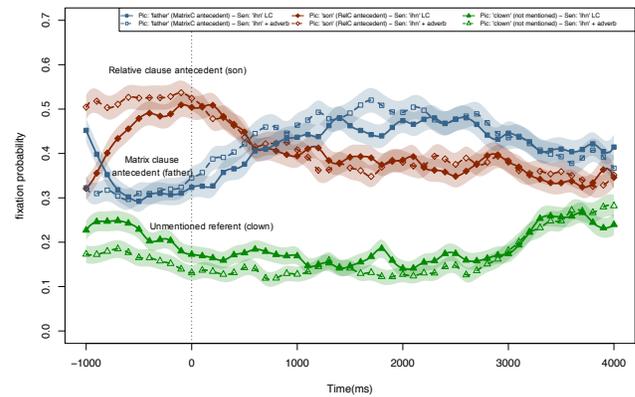


Figure 3: Results for *pronouns*: Proportion of fixations on the *globally correct referent* (blue lines), the *local referent* (red), and on an *unmentioned person* (green). Zero marks the onset of the pronoun in the spoken stimulus. There was no reliable difference in fixations to the local referent (red lines), when a LSC was present (solid) than in the control condition (dashed).

actual local coherence. Another possibility is that participants might have looked at both of the two actors (the *agent* and the *patient*) in transitive events, such that the gazes are not as informative for the pronoun cases as for the reflexive cases, where there is only one actor (the *agent*) in each event. These, and other potential problems of Experiment 1 were dealt with in Experiment 2.

Experiment 2: Depicted events in the visual world

In Experiment 2, we used depicted events instead of referents. By doing so, we were able to provide a unique depiction even for the transitive (i.e. pronoun) cases. We also used a different control condition (i.e. locally non-coherent condition), which is more effective than inserting an adverb before the verb, as in Experiment 1. We used particle verbs, such as *an-ziehen*, in the non-local conditions. In main clauses, the particle *an* would have to be separated and placed at the end of the clause, as in: *der Sohn zieht sich an*.

Materials and design

Again, we used a 2x2-design crossing the type of anaphora (reflexive vs. pronoun) and the presence of the LSC (present vs. absent). Note, that the type of scene (depiction of transitive or reflexive actions) was always presented with the corresponding type of anaphor. So, reflexive scenes like those in Figure 4 were always presented with sentences like (3-a) and (3-b), transitive scenes (Figure 5) were presented with sentences like (3-c) and (3-d). The task was to click on (one of the) correct scenes after hearing the sentence. 36 participants took part in Experiment 2.

Auditory stimuli We recorded 24 sentences and used the double amount of sentence-picture pairs as fillers. Again, the local coherent substring was spliced into the surrounding sentence. To create the control conditions, we swapped the verb of the relative clause (*kämmte/combed*) with the second verb of the sentences (3-a) and (3-c) (*anzog/dressed*), which was always a particle verb. When a particle verb is placed inside the relative clause, the LSC is no longer valid (**Der Sohn anzog sich/ihn im Wohnzimmer ...*).

(3) a. *Reflexive, LSC*

Während der Vater_i, den der Sohn_j kämmte,
 While the father_i, who the son_j combed,
 sich_{i/*j} im Wohnzimmer anzog, ...
 himself_{i/*j} in the living room dressed, ...
While the father who the son combed dressed himself in the living room, ...

b. *Reflexive, no LSC*

Während der Vater_i, den der Sohn anzog, sich_i
 While the father_i, who the son dressed, himself;
 im Wohnzimmer kämmte, ...
 in the living room combed, ...
While the father who the son dressed combed himself in the living room, ...

c. *Pronoun, LSC*

Während der Vater_i, den der Sohn_j kämmte,
 While the father_i, who the son_j combed,
 ihn_{j/*i} im Wohnzimmer anzog, ...
 him_{i/*j} in the living room dressed, ...
While the father who the son combed dressed him in the living room, ...

d. *Pronoun, no LSC*

Während der Vater, den der Sohn_j anzog, ihn_j
 While the father, who the son_j dressed, him_j
 im Wohnzimmer kämmte, ...
 in the living room combed, ...
While the father who the son dressed combed him in the living room, ...

Visual stimuli We used scene depictions instead of referents. Sentences with reflexives were always presented with scenes depicting reflexive actions (Figure 4), whereas sentences with pronouns were always presented with scenes depicting transitive actions (Figure 5). This procedure led to a total of 192 scenes. The positions of the scenes were cross-balanced over all trials and participants.

In the reflexive conditions (Sentences (3-a) and (3-b)) the global interpretation of Sentence (3-a) refers to the scene where *the father is dressing himself* (the lower left scene in Figure 4). The LSC is referring to *the son combing himself* (the upper left scene in Figure 4).

Because we generated the control conditions by swapping the verbs, the depicted target and control events were different. That is, for locally coherent sentences the target scene was *the son combing himself* (the upper left scene in Figure

4), whereas for the non-coherent controls, the corresponding scene was *the son dressing himself* (the upper right scene in Figure 4). Accordingly, the globally correct scene for Sentence (3-b) changes from *the father dressing himself* (the lower left scene in Figure 4) to *the father combing himself* (the lower right scene in Figure 4).

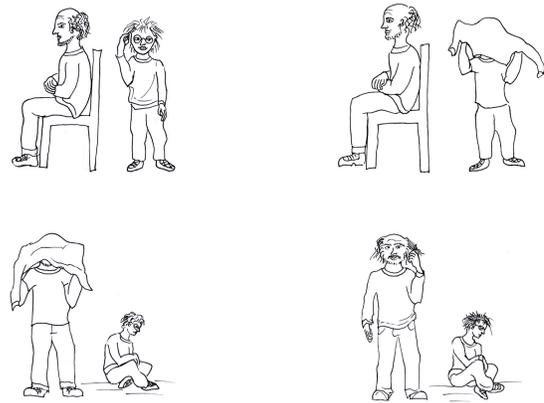


Figure 4: A picture of four scenes depicting reflexive actions from Experiment 2. The son combing himself (upper left), the son dressing himself (upper right), the father dressing himself (lower left), and the father combing himself (lower right).

In the pronoun condition with an LSC (Sentence (3-c)) the global sentence meaning refers to the scene where *the son is combing the father* (the upper left scene in Figure 5) as well as to the scene where *the father is dressing the son* (the lower right scene in Figure 5). The meaning of the LSC is also referring to the scene where *the son is combing the father*. Importantly, in the pronoun condition with an LSC, both global and local meanings are referring to the same scene.

Again, the control condition was generated by swapping the verbs, therefore the depicted target control events were also different. For locally coherent sentences the target scene was the *the son combing the father* (the upper left scene in Figure 5), whereas for the non-coherent controls, the corresponding scene was *the son dressing the father* (the upper right scene in Figure 5). Accordingly, the globally correct scene for Sentence (3-d) changes from *the father dressing the son* (the lower right scene in Figure 5) to *the father combing the son* (lower left scene in Figure 5).

Hypotheses

Reflexive condition We again expect a contrast in fixation probabilities between sentences with LSC (Sentence (3-a)) and sentences without LSC (Sentence (3-b)), such that the picture depicting the event expressed by the LSC (*Der Sohn kämmt sich im Wohnzimmer*) is fixated more often when the LSC is present than when it is not (**Der Sohn anzog sich im Wohnzimmer*). This effect should show up shortly after the offset of the reflexive. The other event, which is described by the main clause (*the father dressing (LSC) or combing (no*

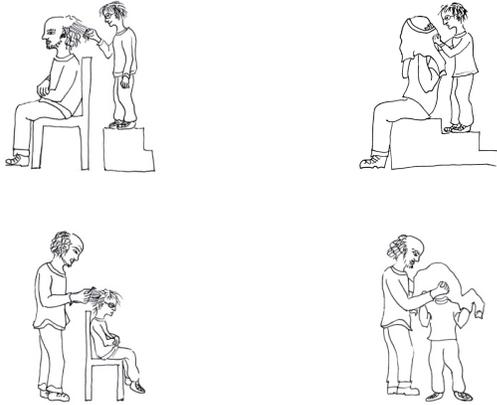


Figure 5: A picture of four scenes depicting transitive actions from Experiment 2. The son combing the father (upper left), the son dressing the father (upper right), the father combing the son (lower left) and the father dressing the son (lower right).

LSC) *himself*) should be fixated later as soon as the event is mentioned.

Pronoun condition In the pronoun condition both the global and the local meaning of the sentence refer to the same scene (*the son combing the father*), therefore we expect a "boost effect" on fixation probabilities on this scene, i.e. we expect fixation probabilities to be high in the control condition (Sentence (3-d)), but even higher in the local coherent condition (Sentence (3-c)). Furthermore, fixation probabilities on the event described in the main clause should rise as soon as the corresponding event is mentioned, i.e. in the LSC condition, the fixations on the main verb action depiction should be delayed until after the offset of the local coherence.

Results

For reflexives (Figure 6)², there are significantly more fixations to the scene denoted by the LSC when the LSC was present than when the LSC was absent. This effect ranges from about 800 ms to after 2000 ms from the onset of the reflexive. Of course, during the matrix clause at the end of the sentence, the highest proportion of fixations is on the scene denoted by the matrix clause (after about 2200 ms from the onset of the reflexive).

For pronouns (Figure 7), the meaning of the LSC coincides with the globally correct meaning of the relative clause, which are therefore depicted by the same scene. The significant difference between the LSC and the non-LSC condition demonstrates the expected "boost effect". This effect lasts from 200 ms to about 1000 ms after the onset of the pronoun. When the event denoted by the matrix clause is described, there are the most fixations on the corresponding scene (after

²For expository reasons only fixation probabilities on the locally and globally denoted scenes were plotted.

about 2200 ms). Again, the indication of significant differences by non-overlapping standard errors were further validated by fitting a linear mixed effects model (all $ps < .05$).

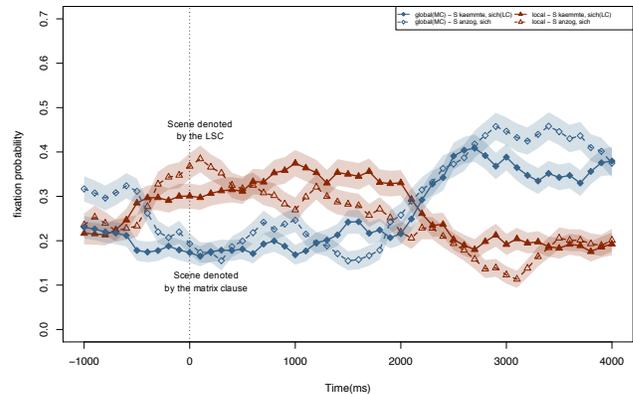


Figure 6: Results for *reflexives*: Proportion of fixations on the depicted event with the *globally correct binding* (blue lines), the *local binding* (red). Fixations on the other depictions are not plotted. Zero marks the onset of the reflexive in the spoken stimulus. There are significantly more fixations to the local binding depiction (red lines), when a LSC was present (solid) than in the control condition (dashed), in the range of about 800 ms to after 2000 ms.

Discussion

Experiment 2 replicates the results for reflexives in Experiment 1. Ignoring for a minute the period where the effect seems to be reversed, about 800 ms after the reflexive we see the local coherence effect, i.e. more looks to the locally coherent, or the control scene, respectively, when there is a local coherence in the speech input, than when there is none. Different from experiment 1, we also found a local coherence effect with pronouns. As discussed above, this might be due to the fact that the target picture in Experiment 2 is a single scene including the directionality of the action, whereas there are two depicted target actors involved in transitive actions in Experiment 1. Most notably, interpreting the pronoun within the local coherence overlaps with the meaning of the relative clause itself, so that the effect amounts to boosting the correct interpretation. The short inverse effect for reflexives starting even before the reflexive and lasting to about 200 ms after the onset of the reflexive might seem worrying at first glance. Note however that this effect is likely due to the fact that different target scenes were used for the target and the control condition (due to swapping the verbs). This difference could hence be attributed to differences in visual saliency between the two depictions. Moreover, the swapped verbs themselves might have added to the effect: particle verbs clearly morphologically indicate a clause boundary, whereas the verbs used for local coherences do not. Detection of a clause boundary might have triggered a short-lived attention-shift towards pic-

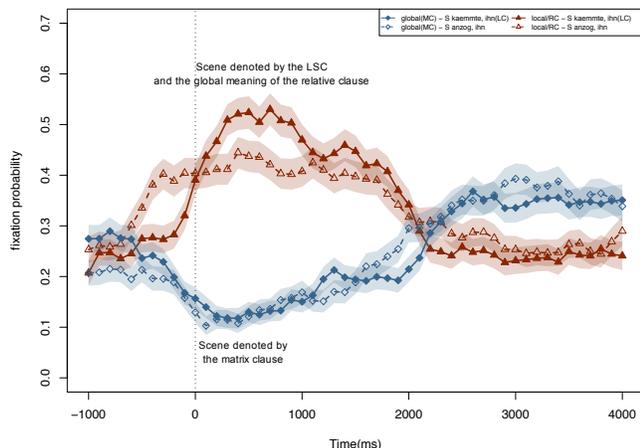


Figure 7: Results for *pronouns*: Proportion of fixations on the *globally correct scene described by the matrix clause* (blue lines) and the *scene described by the LSC and the global meaning of the relative clause* (red). Fixations on the other depictions are not plotted. Zero marks the onset of the pronoun in the spoken stimulus. There are significantly more fixations to the local binding depiction (red lines), when a LSC was present (solid) than in the control condition (dashed), in the range of about 200 ms to after 1000 ms.

tures depicting the verb's action. Note that the same holds for the pronoun cases, where we also see an early and short-lived advantage for the "local" picture in the non-local condition. The early appearance of the effect however, as in the reflexives, renders it unlikely that the effect is due to the local coherence itself.

Conclusions

Both experiments have shown that local syntactic coherences can influence the binding of pronouns and reflexives. The results suggest that LSCs open a short-lived window, during which binding constraints can work both locally and globally. With respect to the time course, shortly after an anaphoric expression is heard, potential referents or scenes corresponding to the binding of the anaphor are fixated. This argues for very early constraints (from both global and local interpretations) exerting their influence on co-reference assignment. This result is in line with Sturt (2003) who showed that *Principle A* operates at the very earliest stages of processing. However, we could also show that globally correct binding can be delayed when a local syntactic coherence interferes.

Furthermore, our findings suggest that the effects for pronouns are more fragile than those for reflexives, replicating similar findings by Kaiser et al. (2009) and Runner et al. (2006) who found that binding constraints for reflexives are harder than those for pronouns. Different to their findings, our results are not dependent on the specific semantic nature of the stimuli, as is the case with picture noun phrases.

On a larger scale, our results can be interpreted in two

ways. They could be seen as an indicator that binding principles for pronouns and – especially – for reflexives are too restrictive because they only capture the by nature global structural characteristics of sentences. On the other hand our results speak for the validity of Binding Theory, because it is even applicable in non-global structures like the local syntactic coherences presented here.

What seems quite clear considering our results, is that LSCs are processed and interpreted in such a profound way that they exert a clear influence, even on the binding of anaphoric expressions.

Acknowledgments

We would like to thank Anne Karina Feldmeth, David Kühner and Minkus Teske for their detail work on the auditory and visual stimuli. We also want to thank Sarah Schwarzkopf for her voice and for recording the auditory stimuli. This work was supported by the German Research Foundation (DFG, grant KO 1932/3-1).

References

- Bates, D. M. (2007). Linear mixed model implementation in lme4. *Manuscript, University of Wisconsin - Madison*.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real time investigation of speech perception, memory and language processing. *Cognitive Psychology*, 6, 84-107.
- Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, 112(1), 55-80.
- Konieczny, L., Müller, D., Baumann, P., Hachmann, W., & Wolfer, S. A. (2009). Local syntactic coherence interpretation, and how prosody modulates it. In *Proceedings of the 22nd annual meeting of the cuny conference on human sentence processing*. Davis, CA.
- R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.
- Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Assigning referents to reflexives and pronouns in picture noun phrases: Experimental tests of binding theory. *Cognitive Science*, 30, 1-49.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355-370.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.