Secret Agents or Sleeping Beauties: What Happens to Repaired Constituents?

Noa Shuval (noa.shuval@gmail.com)
Laboratoire de Psychologie et Neuropsychologie Cognitives, CNRS, Université Paris Descartes,
71 ave Edouard Vaillant, 92774 Boulogne-Billancourt, France

Lars Konieczny (lars.konieczny@frias.uni-freiburg.de)
FRIAS, Universität Freiburg Starkenstr. 44 79098 Freiburg, Germany

Barbara Hemforth (barbara.hemforth@parisdescartes.fr)
Laboratoire de Psychologie et Neuropsychologie Cognitives, CNRS, IUPPD, Université Paris Descartes,
71 ave Edouard Vaillant, 92774 Boulogne-Billancourt, France

Abstract

In two experiments we investigated the comprehension of sentences with repaired NPs. In our first experiment, we applied an acceptability task after speeded auditory presentation of French versions of sentences with and without repairs like "*I will go to (the butcher, uh no,) the baker. I need some bread/meat*". While repairs led to reduced acceptability for consistent continuations, the inconsistent continuation was more acceptable when a compatible but repaired constituent had been mentioned before, suggesting that the to-be-repaired constituent was not fully overwritten by the correction. In our second experiment, the visual world paradigm was used to auditorily present participants with the stimuli compiled for Experiment 1, while they looked at corresponding visual stimuli. This time, evidence from eye fixation patterns suggests that the to-be-repaired constituent was actually suppressed online during sentence processing. To settle this contradicting evidence we would like to suggest that the acceptability judgments are mainly the result of offline reconstruction of memory traces following Gimenes et al., (2009).

Keywords: Sentence Processing; Disfluencies; Acceptability Judgments; Visual World Paradigm

Introduction

The human sentence processing system has to be extremely robust since it does not only have to cope with highly standardized and edited to correct input, but very often (probably more often than not) also with deficient input caused by various, often non-linguistic, situational factors. In this paper, we will look at the comprehension of repaired utterances like (1). It has been proposed that disfluencies such as silent or filled pauses or repairs may lead to undesirable effects in sentence processing, leaving misparses harder to detect, possibly by providing cues which are interpreted as prosodic structuring information (Bailey & Ferreira, 2003; Maxfield, Lyon, & Silliman, 2009). For repairs, it has been proposed that the to-be-repaired constituent may continue influencing listeners' comprehension, the so-called lingering effect (e.g., Lau & Ferreira, 2005). It is this latter effect that we investigate in our experiments.

(1) *J'irai chez le boucher, euh non, le boulanger. J'ai besoin de pain.*
I'm going to the butcher, eh no, the baker. I need some bread.

Disfluencies are highly frequent in natural language production. They include editing terms such ***uh*** and ***um*** as well as repeats ("I – uh - I wouldn't", e.g. Clark & Wasow, 1998) as well as revisions. Typically, in spoken language, disfluencies can be found in about six out of 100 words (Fox Tree, 1995). In the corpus used by Levelt (1983), 25 % of the annotated disfluencies were repairs similar to the structures under investigation in our studies. Of these, 62 % included editing expressions like Dutch versions of "I mean" or "that is" or mostly (30 % of all repairs) the Dutch version of "uh". Since disfluencies in general and repairs in particular are so frequent, listeners have to find ways to process them, they have to detect the disfluency, see what the problem is, and edit out the part of speech to-be-repaired to arrive at the intended meaning of the utterance.

Research on error processing in spelling has provided evidence, that recently processed incorrect information (Brown, 1988; Dixon & Kamisnka, 2007; Jacoby & Hollingshead, 1990) may affect subsequent performance even in cases, where the error has been explicitly recognized as such (Perruchet, Rey, Hivert, & Pacton, 2006). Editing out explicitly marked repairs may, equally, not always work perfectly well. Lau and Ferreira (2005, see also Bailey & Ferreira, 2003; Ferreira, Lau, & Bailey, 2004) claim that the to-be-repaired constituent in repetitions and corrections introduces lexical content and local syntactic structure not fully overwritten by the correction. They studied a disfluency involving the repair of a verb (like chosen vs. selected) in sentences like (2 a, b).

(2) *The little girl a. chosen-uh/b. picked-uh selected for the role celebrated with her parents and friends.*

Sentences like these, with verbs like "selected" which are ambiguous between a main verb and a past participle reading, usually lead to comprehension difficulty (e.g.,

increased reading times), on the disambiguating prepositional phrase ("for the role"), in particular when the verb is biased for a simple main verb imperfect reading. This garden-path disappeared when the ambiguous verb was preceded by an unambiguous past participle ("chosen").

Although a repaired constituent such as *chosen* in (2) should be discarded from the representation of the sentence, it has been shown across a variety of constructions that they can influence offline acceptability judgments (e.g. Lau & Ferreira, 2005). Garden paths as well as semantically inconsistent sentences are judged as more acceptable when containing a repaired element that would have rendered them unambiguous or semantically consistent, respectively.

However, since most of the published studies only present offline data, we cannot be sure that the acceptability judgments reflect an online lingering effect of the repaired element, as previously suggested, or whether they may be due to offline reconstruction of memory traces. Gimenes, Rigalleau, & Gaonac'h (2009), for example, show that positive acceptability judgments do not necessarily mean that no problem has been detected online. In their study, conducted in French, they used a self-paced non-cumulative reading paradigm to present participants with doubly center-embedded object relative sentences. Participants were asked, among other things, to evaluate each sentence. Gimenes et al. compared two conditions: In one condition all three VPs were present in the sentence (3), whereas in the other one the second VP (VP2) was missing (4).

(3) The Mexican meal that the gastronomic critic that the journal hired tasted in the new restaurant had a strange smell.
(4) The Mexican meal that the gastronomic critic that the journal hired had a strange smell.

While the missing-VP2 sentences were rated as better than the All-VPs sentences, they presented longer reading times on the last VP compared to the all-VPs sentences. This means that participants preferred the sentences with the VP2 omissions, even though there was still an online sensitivity to that omission. In other words, the online difficulty was not reflected in the acceptability judgments.

Aim of current study

In our study we wished to find out whether the influence of repaired constituents on offline acceptability judgments are due to online lingering effect (i.e. the parallel construction of a phrase structure compatible with the repaired element) or to offline reconstruction. In order to do so, we chose to first, try and replicate earlier results associated with repairs in an offline experiment in which acceptability judgments were collected after presenting participants with speeded sentence with NP repairs such as "*I will go to (the butcher, uh no,) the baker. I need some bread/meat*". In these constructions, an inconsistency (I go to the baker. I need some meat.) is more acceptable when preceded by a semantically consistent but repaired element (I go to the butcher, uh no, the baker) (Hemforth, Pynte, & Bellengier, 2007). Second, we used the visual world paradigm in order to investigate online processing of these constructions.

Experiment I

In our first experiment, we expect to find evidence for the so-called lingering effect of the to-be-repaired constituent (Bailey & Ferreira 2003, Hemforth et al., 2007) in the acceptability judgments.

Design and procedure:

Participants: 24 native French undergraduate students, from the Paris Descartes University participated in this experiment in exchange for course credits.

Materials: We constructed 16 items each in four conditions as in examples (5a, b) and (6a, b).

(5) Consistent/inconsistent target
J'irai chez le boulanger, J'ai besoin de a. pain/b. viande.
I'm going to the baker, I need some a. bread/b. meat.

(6) Consistent/inconsistent repair
J'irai chez le boucher, euh non, le boulanger.
J'ai besoin de a. pain/b. viande.
I'm going to the butcher, eh no, the baker.
I need some a. bread/b. meat.

In half of the sentences participants were presented with disfluencies in the form of an NP replacement. The second experimental factor was the consistency of the last word of the second sentence with the contents of the first sentence: The object of the second sentence made this sentence either consistent (5a, 6a) or inconsistent (5b, 6b) as a continuation of the first sentence. Specifically, in the repair condition, the consistent continuation was compatible with the repair, where as the inconsistent continuation was consistent with the to-be repaired item. For half of the participants, the NPs were exchanged to control for plausibility effects.

We presented participants with synthesised sentences using the Acapela© Text to Speech software in order to make our experimental design as comparable as possible to Hemforth et al. (2007). All utterances were produced at 22 kHz, using a female voice (Claire). Two sentences were synthesized for each item. The second sentence was speeded up by 30%, using the Audacity® software, keeping all other basic parameters like frequency unchanged. The synthesized materials were inspected by the authors as well as several native speakers of French and judged as highly natural and easily understandable. In a separate experiment using the same materials but with comprehension questions, participants answered questions correctly in more than 97%

of the cases in all conditions (Shuval & Hemforth, submitted).

Procedure: Participants were told that they would hear sentences generated by a computer. For each experimental item, a visual signal indicated that a sentence would be played. Once the sentence was completed, the participant had to judge its grammaticality on an explicit 1-4 scale where 1 corresponded to a very poor sentence and 4 to a very good one. Judgments were automatically recorded by the experimental software (ExperimentBuilder© by SR Research). Each experimental item was presented in one of the four experimental conditions across participants. Eight randomized lists were prepared including 15 fillers. The filler sentences varied with respect to their syntactic and semantic acceptability.

Predictions: Based on earlier evidence, we expect to find that the to-be repaired constituent influences acceptability judgments, so that sentences with inconsistent continuations that are, however, consistent with the to-be repaired constituents, should be judged as more acceptable than sentences without a replacement. We therefore predict an interaction of the experimental factors (repair vs. no repair, and consistent vs. inconsistent).

Results

Acceptability judgments: Repaired sentences were judged less acceptable than unrepaired sentences, though only marginally so across participants ($F1_{1,23}$=3.05, $p < .10$; $F_{1,15}$=4.80, $p < .05$). Inconsistent sentences were generally judged less acceptable than consistent sentences ($F_{1,23}$=39.94, $p < .001$; $F_{1,15}$=45.83, $p < .001$). A reliable Repair*Consistency interaction ($F_{1,23}$=10.75, $p <.01$; $F_{1,15}$= 41.05, $p < .001$) was established: Acceptability of consistent sentences decreased in sentences with repairs ($F_{1,23}$=16.11, $p < .01$; $F_{1,15}$=30.33, $p < .001$), whereas acceptability of inconsistent sentences increased in the Repair condition though only reliably so across items ($F_{1,23}$=1.854, $p > .18$; $F_{1,15}$= 6.01, $p < .03$).

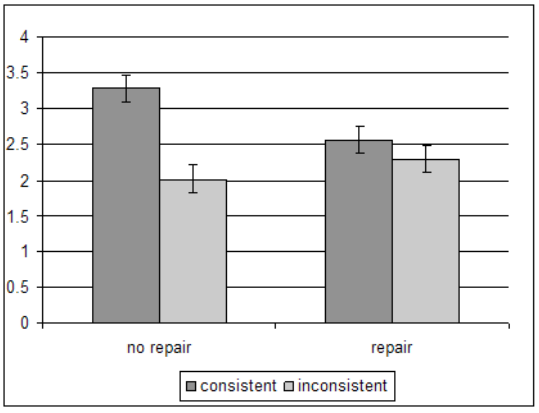

Figure 1: Acceptability judgments

Judgment times: Figure 2 shows the judgment times for all four conditions. Judging the acceptability took reliably longer for repaired sentences than for unrepaired sentences ($F1_{1,23}$=6.15, $p < .03$; $F2_{1,15}$=7.11, $p < .02$). No other effects turned out to be reliable.

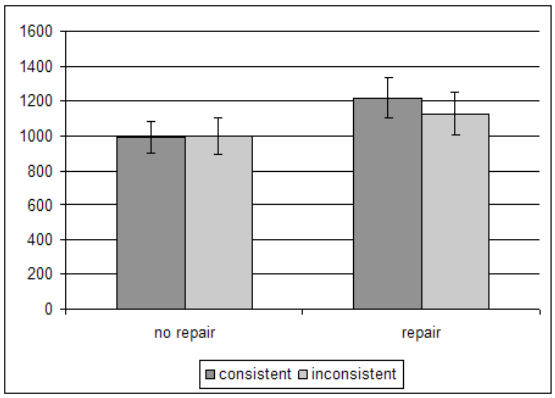

Figure 2: Judgments times

Discussion

Obviously, the to-be-repaired items affect the acceptability of the sentence. Interference is found in sentences with repairs leading to increased judgment times (although these judgment times were taken off-line, increased sentence length in the repaired conditions might possibly contribute to the increased judgment times). It shows up in acceptability judgments in the consistent condition, reducing acceptability, as well as in the inconsistent condition, enhancing acceptability. Inconsistent continuations become more acceptable following repairs presumably due to the interfering constituent, whereas consistent continuations are becoming less acceptable possibly for the same reasons. (A plausible alternative interpretation of the decreased acceptability of consistent targets following repairs may simply be that repaired constructions are less acceptable than non-repaired ones.) Following Lau and Ferreira (2005), this interfering effect might be due to the lingering of a partial interpretation compatible with the to-be-repaired constituent.

There are at least two ways to conceptualize "lingering", however. One possibility is that the partial interpretations that should have been suppressed remain active and thus influence the processing of the sentence at any moment. A second possibility compatible with the current data would be that the partial interpretations are actually suppressed from current working memory staying dormant until they are reactivated in sentence wrap-up processes in cases of emergency (garden paths, inconsistencies) or in situations with heightened cognitive load (such as speeded presentations). Active lingering predicts that the to-be-repaired item interferes measurably with on-line sentence processing, whereas dormant lingering predicts mostly off-line effects. In order to find out whether the to-be-repaired constituents behave more like secret agents waiting to jump

in whenever possible or like sleeping beauties, only to be woken up when necessary, we ran Experiment 2

Experiment II

In this experiment we wanted to tap into the time course of comprehension of the repaired sentences we used in Experiment 1. In order to do so, we used exactly the same items as the basis for a visual world experiment (Tanenhaus, Spivey-Knowlton, Eberhard, Sedivy, 1995). In this experiment, we made the linguistic input as natural as possible, using a natural speaker and non-speeded presentation.

Design and Procedure: In our experiment, 32 native French undergraduate students, from the Paris Descartes University with normal hearing and normal or corrected to normal vision listened to auditorily presented sentences that were constructed for Experiment 1. This time, they were recorded by a natural speaker and the critical regions were cross-spliced using the PRAAT speech software for Windows© (Boersma & Weenink, 2009) in order to control for intonation differences between conditions (5, 6; repeated here).

(5) Consistent/inconsistent target
J'irai chez le boulanger, J'ai besoin de a. pain/b. viande.
I'm going to the baker, I need some a. bread/b. meat.

(6) Consistent/inconsistent repair
J'irai chez le boucher, euh non, le boulanger.
J'ai besoin de a. pain/b. viande.
I'm going to the butcher, eh no, the baker.
I need some a. bread/b. meat.

Each item was accompanied by a visual stimulus depicting four objects (Figure 3a): the consistent (bread) and the inconsistent (meat) objects together with a semantically related object (lettuce) and an unrelated object (glasses).

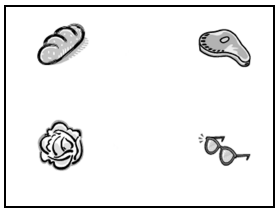

Figure 3a: Visual Stimulus

We tracked eye movements with an SR Research EyeLink© II eye-tracker. For each participant the dominant eye was tracked as determined by the Miles (1930) test. The participants were presented with the visual stimuli on a 21'' screen and with the corresponding auditory stimuli via earphones. Each session started with a calibration of the eye tracking system.

Each trial set began with a short tone to mark the beginning of the trial. This tone was immediately followed by an introductory slide that disappeared after four seconds (see Figure 3b). It was replaced by a blank screen with a centered fixation cross ''+''. The participants were asked to fixate the cross while pressing the space bar in order to control for the calibration before the critical image (drift correction). The drift correction also guaranteed that the very first fixation was always on the center of the screen. The experimental slide appeared simultaneously with the beginning of the auditory stimulus. Following the experimental slide, a multiple choice comprehension question was presented. Each experimental session took about 30 minutes.

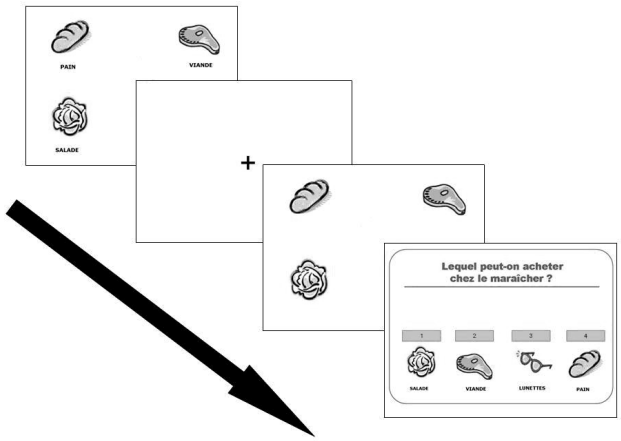

Figure 3b: Order of presentation

Predictions: The active or dormant nature of the lingering effect of the repaired constituent should be reflected in eye movement behavior elicited by the inconsistent repair condition in the following manner: In any case, we expect more anticipatory eye movements to the consistent object than to the inconsistent one even before the onset of the target noun (bread/meat) (Altmann & Kamide, 1999). In the case of an active lingering effect, we expect more anticipatory eye movements to the inconsistent objects than to the related baseline objects (lettuce). Following the onset of the target noun, eye movement to the inconsistent objects should be facilitated (start earlier and/or be more frequent) in the repair condition compared to the non-repair condition. If, however, the lingering effect is dormant, we expect to see evidence of suppression as follows: Suppression of the inconsistent object should lead to similar or lower proportions of fixations to the inconsistent objects compared to the related-baseline objects before the onset of the target noun. After the onset of the target noun, fixations to the inconsistent objects should be slowed and reduced compared to the non-repair condition.

Results:

We calculated the log odds for gazes on the target object at each time step, using the formula in (7).

(7) $Y = \log2(P_{(CO)} / P_{(ICO)})$

$P_{(CO)}$ refers to the likelihood of a gaze on the consistent object and $P_{(ICO)}$ to the likelihood of a gaze on the inconsistent object.

Figure 4 shows the time course of fixations starting with the onset of the verb of the second sentence. The vertical line at 663ms marks the mean onset of the target noun. Values below zero represent a higher number of fixations to the inconsistent object (meat), values above zero represent more looks to the consistent object (bread). Error bars mark the 95 % confidence interval. Comparing the inconsistent conditions (filled circles: non-repaired, empty circles: repaired), we do not see any indication of an active lingering effect. However, we can see clear effects of inhibition or suppression in the case of sentences with repairs. The inconsistent target attracts reliably less fixations in the repaired sentences than in the non-repaired sentences. Moreover, even before target onset, there are more fixations on the consistent target in the repair conditions, giving an additional indication of suppression.

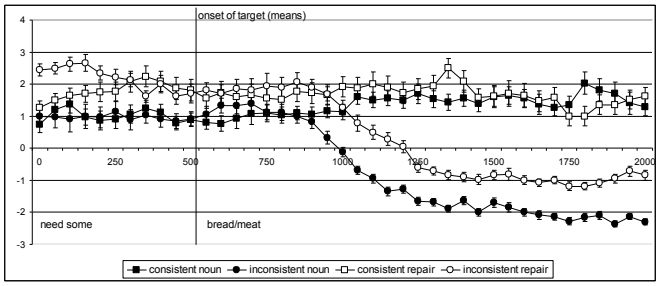

Figure 4: Log2odds of fixation probabilities of consistent/inconsistent objects

In Figure 5 values above zero represent more fixations to the inconsistent object (meat), values below zero represent more fixations to the related object (lettuce), which serves as baseline. The vertical line at 663ms marks the mean onset of the target noun. Error bars mark the 95 % confidence interval. Comparing the inconsistent conditions (filled circles: non-repaired, empty circles: repaired), we see further evidence for suppression in the case of sentences with repairs. Before the inconsistent target is mentioned, it is not fixated more often than the semantically related but unmentioned object (all Fs < 1). In the region between 500 and 1000 ms, it is even fixated less often ($F_{1,31}= 4.99$, $p<.04$; $F_{21,15}= 6.85$, $P<.02$).

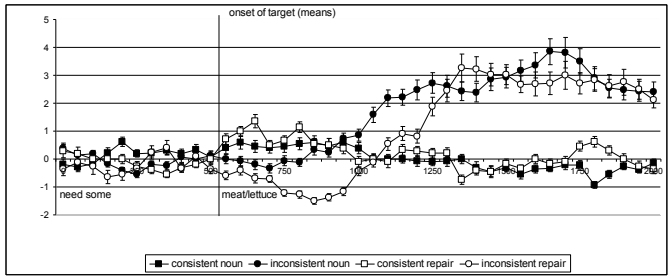

Figure 5: Log2odds of fixation probabilities of inconsistent/related objects

Discussion

In sum, although acceptability judgments in the constructions under investigation in this paper show a lingering effect of repairs, there is no online evidence for increased accessibility of the inconsistent target in repair sentences. In other words, there is no evidence for active lingering. To the contrary, the inconsistent target seems to be even less accessible in sentences with repairs than in sentences where no semantic expectation for it has ever been generated. The inconsistent target is no more accessible (and at times even less so) than the baseline object (semantically related to the target objects, but not predicted by the first sentence). This pattern of results suggests that the inconsistent target may have undergone suppression (Gernsbacher, 1990). Shuval & Hemforth (submitted) present evidence that reducing the quality of the input by using synthesized speech also reduces the effect of suppression, so that we can actually conclude that the naturalness of the materials of the current experiment plays an important role for our results.

General Discussion

In this paper we presented two experiments investigating the processing of repaired versus non-repaired sentences. In the first experiment, sentences were presented in a speeded-up version and participants had to provide acceptability judgments at the end of the sentence. With this paradigm and task, we found evidence for interference of the to-be-repaired constituent in acceptability judgments as well as in acceptability judgment times similar to the experiments reported in Hemforth et al. (2007) or Lau and Ferreira (2005). This evidence would be fully compatible with a lingering effect of the to-be-repaired constituent.

In our second experiment, we measured online comprehension of repaired versus non-repaired sentences using the Visual World Paradigm. In this experiment, the auditory stimuli were presented in normal speed. The participants' task was to listen to the sentences for comprehension and to answer comprehension questions following the sentence. In this experiment, we did not find any evidence for active online lingering of the to-be-repaired constituent. To the contrary, the to-be-repaired constituents seemed to be suppressed or inhibited, as

fixations to target items associated to them were fewer following repairs. Inconsistent targets were at times even less fixated than the semantically related baseline object. This pattern of results is not compatible with active lingering, which should have resulted in increasing the accessibility of the inconsistent target after repairs as well as compared to the baseline.

Following Gimenes et al (2009), we propose that offline acceptability judgments do not necessarily reflect the same processes at stake while a sentence is processed. During sentence wrap-up, in particular in cases of increased cognitive load and reduced perceptibility (speeded presentation) memory traces not active in current working memory may be reactivated in order to arrive at a decent judgment. This is to be expected in particular when the judgment is particularly difficult, as for garden-path sentences or in the case of inconsistencies. In "ordinary" circumstances, repairs work highly efficiently. The to-be-repaired constituent is actually suppressed from the current space of interpretations given good quality of the input and sufficient processing time. In more demanding situations, comprehenders may just content themselves with "good enough" representations (Ferreira et al. 2002). This would be compatible with an "any-time" algorithm rendering good enough solutions within restricted time as it has been proposed for a variety of cognitive processes (e.g.: Horsch & Poole, 1998). Repairs do thus not act as a secret agent, interfering with online processing, but more like the sleeping beauty, just to be woken up to make things mend in the end.

Acknowledgments

We would like to thank Joël Pynte for his highly valuable contribution to many discussions held at the beginning of this project. Many thanks to Catherine Rigaut for helping us prepare the materials for the visual world experiment. We would also like to thank the anonymous reviewers for their insightful comments.

References

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

Bailey, K.G.B, & Ferreira, F. (2003). Disfluencies influence syntactic parsing. *Journal of Memory and Language, 49*, 183-200.

Bailey, K., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray and R. L. Hill, *Eye Movements: A Window on Mind and Brain*, Elsevier LTD, 486-500.

Boersma, P. & Weenink, D. (2009): Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/

Brown, A. S. (1988). Encountering misspellings and spelling performance: Why wrong isn't right. *Journal of Educational Psychology*, *80*, 488-494.

Clark, H.H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201-242.

Dixon, M., & Kaminska, Z. (2007). Does exposure to orthography affect children's spelling accuracy? *Journal of Research in Reading, 30,* 184-197.

Ferreira, Fernanda, Bailey, Karl.G.D, , Ferraro, Vittoria (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferreira, F., Lau, E.F., & Bailey, K.G.D. (2004). Disfluencies, parsing, and tree-adjoining grammars. *Cognitive Science*, 721-749.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34,* 709–738.

Gernsbacher, M.A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gimenes, M., Rigalleau, F. & Gaonac'h D. (2009). When a Missing Verb Makes a French Sentence more Acceptable. *Language and Cognitive Processes, 24 (3),* 440-449.

Hemforth, B., Pynte, J., & Bellengier, E. (2007). Making sense out of nonsense : The acceptability of repairs. In H. Christiansen & J. Villadsen, *Proceedings of the 4th International Workshop on Constraints and Language Processing*. Roskilde : CSRR 113, pp. 93-100.

Horsch M., C. and Poole, D. (1998). An Anytime Algorithm for Decision Making under Uncertainty, *In Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI-98), Madison, Wisconsin, USA, July 1998*, pp 246-255.

Jacoby, L. L., & Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology*, *44*, 345-358.

Lau, E., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, 2005, 20 (5), 633–666.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1):41-104.

Maxfield, N.D., Lyon, J.M., & Silliman, E.R. (2009). Disfluencies along the garden path: Brain electrophysiological evidence of disrupted sentence processing. *Brain and Language, 111,* 86-100.

Perruchet, P., Rey, A., Hivert, E. & Pacton, S. (2006). Do Distractors Interfere With Memory for Study Pairs in Associative Recognition ? *Memory and Cognition*, 34, 1046-1054.

Shuval, N. & Hemforth, B. (Submitted). Understanding repairs: The role of task, visual and auditory stimuli in the visual world paradigm.

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.