Dario Paape | Shravan Vasishth

Local Coherence and Preemptive Digging-in Effects in German

Local Coherence and Preemptive Digging-in Effects in German

Dario Paape
University of Potsdam, Germany

Shravan Vasishth
University of Potsdam, Germany; University of Sheffield, UK

Abstract
SOPARSE predicts so-called local coherence effects: locally plausible but globally impossible parses of substrings can exert a distracting influence during sentence processing. Additionally, it predicts digging-in effects: the longer the parser stays committed to a particular analysis, the harder it becomes to inhibit that analysis. We investigated the interaction of these two predictions using German sentences. Results from a self-paced reading study show that the processing difficulty caused by a local coherence can be reduced by first allowing the globally correct parse to become entrenched, which supports SOPARSE's assumptions.

Keywords
Local coherence, digging-in effects, self-paced reading, SOPARSE, sentence processing, German

Introduction

According to standard models of parsing, incoming words are immediately integrated with preceding material according to the rules of grammar (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995); in cases where these rules do not disambiguate between multiple attachment possibilities, integration can nevertheless proceed if a heuristic such as Minimal Attachment or Late Closure is used, as long as no grammatical principles are violated (Frazier, 1979).

One such principle has been suggested by Chomsky (1981) in the form of the Theta Criterion, which requires: (a) that a verb discharge all of its associated theta roles; and (b) that each available argument receives exactly one theta role. Consider the following sentence from Frazier and Rayner (1982):

(1) Though George kept on reading that stupid science fiction story really bothered him.

Corresponding author:
Dario Paape, Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany.
Email: paape@uni-potsdam.de

This sentence creates a garden path effect due to the parser's tendency to integrate the noun phrase (NP) *that stupid science fiction story* into the initial adjunct clause, where it receives a theta role from *reading*. This behavior is predicted by Frazier's (1979) Late Closure principle, which mandates that incoming material be attached to the phrase that is currently being processed rather than open a new phrase. When the verb *bothered* arrives, it will fail to discharge both of its theta roles. This grammatical violation, in combination with the reader's knowledge that an English main clause cannot start with an adverb–verb sequence, signals that part of the existing structure needs to be altered, explaining why (1) causes processing difficulty. Retracting the attachment of *that stupid science fiction story* to *reading* and thus freeing the NP for theta assignment is possible in (1) due to the fact that *read* can also be intransitive.

Consider now the sentences in (2a, b), taken from Tabor, Galantucci, and Richardson (2004).

> (2)　a.　The coach smiled at the player (who was) tossed a frisbee by the opposing team.
> 　　　b.　The coach smiled at the player (who was) thrown a frisbee by the opposing team.

In a self-paced reading study, Tabor et al. (2004) found that when the relative clause appeared in reduced form, that is, when the words *who* and *was* were missing, the region between the passive participle (*tossed*/*thrown*) and the word *by* was read more slowly in sentence (2a) than in (2b). When the relative clause was unreduced, no such difference appeared. The authors explain the effect through the presence of local coherence in the reduced version of (2a): due to the ambiguous nature of the participle *tossed*, the string *the player tossed a frisbee* could, in principle, constitute a well-formed subject–verb–object (SVO) sequence with an active reading while the string *the player thrown a frisbee* could not.

However, in order to allow the consideration of *the player* as the agent of *tossed*, the criterion demanding that an NP must not have two theta roles has to be weakened. Additionally, under standard assumptions about phrase structure, there is no way of deriving a tree in which *the player* is embedded under *smiled at* and dominates *tossed* in an active construction at the same time. Neither is it possible to reanalyze the sequence as in (1), as this would leave the obligatorily transitive *smiled at* without an object. Finally, *tossed* can be integrated into the existing structure without attempting to detach *the player*, namely by being analyzed as belonging to a reduced relative construction.[1] If the parser realizes that this option is available, reanalysis should not be triggered.

The problem lies in the assumption that the parser needs to be, as Tabor et al. (2004) call it, self-consistent. There are, however, processing models which relax this assumption, among which Tabor et al. (2004) specifically mention the Sausage Machine (Frazier & Fodor, 1978), multi-layered probabilistic accounts (Crocker & Brants, 2000) and, most importantly for present purposes, self-organized parsing accounts such as the Unification Space model (Vosse & Kempen, 2000) and SOPARSE (Tabor & Hutchins, 2004).[2]

SOPARSE belongs to the class of constraint-based competition models and, as is typical of such models, assumes that lexical items carry syntactic structure. Once a lexical item enters the computation, it will attempt to form attachments to other items wherever this is locally possible. Attachments (or 'links') are assumed to have continuous strength values and compete with other, incompatible attachments, inhibiting them until a single set of links remains. A consistent parse will consist of links which do not (or only weakly) compete with each other, but rather 'conspire' together to prevail. SOPARSE explains local coherence effects by assuming that a linkage with high local consistency enters into competition with the globally licensed one, increasing the time needed to consolidate the correct parse.

SOPARSE predicts that semantic as well as syntactic local coherence influences the amount of competition. Consider (3a, b).

(3) a. The bandit worried about the prisoner (that was) transported the whole way by the capricious guards.
   b. The bandit worried about the gold (that was) transported the whole way by the capricious guards.

The prisoner is a plausible agent of *transported* while *the gold* is not. In another self-paced reading study, Tabor et al. (2004) found that when the sentences appeared in reduced form, reading times for (3a) increased compared to (3b) starting at *transported*, thus lending support to the competition account. Using a visual-world paradigm and German stimuli, Konieczny, Müller, Hachmann, Schwarzkopf, & Wolfer (2009) further substantiated the notion that locally coherent linkages are semantically interpreted.

Apart from local coherence effects, SOPARSE also predicts digging-in effects: a linkage – normally called an attachment – grows in strength both due to the passage of time as well as due to additional consistent linkages being formed. A stronger linkage will inhibit its competitors more strongly. Furthermore, SOPARSE employs a 'rich get richer' principle: the more activation a linkage has already accumulated, the faster its activation will continue to grow. To demonstrate digging-in effects, Tabor and Hutchins (2004) investigated sentences such as (4a, b).

(4) a. As the author wrote (the essay) the book grew.
   b. As the author wrote (the essay) the book describing Babylon grew.

The addition of the modifier after the NP *the book* resulted in longer reading times at the verb *grew* when the sentences appeared in their ambiguous versions. This effect can be explained by assuming that the erroneous linkage between *wrote* and *the book* grows stronger while the relative clause is being processed, thus being able to compete more strongly with the correct linkage between *the book* and *grew*.

Since digging-in should apply to ultimately correct analyses as well, a novel prediction of the SOPARSE architecture can be derived: allowing the globally consistent parse to gather strength before introducing a local coherence should ameliorate the competition effect, as it will win the competition against the 'intruding' parse more easily.

As explained by Tabor and Hutchins (2004, p. 445ff.), SOPARSE views a lexical item as successfully integrated when either (a) the activation of an associated set of links has reached a threshold value, or (b) some maximum amount of time has passed. The presence of competing links causes the activation of a given set of links to grow more slowly. Thus, the globally correct parse and the locally coherent parse will inhibit each other. If the globally consistent parse has had time to become entrenched, it will have an advantage in this competition, alleviating the disrupting effect of the local coherence. We will refer to this idea as *preemptive digging-in*.

German is a particularly interesting language for investigating preemptive digging-in. One peculiarity of German is that it features SVO main clauses and subject–object–verb subordinate clauses. Compare (5a, b).

(5) a. Der General   belagerte die Stadt.
      the general   besieged   the city
   b. Viele starben, als   der General die Stadt   belagerte.
      many   died   when   the general the city   besieged
      'Many died when the general besieged the city.'

As there are no exceptions to this rule, as soon as a subordinating conjunction is encountered a native speaker of German should have a strong expectation of the verb appearing at the end of the clause, after all its (non-sentential) arguments.

There is evidence that readers of verb-final sentences attach arguments before they encounter the verb (Bader & Lasser, 1994; Kamide & Mitchell, 1999; Konieczny, Hemforth, Scheepers, & Strube, 1997). In order to account for such pre-head attachment, models which assume that syntax comes from the lexicon need to be augmented. What is needed is a mechanism that can introduce phrases ahead of time if their associated lexical head is certain to appear downstream. For this purpose, Kamide and Mitchell (1999) suggest introducing a *generic sentence-structure frame*, which is ''something like a standard verb-frame but lacking the lowest level information which instantiates the frame as the product of a particular verb'' (p. 635). A similar approach to verb-final structures is suggested by Konieczny (1996), who assumes a *prediction stack* that allows the attachment of incoming material to predicted syntactic nodes.

Once the generic verb-frame has been introduced into the computation space, it should be able to form linkages, which should in turn be able to accumulate strength. We implemented this idea by adding adverbial material, which is assumed to be attached to the generic verb-frame, before introducing a local coherence (see examples in (6)).

2 Experiment

2.1 Design and Materials

We conducted a self-paced reading study with a 2 × 2 design with the factors local coherence (locally coherent vs. not locally coherent) and digging-in time (short vs. long). Following the schema exemplified in (6), we created 32 sentences containing ambiguous participle forms, along with 96 filler sentences containing a variety of structures. A complete list of the experimental items is given in the Appendix.

(6) Heute   weiß   man, dass …
    today   knows   one   that …
    a. **short, locally coherent**
       einer der   Generäle belagerte Städte mit   großer Verbissenheit verteidigte,
       one of the generals besieged cities with great   determination defended. Singular (SG)
       um   Zeit   zu gewinnen.
       in order to time   to gain
    b. **long, locally coherent**
       entlang der Nachschubrouten   anfänglich einer der   Generäle belagerte Städte mit
       along   the supply lines   initially   one of the generals besieged cities with
       großer   Verbissenheit   verteidigte,   um   Zeit   zu   gewinnen.
       great   determination defended.SG   in order to   time   to   gain
    c. **short, not locally coherent**
       einige der   Generäle belagerte Städte mit großer Verbissenheit verteidigten,
       some   of the generals besieged cities with great   determination defended. Plural(PL)
       um   Zeit   zu   gewinnen.
       in order to time   to   gain
    d. **long, not locally coherent**
       entlang der Nachschubrouten anfänglich   einige der   Generäle belagerte Städte mit
       along   the supply lines   initially   some of the generals   besieged cities with

großer	Verbissenheit	verteidigten,	um	Zeit	zu	gewinnen.
great	determination	defended.PL	in order to	time	to	gain

'Today we know that one/some of the generals defended besieged cities with great determination in order to buy time.'

The participle form *belagerte*, 'besieged' is ambiguous between a plural-inflected adjective and a verb marked for past, singular and third or first person. We will refer to the participle as the critical word as it introduces the local coherence.

In the globally correct parse of these stimuli, *belagerte*, 'besieged' acts as a plural-inflected adjective modifying a determinerless object NP (*belagerte Städte*, 'besieged cities'). Local coherence was manipulated via the number feature of the embedded subject, which either did or did not fit with an intruding SVO sequence. In this local linkage, the participle acts as a verb and *Städte*, 'cities' acts as a direct object.[3] (7) illustrates how the availability of the locally coherent parse is influenced by the number feature of the subject by presenting it as an isolated sentence.

(7)   **Locally coherent**

　　a. | Einer | der | Generäle | belagerte | Städte. |
|-------|-----|----------|-----------|---------|
| one | of the | generals | besieged. Singular (SG) | cities. |

'One of the generals besieged cities.'

Not locally coherent

　　b. | *Einige | der | Generäle | belagerte | Städte. |
|---------|-----|----------|-----------|---------|
| some | of the | generals | besieged. SG | cities. |

*'Some of the generals besieged cities.'

The manipulation of *eine/einige*, 'one/some' marks the subject *einer/einige der Generäle*, 'one/some of the generals' as either singular or plural, respectively. In the case of *einige*, 'some', this results in an agreement mismatch between the subject and the would-be verb of the locally coherent parse in (6c, d), which is marked for singular, as plural would require the suffix *-n*. Assuming that local coherence is based on the feature match between the verb and its arguments, the effect should be shut down or at least significantly weakened by this change, which is why the plural version is referred to as not locally coherent.[4]

Digging-in time for the globally correct parse was manipulated by adding one or more adjuncts right after the complementizer, as in (6b, d). Assuming that the presence of a complementizer will allow attachments to the upcoming verbal structure, as explained above, attaching these adjuncts to the generic verb-frame should grant it an advantage in the competition against the intruding SVO sequence.

The adjunct phrase which appears after the object NP (for instance, *mit großer Verbissenheit*, 'with great determination') never disambiguates towards either the locally coherent or the global linkage due to the fact that sentence-final adverbial material is possible in German main clauses. The semantic content of these phrases was designed to fit with both interpretations as well: cities can be besieged or defended with great determination. The local coherence ends at the verb of the global parse, that is, *verteidigte(n)*, 'defended' in (6).

As some accounts suggest that lexical biases may cause or strengthen local coherence effects (Crocker & Brants, 2000; Gibson, 2006), a numeric measure of the lexical bias of the ambiguous participle was calculated. This measure was obtained by querying the newspaper corpus of the Digitales Wörterbuch der Deutschen Sprache (Geyken, 2006; Geyken, Klein & Wolz, 2004; queried online at http://www.dwds.de/) for occurrences of the ambiguous word's lemma tagged as either a finite verb or as adjective. The weighted average of counts in four regional and national

newspapers (894 million tokens) was calculated for each tag and the verb count was divided by the adjective count. The resulting bias ratio was logarithmized, such that verb-biased participles received a positive value and adjective-based participles received a negative value.

2.2 Predictions

SOPARSE predicts that a locally coherent SVO (Adjunct) sequence should compete with the globally correct SO (Adjunct)V structure from the critical ambiguous participle onwards. This should result in increased reading times in comparison to the not locally coherent conditions.

Furthermore, once we assume that Kamide and Mitchell's (1999) generic sentence-structure frame is part of the model, attaching phrases to this verb-final structure should create a digging-in effect: the generic verb-frame should be a stronger competitor for the intruding SVO parse, thus weakening the local coherence effect. Note that it is not possible to integrate the additional preverbal adjuncts into the SVO parse, as in German main clauses the finite verb can only be preceded by a single phrase (Speyer, 2004), which in the case of (6, 7) would be the subject.

2.3 Participants

Forty students from the University of Potsdam took part in the experiment. All were native speakers of German and were either paid 5 € or received course credit for their participation.

2.4 Procedure

The experiment was run on a PC using the Linger software developed by Douglas Rohde,[5] using the setting for centered self-paced reading. We opted for the centered presentation method instead of a 'moving window' approach as the experimental sentences were rather long and the introduction of line breaks may have unduly favored one of the two assumed readings. Participants were instructed to read silently and at their normal pace. They were also informed that they would be asked comprehension questions which they should answer based on the information given in the sentence without making additional assumptions. The program was configured to present the experimental sentences in a counterbalanced manner, such that each participant read eight sentences for each condition, but each sentence in only one condition, combined pseudo-randomly with all 96 filler sentences. Each trial was started by a one-second display of a fixation cross in the middle of the screen which could not be skipped. Participants then used the space bar to move through the sentence word by word. Yes–no comprehension questions appeared randomly after one-third of all sentences. Which sentences were followed by a question was determined pseudo-randomly at runtime. Multiple questions were available for each experimental item, with a maximum of one being presented per trial. For the experimental items, questions targeted the global reading and/or the plurality of the subject NP, but never the locally coherent reading. Half of the comprehension questions required 'yes' answers; feedback was given after incorrect answers. At three points during the experiment, participants were offered the opportunity to take a break. Experimental sessions lasted 30 minutes on average.

2.5 Data Analysis and Results

Data analysis was carried out using the R environment (R Core Team, 2012). Participants' accuracy on comprehension questions for experimental items was 90%; the overall accuracy was 94%.
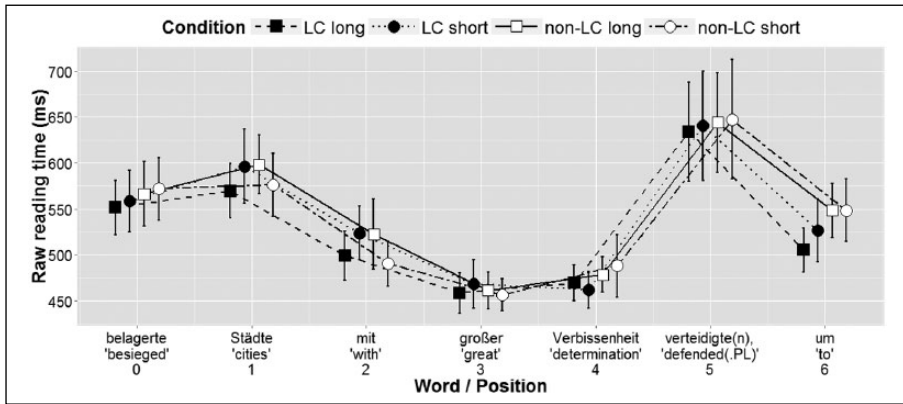
Figure 1. Mean reading times for word positions 0 through 6. Error bars represent 95% confidence intervals.

For the reading time analyses, word positions were labeled relative to the position of the critical ambiguous participle, which was labeled as 0. Figure 1 shows the raw reading times for positions 0 through 6.

Linear mixed-effects models were fit at each region of interest using the lme4 package (Bates, Maechler, & Bolker, 2012) and Stan (Stan Development Team, 2014); mixed-effects modeling offers several advantages over the classical analysis of variance approach, including elimination of the need to analyze subject and item means separately (Baayen, 2008; Baayen, Davidson, & Bates, 2008). The lme4 models included varying intercepts for subjects and items, and varying slopes (where possible – in some cases, the estimates were either zero or a convergence failure occurred). The Stan model included full variance–covariance matrices for both subject and items, using non-informative LKJ priors (see Sorensen & Vasishth, 2014 and Stan Development Team, 2014 for details). The code and data will be released with the publication of this paper.

Sum contrasts were defined for the fixed effects local coherence and digging-in time. The locally coherent (that is, agreement match) conditions were coded as 1 and the not locally coherent (that is, agreement mismatch) conditions were coded as –1. Likewise, the short conditions (that is, those without adverbial material inserted) were coded as 1 and the long conditions as –1, respectively. The corpus-based lexical bias factor was also included as a fixed effect after first being centered. No varying slope was added for the factor bias because it was a between-items factor.

In order to carry out statistical inference with mixed-effects models (and linear models in general), model assumptions need to be satisfied. An important assumption is that the residuals have an approximately normal distribution. Since the residuals of the models fit to raw reading times were noticeably non-normal, the Box–Cox procedure (Box & Cox, 1964; Venables & Ripley, 2002) was applied to find the appropriate transformation such that the residuals would be approximately normal. The procedure suggested reciprocal reading times as an appropriate variance-stabilizing transform; accordingly, reading times were converted to 1/reading times, and then multiplied by 1000 to obtain a rate measure on 1/seconds scale. The sign was also changed to negative in order to allow an intuitive interpretation of the estimated effects. However, there were still some influential values in the form of very short reading times (< 167 ms), so an additional criterion of reciprocal reading time greater than –6 was applied, which resulted in a loss of less than 1% of data at each word position. *P*-values were computed using the likelihood ratio test, by

comparing a model that included the term in question with an otherwise identical model in which the term was not present.

One significant interaction appeared at position 2, the first word of the adverbial phrase following the locally coherent SVO sequence. In the locally coherent conditions, reading times on this word were elevated when digging-in time was short ($X_1^2 = 4.5$, $p = 0.03$). We believe this to be a spillover effect from the critical intruder region. There were no significant main effects and no other interactions at this position.

At position 4, the final word of the post-critical adverbial, a main effect of local coherence was found such that reading times were significantly shorter in the locally coherent conditions ($X_1^2 = 4.4$, $p = 0.04$). This pattern re-emerged at position 6, the word following the clause-final verb ($X_1^2 = 10.6$, $p = 0.001$). For this position, in order to meet the potential objection that there might be differential amounts of processing spillover from position 5 due to the presence of a plural suffix in the not locally coherent (not-LC) conditions, in a separate model the reciprocal reading time of the preceding word was entered into the model as a covariate; this model did not change the result. Table 1 shows the estimated parameters for the positions at which significant effects were found.

2.6 Discussion

The results validate SOPARSE's prediction that digging-in time for the globally correct parse should ameliorate local coherence effects. Adding material after the complementizer led to decreased reading times in the post-critical region when a local coherence was present, most likely due to processing spillover from the critical region. This effect was reversed in the conditions without local coherence. Our speculation is that this reversal is due to increased semantic processing effort when the reader has to integrate multiple unrelated events (for example, a group of generals besieging different sets of cities) with the additional adverbial material.

While the lexical bias of the critical participle did not significantly influence reading times at any position, there was a trend towards faster reading times at position 2 when the adverbial material was present, but only for verb-biased participles. This supports Gibson's (2006) proposal that lexical frequencies and syntactic context interact in word category disambiguation, but only under the assumption that the presence of optional modifiers is factored in.[6] Since SOPARSE is a highly lexicalized model, it is to be expected that lexical frequencies should influence parsing, even though Tabor and Hutchins (2004) make no explicit mention of this. However, an anonymous reviewer informs us that biasing the dynamics of the relevant lexical entries would predict the observed behavior.

The speedup in the locally coherent conditions that was observed at word positions 4 and 6 is surprising. It is especially mysterious why such an effect should appear right after the local coherence clashes with the appearance of the clause-final verb, even when spillover is taken into consideration. One possibility is that on at least some occasions participants failed to parse sentences that contained a local coherence even when the additional material was present. In SOPARSE, the activation of syntactic fragments is subject to a certain amount of randomness or noise, which can cause an illicit analysis to outperform the correct one, resulting in parsing failure.[7] That the effect occurred in both the short and the long conditions would imply that the locally coherent linkage managed to outcompete the globally consistent parse in the long run even if the latter had be given time to consolidate.

To explore this hypothesis, a linear mixed-effects model with subjects and items as varying intercepts was fit to question responses (for experimental items only) using the Laplace approximation. Coherence and digging-in time were entered as fixed effects, using the same coding as for the reading time analysis, along with the lexical bias predictor and all two-way interactions. The fit revealed a significant effect of local coherence on response accuracy ($z = -3.69$,

Table 1. Coefficient estimates, standard errors and *t* values for the linear mixed-effects models, and lower/upper bounds on 95% credible intervals for the Stan models fit to reciprocal reading times at word positions 2, 4 and 6. The 95% credible interval marks the region within which the true parameter value lies with probability 0.95 (given the data).

Position 2

	Estimate	Standard error	*t* value	CrI lower	CrI upper
(Intercept)	−2.25	0.09	−26.59	−2.43	−2.10
coh	0.02	0.01	1.28	−0.01	0.05
len	−0.00	0.01	−0.05	−0.03	0.03
bias	−0.01	0.02	−0.36	−0.06	0.04
coh:len	0.04	0.02	2.17	0.00	0.07
coh:bias	0.01	0.01	0.59	−0.02	0.04
len:bias	0.02	0.01	2.07	−0.00	0.04
coh:len:bias	0.00	0.01	0.13	−0.02	0.03

Position 4

	Estimate	Standard error	*t* value	CrI lower	CrI upper
(Intercept)	−2.37	0.10	−24.25	−2.58	−2.15
coh	−0.03	0.01	−2.10	−0.06	0.00
len	−0.02	0.01	−1.60	−0.05	0.01
bias	−0.01	0.03	−0.42	−0.06	0.04
coh:len	−0.01	0.02	−0.28	−0.04	0.03
coh:bias	−0.00	0.01	−0.24	−0.04	0.03
len:bias	0.02	0.01	1.62	−0.01	0.04
coh:len:bias	0.01	0.01	0.57	−0.02	0.03

Position 6

	Estimate	Standard error	*t* value	CrI lower	CrI upper
(Intercept)	−2.15	0.07	−28.61	−2.31	−2.00
coh	−0.05	0.01	−3.27	−0.08	−0.01
len	−0.01	0.02	−0.54	−0.04	0.03
bias	0.00	0.01	−0.08	−0.04	0.04
coh:len	0.01	0.02	0.37	−0.04	0.05
coh:bias	−0.00	0.01	−0.36	−0.03	0.03
len:bias	0.00	0.01	0.06	−0.03	0.03
coh:len:bias	−0.00	0.01	−0.11	−0.04	0.03

$p < 0.01$), but no effects of either digging-in time or bias and no interactions. According to the model, local coherence resulted in an 11% drop in accuracy (98% vs. 87%). As questions appeared randomly on one-third of trials, the effect should be interpreted with caution. It could, however, indicate that subjects were sometimes distracted so much by the local coherence that they gave up on trying to find a globally consistent analysis and kept on pressing the space bar to advance to the next stimulus.

2.7 Alternatives to SOPARSE

Since the publication of Tabor et al.'s (2004) original results, several alternative explanations of local coherence effects that do not rely on self-organization have been offered.

The model of Gibson (2006) is able to explain the result of Tabor et al.'s (2004) Experiment 1, as well as those of Experiment 2 if local plausibility information is taken into account (cf. Gibson, 2006, p. 383). As stated above, digging-in effects need to be explained through changes in contextual frequencies. The same is true for the approach presented by Bicknell and Levy (2009) and Bicknell, Levy, and Demberg (2009), in which local coherence effects are the consequence of a clash between bottom-up prior beliefs about local structure and top-down information provided by the sentential context: if the additional material does not make the local structure less likely, the effect of the length manipulation remains unexplained. It should also be noted that the predictions of this approach change depending on how 'local' is defined, as a small enough 'window' would allow for a local coherence effect in both the short and the long variants.

Morgan, Keller, and Steedman (2010) present an approach to local coherence that is based on purely bottom-up parsing and factors in the plausibility of partial interpretations. It is, however, unclear how this model would explain the interaction between local coherence and digging-in time observed in the current study, since the local plausibility of the SVO parse should be unchanged.

Roger Levy and colleagues (Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009) take a completely different view of local coherence. Here, the phenomenon is explained in terms of a 'noisy channel' model of sentence comprehension, which introduces a measure of uncertainty as to the identity of past words. If readers doubt their current interpretation of a sentence, they may retroactively 'edit' the input in concordance with their linguistic expectations. The more likely readers are to adopt this strategy, the more processing difficulty arises. Compare (8).

(8) a. The coach smiled at the player tossed a frisbee.
    b. The coach smiled as the player tossed a frisbee.
    c. The coach smiled and the player tossed a frisbee.

(8a) is a shortened version of the input presented in Tabor et al.'s (2004) Experiment 1, while (8b, c) feature edits which result in a reading where *the player* is the agent of *tossed*. Since it is presumably easy to confuse *at* with the orthographically similar *as* or *and*, a comparatively large amount of processing difficulty is expected. Indeed, Levy et al. (2009) found that replacing *at* with *toward* in the stimulus sentence lead to decreased processing difficulty, presumably because the edits shown in (8b, c) are less likely to be made.

If the noisy channel approach is to explain the current results, one needs to take into account that the edited input as a whole must receive a grammatical analysis. The only way to achieve this for (6) would be to delete the complementizer and replace the comma with a colon:

(9) Heute weiß   man: einer   der   Generäle   belagerte   Städte …
    today knows   one   one   of the generals   besieged   cities

Since incorporating the additional adverbial material in the long conditions into this structure is impossible, it would have to be deleted completely. Given that the probability of making this extensive change to the input should be low, the model could explain the interaction observed in the current study. A problematic aspect of this view, however, comes from the noisy channel model's assumption that when an edit is made, the reader comes to believe that the input actually looked different from the start (Levy, 2008). If readers are convinced that they are processing a main

clause, the arrival of the clause-final verb should thus lead to considerable disruption, contrary to what we observed.

Yet another approach to local coherence comes from Hale (2011). He explains the disruption in terms of the parser trying to minimize processing costs: it shuns complex structures such as the reduced relative construction and may instead temporarily consider illicit analyses. In order for this model to predict the current results, it would need to specify (a) how a main clause analysis is less complex than a subordinate clause analysis and (b) how the complexity of the two analyses in question is influenced by the length manipulation.

The Sausage Machine of Frazier and Fodor (1978) and the n-gram-based approach of Crocker and Brants (2000) both feature multiple layers of syntactic processing. The former incorporates a preliminary parser which can 'see' only a limited number of words at a given time. If the incompatible part of the input is lost from view before the local coherence arrives, the pre-parser will pass the locally computed misanalysis to the 'real' parser, creating disruption. Like the Morgan et al. (2010) model, the Sausage Machine cannot straightforwardly account for the effect of the digging-in manipulation: in order for the misanalysis to become possible, material which cannot be locally integrated has to have left the pre-parser's 'viewing window'. In both the long and the short conditions, the left edge of the critical window would thus be the subject NP, and consequently there should be no processing difference.

The cascaded probabilistic model of Crocker and Brants (2000) assumes more than two processing layers. At each layer, several possible analyses of the input are generated and then passed to the next layer along with their associated probabilities. The lowest layer is a part-of-speech tagger that takes both lexical frequencies and contextual probabilities (in this case trigram probabilities) into account, much like Gibson's (2006) approach does. If this first layer favors the wrong set of tags, reanalysis becomes necessary at higher layers, explaining the disruption caused by local coherence and, possibly, lexical bias. The model could stipulate that the long and short conditions in our study differ because reanalysis is triggered at different layers. If the intruding SVO structure starts out as a verb phrase, it will not be able to accommodate the adverbial material in the long conditions and a reanalysis must take place. If the modifiers are absent, however, the structure's malformedness only becomes apparent at the clausal layer, where the complementizer demands a verb-final structure. We are somewhat skeptical of this explanation as the complementizer is encountered before the local coherence and should set up a strong expectation for a verb-final sentence in any case.

3 General discussion

We found that German readers are distracted by a locally coherent ungrammatical SVO sequence embedded in a verb-final subordinate clause, in accord with the findings of Konieczny et al. (2009). Furthermore, results showed that the addition of adverbial material between complementizer and subject modulated the local coherence effect, which can be attributed to increased digging-in time for the globally correct parse, as predicted by SOPARSE. Other proposed accounts fare less well in accounting for our results, mainly because they do not assume that structures will automatically 'solidify' over time and thus gain an advantage during competition.

Evidence that the end result of a parsing operation may not always be entirely self-consistent has been accruing for some time now. In addition to the findings of Tabor et al. (2004) and Konieczny et al. (2009), Konieczny (2005) found that readers were less likely to spot errors in sentences in which a surplus argument was part of a local coherence. It appears that plausibility considerations play a large role in the emergence of non-consistency. Compare (10), taken from Cai, Sturt, and Pickering (2011).

(10) Because it was John that Ralph threatened(,) the neighbor/camera next door recorded their conversation.

In a self-paced reading study, the authors found that reading times for the word next were elevated when the sentence contained *neighbor* instead of *camera*, but only when the comma was absent. The result can be explained through the presence of the locally coherent substring *Ralph threatened the neighbor* in the difficult version.

An alternative explanation considered by Cai et al. (2011) is the so-called dual-route account of Ferreira (2003) and Christianson, Luke, and Ferreira (2010). In this particular instantiation of the 'good enough' approach to sentence processing, it is assumed that semantics-based heuristics can override syntactic rules. The dual-route account can explain why subjects sometimes interpret implausible passive sentences such as *The dog was bitten by the man* according to their world knowledge, that is, as *The dog bit the man*, instead of following the rules of grammar. We believe that SOPARSE fails to account for the effect since no local coherence is present in these cases. An anonymous reviewer suggests, however, that SOPARSE can explain the reversal by assuming that passive structures are not highly activated, allowing the subject and object NPs to switch roles.

Meanwhile, the effect of the length manipulation observed in the current study cannot be straightforwardly explained by the dual-route account. Ferreira (2003) assumes that on the semantic route, NPs are assigned thematic roles based on their typicality. Looking at the current results, it is not clear why the presence of an adjunct should interact with typicality, especially since the additional material could not be integrated into the partial SVO parse. The dual-route account does not specify how exactly syntactic computation and semantic heuristics interact. If anything, it seems to us that the syntactic and the local semantic information should clash more strongly in the long conditions, creating more instead of less disruption.

Incidentally, our findings also shed some light on the nature of distance effects in parsing; compare sentence (11a) from Tabor and Hutchins (2004) with sentence (11b) from Ferreira and Henderson (1991).

(11) a. As the author wrote the book (describing Babylon) grew.
      b. While the boy scratched the dog (that is hairy) yawned loudly.

Both Tabor and Hutchins' Experiment 1 and Ferreira and Henderson's Experiment 5 manipulated the length of the ambiguous region in a Late Closure ambiguity. In reading times and grammaticality judgments, respectively, both studies found that stimuli with longer ambiguous regions were harder to process than those with shorter ambiguous regions. While Tabor and Hutchins attribute the increased difficulty with the long version to digging-in of the erroneous attachment between the verb and the second NP, Ferreira and Henderson offer a different explanation. Their account assumes that the verb *scratch* in (11b) has two separate thematic structures, one for the transitive and one for the intransitive reading. Both structures enter the computation when the verb is first read, but the intransitive variant starts to 'decay' in memory as a direct object becomes available immediately. When the main clause verb *yawned* arrives, the additional decay incurred by the modifier makes it harder to reactivate the intransitive thematic structure and reanalysis becomes more difficult. The same reasoning can be applied for Tabor and Hutchins' stimuli, which calls the idea of a digging-in effect into question.

That said, it is not clear how decay would explain the preemptive effect we observed in the current study. The only relevant word whose activation could have decayed, the complementizer *dass*, 'that', has no alternative reading. Furthermore, had the complementizer somehow faded from memory due to the length manipulation, participants should have been more instead of less likely

to fall victim to the local coherence. It is the digging-in assumption, combined with Kamide and Mitchell's (1999) generic sentence-structure frame, that makes the correct prediction in this case.

Acknowledgements

Funding

Notes

1. More precisely, the NP *the player* will need to be reanalyzed as the head of the reduced relative clause in any case, but the relationship with *smiled* at should remain intact.
2. An anonymous reviewer informs us that the model of van der Welde and de Kamps (2006) also implements self-organized parsing.
3. Konieczny et al. (2009) employed a very similar manipulation, but in their design the participle acted as an adverb in the globally correct parse.
4. In principle, it is possible for the position of the critical participle in the experimental materials to be occupied by a verb. That verb would need to be intransitive or take a sentential complement, given the aforementioned requirement that all non-sentential arguments precede it. However, all participles in the current study were obligatorily transitive and unable to accept sentential complements under the verb reading.
5. See http://tedlab.mit.edu/~dr/Linger/
6. More precisely, encountering a verb at the participle's position would need to be less likely than encountering an adjective.
7. The noise feature is taken from the Unification Space model of Vosse and Kempen (2000) and needed by Tabor and Hutchins (2004) to explain their own grammaticality judgment results.

Supplemental Material are available at http://las.sagepub.com/supplemental.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 225–242). Hillsdale, NJ: Lawrence Erlbaum.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999–0. Retrieved from http://CRAN.R-project.org/package=lme4

Bicknell, K., & Levy, R. (2009). A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 665–673). *Boulder, CO*: Association for Computational Linguistics.

Bicknell, K., Levy, R., & Demberg, V. (2009). Correcting the incorrect: Local coherence effects modeled with prior belief update. In I. Kwon, H. Pritchett, & J. Spence (Eds.), *Proceedings of the 35th annual meeting of the Berkeley linguistics society* (pp. 13–24). *Berkeley, CA*: Berkeley Linguistics Society.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, *26*(2), 211–252.

Cai, Z. G., Sturt, P., & Pickering, M. J. (2011). The effect of nonadopted analyses on sentence processing. *Language and Cognitive Processes*. doi:10.1080/01690965.2011.599657.

Chomsky, N. A. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538–544.

Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, *29*(6), 647–669.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Doctoral dissertation. Bloomington, IN: Indiana University Linguistics Club.

Frazier, L., & Fodor, J. D. (1978). The Sausage Machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210.

Geyken, A. (2006). *DWDS-Corpus*. In C. Fellbaum (Ed.), *Collocations and idioms: Linguistic, lexicographic, and computational aspects* (pp. 23–40). London, UK: Continuum Press.

Geyken, A., Klein, W., & Wolz, R. (2004). Projekt Digitales Wörterbuch der deutschen Sprache des 20. Jh. In *Jahrbuch der BBAW 2003* (pp. 238–320). Berlin, Germany: Akademie Verlag.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, *54*(3), 363–388.

Hale, J. T. (2011). What a rational parser would do. *Cognitive Science*, *35*(3), 399–443.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes*, *14*(5/6), 631–662.

Konieczny, L. (1996). *Human sentence processing: A semantics-oriented parsing approach*. Doctoral dissertation, Albert-Ludwigs-Universität zu Freiburg.

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual meeting of the Cognitive Science Society* (pp. 1178–1183). Mahwah, NJ: Lawrence Erlbaum.

Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes*, *12*(2/3), 307–348.

Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. A. (2009). Local syntactic coherence interpretation: Evidence from a visual world study. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1133–1138). Austin, TX: Cognitive Science Society.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 234–243). Madison, WI: Omnipress.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. In *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21086–21090. (Correction in Proceedings of the National Academy of Sciences of the United States of America, 107, 5260)

Morgan, E., Keller, F., & Steedman, M. (2010). A bottom-up parsing model of local coherence effects. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, *Portland, OR* (pp. 1559–1564).

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/.

Sorensen, T., & Vasishth, S. (2014). A tutorial on fitting Bayesian linear mixed models using Stan. Submitted to *Journal of Mathematical Psychology*, arXiv, 1506.06201v1.

Speyer, A. (2004). Competing constraints on Vorfeldbesetzung in German. In B. Shaer, W. Frey, & C. Maienborn (Eds.), *Proceedings of the Dislocated Elements Workshop. ZASPiL 35, Vol. 2.* (pp. 519–541).

Stan Development Team (2014). *Stan Modeling Language: User's Guide and Reference Manual*. Version 2.5.0.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

van der Welde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavorial and Brain Sciences*, *29*(1), 37–108.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. 4th Edition. New York, NY: Springer.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143.

Appendix: Experimental Materials

As in Tabor, Galantucci, & Richardson's (2004) Experiment 2, comparing an ambiguous with an unambiguous word (as in the case of tossed vs. thrown) was avoided by using a critical word with the same surface form in all conditions.

1. Es wurde berichtet, dass (während der Krise abermals) eine/einige der Banken beantragte Kredite aus reinem Leichtsinn bewilligte(n), anstatt größere Bargeldreserven anzulegen.

2. Heute weiß man, dass (entlang der Nachschubrouten anfänglich) einer/einige der Generäle belagerte Städte mit großer Verbissenheit verteidigte(n), um Zeit zu gewinnen.

3. Man fand heraus, dass (trotz des Munitionsmangels schließlich) eines/einige der Schiffe zerstörte Brücken aus sicherer Entfernung beschoss(en), um Trümmer zu zerstreuen.

4. Die Presse schrieb, dass (ohne richterliche Anordnung offensichtlich) eine/einige der Abteilungen zensierte Akten nach dem Putsch vernichtete(n), um weitere Nachforschungen auszuschließen.

5. Es wurde bekannt, dass (im untersuchten Zeitraum offenbar) einer/einige der Mitarbeiter übersetzte Texte nach eigenem Gutdünken kürzte(n), um Papier zu sparen.

6. Die Öffentlichkeit erfuhr, dass (abseits des Prozesses überdies) einer/einige der Agenten entlarvte Spione dem Auftrag gemäß umbrachte(n), damit sie nicht redeten.

7. Die Polizei meldet, dass (trotz der Videoüberwachung tatsächlich) einer/einige der Zauberkünstler überraschte Zuschauer abseits der Veranstaltung bestahl(en), als sie abgelenkt waren.

8. Die Zeitung schrieb, dass (gerade am wichtigen Ausstellungsabend) eine/einige der Künstlerinnen schockierte Kritiker mit voller Absicht anrempelte(n), um Aufmerksamkeit zu erregen.

9. Jüngste Untersuchungen ergaben, dass (entgegen der Behauptung offensichtlich) einer/einige der Vorstände diktierte Briefe in regelmäßigen Abständen vergaß(en), weil er betrunken war/sie betrunken waren.

10. Edgar musste feststellen, dass (nur aufgrund von Faulheit) einer/einige der Lektoren markierte Textstellen in vielen Fällen überlas(en), anstatt sie zu korrigieren.

11. Später wurde klar, dass (angesichts seiner bitteren Armut) einer/einige der Kellner entsorgte Abfälle nach der Arbeit mitnahm(en), was er nicht durfte/sie nicht durften.

12. Zeitungen berichten übereinstimmend, dass (östlich des Grenzpostens immerhin) einer/einige der Bautrupps blockierte Straÿen gemäß amtlicher Anweisung freiräumte(n), nachdem Frieden eingekehrt war.

13. Lobend wurde erwähnt, dass (übereinstimmenden Berichten zufolge zumindest) einer/einige der Soldaten evakuierte Zivilisten nach der Explosion beruhigte(n), um Panik zu verhindern.

14. Historiker fanden heraus, dass (fernab der Zivilisation schließlich) einer/einige der Piraten erbeutete Münzen auf seinen/ihren Raubzügen einschmolz(en), um goldene Säbel herzustellen.

15. Die Geschäftsleitung erfuhr, dass (allem Anschein nach leider) einer/einige der Mechaniker reparierte Flugzeuge nicht immer ordnungsgemäß überprüfte(n), was zur Entlassung führte.

16. Die Videoaufnahmen beweisen, dass (mangels anderer Gelegenheiten tatsächlich) einer/einige der Praktikanten kopierte Dokumente während seiner/ihrer Mittagspause durchlas(en), um Informationen zu sammeln.

17. Im Bericht steht, dass (trotz des gesetzlichen Feiertags) eine/einige der Firmen installierte Lampen dem Vertrag entsprechend austauschte(n), was den Mieter freute.

18. Man erfuhr später, dass (inmitten der Unruhen offensichtlich) einer/einige der Spitzel enttarnte Informanten durch raffinierte Tricks warnte(n), damit sie fliehen konnten.

19. Es ist verständlich, dass (trotz spezieller Ausbildung anfänglich) einer/einige der Psychiater hypnotisierte Patienten nur äußerst ungern befragte(n), wenn es Gewalttäter waren.

20. Man muss bedenken, dass (nach wie vor auch) eine/einige der Maschinen zerkleinerte Marmorblöcke unter großem Zeitaufwand sortierte(n), was Verzögerungen unvermeidlich machte.

21. Die Befragung ergab, dass (innerhalb der Mauern glücklicherweise) einer/einige der Wärter misshandelte Häftlinge Tag um Tag beschützte(n), wofür er belohnt wurde/sie belohnt wurden.

22. Es wurde verschwiegen, dass (während der Blockade offenbar) einer/einige der Chirurgen amputierte Körperteile als letzte Lösung einfror(en), um Transplantationen zu ermöglichen.

23. Anwohner konnten beobachten, dass (unweit des Palastes plötzlich) einer/einige der Bagger erweiterte Fahrrinnen binnen kurzer Zeit zuschüttete(n), weil ein Planungsfehler vorlag.

24. Niemand hatte geahnt, dass (angesichts verzweifelter Aussichten schließlich) einer/einige der Rebellen entführte Diplomaten entgegen der Befehle freiließ(en), weil er Mitleid hatte/sie Mitleid hatten.

25. Der Studienleiter berichtet, dass (trotz der Infektion immerhin) eines/einige der Medikamente gefährdete Probanden bei äußerlicher Anwendung immunisierte(n), was neue Rätsel aufgibt.

26. Der Chef hörte, dass (nach mehrfacher Ermahnung wenigstens) einer/einige der Köche frittierte Teigtaschen mit großem Pflichtbewusstsein aussortierte(n), wenn sie geplatzt waren.

27. Das Personal berichtete, dass (aufgrund geringerer Risiken neuerdings) einer/einige der Tierärzte kastrierte Katzen unter örtlicher Betäubung operierte(n), weil sie gleichmütiger waren.

28. Der Scheich hörte, dass (in der Sitzung zumindest) einer/einige der Architekten modernisierte Gebäude aus ästhetischer Überzeugung verteufelte(n), was ihn/sie sympathisch machte.
29. Der Vertreter sagte, dass (entgegen jeder Erwartung offenbar) eines/einige der Geräte enttäuschte Verbraucher schließlich doch noch überzeugte(n), was ihn zufrieden mache.
30. Unabhängige Beobachter erzählen, dass (kurz vor Verhandlungsschluss erwartungsgemäß) einer/einige der Zeugen bestätigte Berichte unter bitteren Tränen widerrief(en), weil er bedroht wurde/sie bedroht wurden.
31. Die Kontrolle ergab, dass (trotz vieler Beschwerden offensichtlich) einer/einige der Autoren veröffentlichte Artikel immer noch voreilig kritisierte(n), anstatt sie sorgfältig durchzulesen.
32. Die Menschen wussten, dass (angesichts der Zwangslage zuletzt) einer/einige der Aufständischen versteckte Lebensmittel ungeachtet der Konsequenzen herausgab(en), wenn eine Durchsuchung stattfand.