Speech planning interferes with language comprehension: Evidence from semantic illusions in question-response sequences

Mathias Barthel
Humboldt University Berlin
`mathias.barthel@hu-berlin.de`

Abstract

In conversation, speakers need to plan and comprehend language in parallel in order to meet the tight timing constraints of turn taking. Given that language comprehension and speech production planning both require cognitive resources and engage overlapping neural circuits, these two tasks may interfere with one another in dialogue situations. Interference effects have been reported on a number of linguistic processing levels, including lexico-semantics. This paper reports a study on semantic processing efficiency during language comprehension in overlap with speech planning, where participants responded verbally to questions containing semantic illusions. Participants rejected a smaller proportion of the illusions when planning their response in overlap with the illusory word than when planning their response after the end of the question. The obtained results indicate that speech planning interferes with language comprehension in dialogue situations, leading to reduced semantic processing of the incoming turn. Potential explanatory processing accounts are discussed.

1 Introduction

When speakers are in conversation, they notoriously take turns at talking, switching their roles of speaker and listener within short intervals of time (Sacks et al., 1974). For the greatest part of the conversation, only one of the speakers talks while the other stays silent, and stretches of mutual silence and overlapping talk are mostly very brief (Heldner and Edlund, 2010; Stivers et al., 2009). Among the greatest driving forces for fast responses, next to the possibility of completely missing out on the turn, is the semiotics of turn-timing, whereby long gaps are interpreted to be meaningful, signalling, for example, reduced willingness to comply with a request (Kendrick and Torreira, 2014; Roberts and Francis, 2013; Roberts et al., 2011). In order to achieve this remarkably precise orchestration of speaking turns, the next speaker needs to start planning his utterance while the current speaker is still delivering her turn. Planning-in-overlap has indeed been found to be the default strategy of speech planning in conversational situations, where speakers start to plan their response as soon as they can anticipate the message of the incoming turn (Barthel and Levinson, 2020; Barthel et al., 2016, 2017; Bögels et al., 2015; Bögels, 2020; Corps et al., 2018). While planning in overlap makes seamless responses possible (Barthel, 2020; Levinson and Torreira, 2015), it comes with the cost of increased processing load during speech planning as compared to planning during the silence between turns (Barthel and Sauppe, 2019). That means that, in dialogue, processing load in next speakers usually peaks just before turn transitions, which seems reasonable for two related reasons. Firstly, turn-transitions are dual-task situations, with response planning being executed during ongoing language comprehension. And secondly, the related nature of the two tasks can create interference between them, making them less efficient as they become computationally harder. Such interference effects can occur on any possible level of language processing, from lexical selection over word form retrieval and phonetic encoding down to motor preparation (Abdel Rahman and Melinger, 2019; Barthel and Levinson, 2020; Boiteau et al., 2014; Bürki et al., 2020; Fargier and Laganaro, 2016; He et al., 2021; Jescheniak et al., 2014; Konopka, 2012; La Heij et al., 1990; Meyer, 1996; Schriefers et al., 1990, inter alia). These cross-talk effects have been assumed to be rooted in shared representations for production and comprehension and/or partly overlapping neural architecture underpinning these tasks (Buchsbaum et al., 2001; Hagoort and Indefrey, 2014; Indefrey and Levelt, 2004; Kempen

Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue, September 20–22, 2021, Potsdam / The Internet.

et al., 2012; MacKay, 1987; Menenti et al., 2011; Silbert et al., 2014). While the interference effects of language comprehension on speech production have received increasing attention during the last decades, the effects of speech planning on parallel comprehension remain comparatively understudied (Daliri and Max, 2016; Fargier and Laganaro, 2019; Levelt et al., 1991; Roelofs et al., 2007).

The present study focuses on semantic processing of the incoming speech in experimentally elicited question-answer sequences, exploiting the well-known effect of semantic illusions – the acceptance by a comprehender of a fallacious question or statement containing a word that makes the question or statement wrong but is semantically related to the correct word that was to be expected. A classic example used in the seminal study by Erickson and Mattson (1981) is the question *"How many animals of each kind did Moses take on the Ark?"*. In their study, Erickson and Mattson found that participants frequently failed to spot the illusion in the questions, even though they actually knew the correct facts, e.g., that it was not Moses but Noah who prepared himself for the great flood. The origin of the illusion is commonly explained by frugal, superficial semantic analysis and partial semantic feature-matching between the critical word and its sentence co-text – the greater the intersection of the semantics of the critical word with that of the correct word, the greater the chance that the illusion passes undetected (Song and Schwarz, 2008; Speckmann and Unkelbach, 2020; Van Oostendorp and De Mul, 1990).[1]

The conversational counterpart to the carefully constructed illusions used in experimental studies are word substitution errors, where an intended word accidentally gets replaced by an unintended word during speech planning. Such word substitutions are a very common kind of speech error (Meringer, 1908). Of the Fromkin Speech Error Corpus[2], a collection of 8673 spontaneous speech errors, 1083 errors (12.5%) are word substitutions, many of which replace the target word with a semantically related word that is generally of the same part of speech. In one example produced by Vicki Fromkin herself (and recorded by Robert Rodman), she produced "*Jack was going to build a YACHT on the 38th day*", instead of the intended and semantically related '*an ark*', when talking about a long period of rain in Oxford. As speech error corpora have traditionally mainly been used to draw inferences on the processes of speech production, the perception of these errors, including common detection rates, received far less attention (Bond, 1999).

In conversational situations, two different general language processing strategies are conceivable, predicting different effects on the detection rate of semantic illusions. As discussed above, in order to secure a timely response, speech planning needs to proceed swiftly already during the incoming turn. Therefore, delays due to capacity limits could lead to undesirably long turn-transition times that might communicate unintended meanings. Because planning in overlap is cognitively more demanding than planning the next turn in silence, next speakers might need to prioritize planning speed over comprehension accuracy after the point when response planning begins in order to secure smooth turn-transitions. We will call this the *turn-timing-prioritized hypothesis*. Equally theoretically conceivable is an alternative strategy that focuses on language comprehension in processing-heavy situations in turn taking. This strategy appears reasonable in view of the differences in the temporal dynamics of speech input and output processing. While the rate of speech input is defined by the speech rate of the current speaker, the rate of progress in speech output planning is under the control of the next speaker. That means that while the input, if not processed upon reception, is soon gone from perceptual memory, delays in response planning can be handled more flexibly by the next speaker, which is a potential incentive to prioritize the processing of incoming speech over speech planning in conversational situations, so that sufficient processing resources are available for comprehension. We will call this the *comprehension-prioritized hypothesis*.

The present study tests these competing hypotheses, using a quiz task with questions containing semantic illusions. While actual conversational situations are arguably way more complex than an experimental quiz situation, responding to questions is a very common action in social encounters. Using pre-recorded questions thus strikes a balance between exerting sufficient experimental

---

[1] The illusion effect is also boosted by additional phonological overlap between the target word and its replacement (Shafto and MacKay, 2000). Yet, the illusions presented in the present study were only semantically related to the target word, not phonologically.

[2] accessible at `https://www.mpi.nl/dbmpi/sedb/sperco_form4.pl`

control and tapping into the target speech production and comprehension processes. The presented questions differ in the point in time when response planning can begin. In one version of the question, response planning can begin already in overlap with the question, while in the other version, the answer to the question can only be known at the very end of the question. E.g., *Welche Tiere helfen dem Weihnachtsmann beim Verteilen der Neujahrsgeschenke?* ("What animals help Santa Claus to distribute New Year's presents?"; early planning) vs. *Der Weihnachtsmann verteilt die Neujahrsgeschenke mit der Hilfe von welchen Tieren?* ("Santa Claus distributes New Year's presents with the help of what animals?"; late planning). The two hypotheses outlined above make different predictions about the detection of semantic illusions in these two versions of the question. If comprehension is prioritized in dialogue situations, detection rates should not depend on whether the response is already being planned in overlap with the critical word or not. If fast turn-timing is prioritized on the other hand, detection rates should be lower in the early planning condition than in the late planning condition.

The point in time when response planning can start is confounded with the questions' sentence structure as well as with the position of the critical word within the question. To test the influence of these confounding factors, a control experiment was tested that did not require participants to actually answer the question. If a difference in detection rates between the two conditions in the main experiment is due to planning vs. not planning in overlap, this difference should not show in the control experiment, which does not involve response planning. If, on the other hand, differences in detection rate are due to differences in the form of the question itself, they should replicate in the control experiment.

2 Method

2.1 Participants

For the Main Experiment, 24 participants (age between 18 and 40 years) were recruited via Prolific and were paid to take part in the experiment online using their own computers. Another set of 24 participants was recruited for the Control Experiment.

2.2 Materials and Design

60 questions were composed, 30 of which were critical questions containing a semantic illusion. Of each question, two versions were composed, manipulating the point in time when the answer to the question can be known so that response planning can begin (planning: early / late; see example in (1)).

(1) (a) Early question: *Von welchem Tier wurde Rotkäppchen gefressen, | als sie ihre* ***Tante*** *besuchte?* ("What animal ate Little Red Riding Hood | when she visited her **aunt**?")

(b) Late question: *Als Rotkäppchen ihre* ***Tante*** *besuchte, wurde sie von welchem Tier gefressen |?* (When Little Red Riding Hood visited her **aunt**, what animal ate her |?)

In the early question in (1a), subjects can begin to plan their verbal response to the question already in the middle of the question (marked by the | symbol in (1)), whereas planning the response in the late question only becomes possible at the end of the question. The critical word containing the semantic illusion (printed in bold in (1)) is therefore located in a later part of the question in the early version, where response planning can be expected to be already ongoing, and in the earlier part of the question in the late version, where response planning cannot have started, yet.

Additionally, 30 filler questions were created. Filler questions were also composed in two versions, allowing for either early or late planning, but they did not contain any illusion (e.g., early planning: *Welcher Mann, der das Unternehmen Apple gründete, | war ein fortschrittlicher Boss?* (Which man, who founded the company Apple, | was a progressive boss?); late planning: *Welcher Mann, der ein fortschrittlicher Boss war, gründete das Unternehmen Apple |?* (Which man, who was a progressive boss, founded the company Apple |?)).

All questions were recorded in a female voice. They had a mean duration of 6 seconds (including 200 ms of initial silence), with a standard deviation of 1.18 seconds. While early and late questions did not differ greatly in length (early questions: mean (sd) = 5.84 s (1.0 s); late questions: mean (sd) = 6.16 s (1.33 s)), filler questions were slightly longer than critical questions (critical: mean (sd) = 5.48 s (0.88 s); fillers: mean (sd) = 6.53 s (1.21 s)).

Two balanced experimental lists were composed, so that questions were presented in only one of the conditions to each subject. The order of presentation of items within the list was random for every participant.

2.3 Procedure

2.3.1 Main Experiment

In the Main Experiment, participants were instructed to use their headphones to listen to the questions during the quiz part of the experiment and to respond verbally to the questions as fast and accurately as possible. They were made aware that not all the questions they would hear during the experiment were correct, using the example question *“Who assassinated US President Clinton?”* and the correct answer *“Nobody. Clinton has not been assassinated.”*. They were further instructed to pay special attention to the fact that some questions would be incorrect and to answer them accurately. Lastly, they were instructed to say *I don’t know* in response to any question that they did not know the answer to.

Each trial began with a fixation cross in the center of the screen for one second, followed by the auditory presentation of the question, which participants had to answer verbally. Participants’ responses were recorded using their PC microphones and participants were instructed to press the space bar after they gave their response. One second after they pressed the space bar, the next trial started. Before the experiment, participants did four practice trials to get to know the procedure.

The quiz part was followed by a post-test questionnaire testing participants’ knowledge of the correct versions of the 30 critical questions used in the quiz. Participants read questions asking about the critical information in each of the critical questions of the quiz (e.g., *“Wen besuchte Rotkäppchen, als sie vom Wolf gefressen wurde?”* (Whom did Little Red Riding Hood visit when she was eaten by the wolf?)) and typed their response in a text box. Questions appeared one at a time, replacing each other each time participants pressed ‘enter’ to confirm their response. The whole experiment took about 20 minutes.

2.3.2 Control Experiment

In the first part of the Control Experiment, participants were instructed to not respond to the questions they heard, but to judge whether the question was correct or erroneous and indicate their choice by clicking on the respective radio button. As a third alternative, participants could indicate that they did not know whether the question was correct or not. One second after participants confirmed their response by pressing the space bar, the next trial started. As well as in the Main Experiment, participants were made aware that not all the questions they would hear during the experiment were correct, using the same example. Participants did four practice trials prior to the experiment. The post-test questionnaire was the same as in the Main Experiment.

3 Results

3.1 Response Latencies

Response latencies were annotated manually in Audacity with respect to question offset and response onset. Of the total of 1440 trial recordings, 5 did not contain any response and were thus discarded. The remaining responses had a mean latency of 1650 ms (sd = 1220 ms; see **Figure 1**). Response latencies were fitted with a Bayesian mixed effects regression model with the R package *brms* (Bürkner, 2017; R Core Team, 2021), with Condition (early planning / late planning) and Type (critical / filler) plus their interaction as fixed effects and as random effects by subject and by item (see Appendix). Both factors were dummy coded, with early planning and critical condition as reference levels. Additionally, the centred duration of the questions in seconds was added as a control variable to the fixed effects structure of the model, since turn duration has been shown to affect turn transition times (Barthel et al., 2016; Magyari, 2015; Roberts et al., 2015), an effect that replicated here (β = -171.54 ms, CI = [-277.65 ms; -72.27 ms]). The prior for the Intercept was set to be normally distributed, with a mean of 1600 ms and a standard deviation of 2500 ms. Priors for the coefficients were vaguely informative as they were set to be normally distributed, with a mean of 0 and a standard deviation of 500 ms.

The model detected both a decisive main effect of Condition (β = -282.16 ms; CI = [-398.36 ms; -166.92 ms], BF_{10} = 5999)[3] as well as a strong main effect of Type (β = -211.55 ms; CI = [-422.47 ms; 6.98 ms]; BF_{10} = 16.86), indicating that early questions were responded to faster than late questions and that filler questions were responded to faster

---

[3] For a guideline to the interpretation of Bayes factors, see Andraszewicz et al. (2015).

than critical questions. However, the model also attested a decisive interaction effect of Condition × Type (β = 496.19 ms; CI = [251.73 ms; 744.07 ms]; BF$_{10}$ = 5999). To investigate the origin of this interaction effect, hypothesis specific tests were conducted using the *hypothesis* function built into *brms*, which revealed that Condition had a decisive effect on response latencies only in filler trials (β = 530.26 ms; CI = [368.85 ms; 690.32 ms]; BF$_{10}$ $>$ 6000), but no effect in critical trials (β = 34.07 ms; CI = [-116.85 ms; 184.98 ms], BF$_{10}$ = 1.88), indicating that responses were given faster in early than in late questions in filler trials but equally fast in critical trials. However, an additional model that was run only on the subset of critical trials in which participants accepted the illusion (see Appendix) revealed that in these trials Condition did have an effect on response latencies (β = 126.01 ms; CI = [-97.65 ms; 346.11 ms], BF$_{10}$ = 4.88), indicating that response latencies in late questions were longer than in early questions when participants gave the expected response. While going in the same direction, this effect was smaller and statistically weaker than in filler trials, possibly because critical questions were slightly more difficult than filler questions and because this test relied on a lower number of observations (N_{early}= 118; N_{late} = 90).

3.2 Semantic Illusions

For an analysis of the proportions of semantic illusions that were detected or not, only critical trials were analysed. 25 trials with unintelligible or nonsense responses, 143 trials in which participants responded that they did not know the answer to the question, and 111 trials where participants revealed in the post-test questionnaire that they did not have the necessary factual knowledge to spot that the original quiz question contained an illusion were discarded, leaving 439 responses for analyses. Responses were coded as accepting the illusion when the expected answer was given, or as rejecting the illusion when the illusion was spotted. Descriptively, illusions were accepted in 84 out of 226 trials (37%) in the early planning condition and in 58 out of 213 trials (27%) in the late planning condition. The probability of the illusion being accepted was fitted with a Bayesian mixed effects regression model (family = bernoulli) with Condition as a dummy-coded fixed effect and as a random effect by subject and by item, with 'rejected' as the reference level (see Appendix). Condition was found to have a strong effect on the proportion of accepted illusions (β = -0.49; CI = [-0.96; -0.03]; BF$_{10}$ = 22.67; with the best estimate for the Intercept at β = -0.67), showing strong evidence for the probability of illusions to be accepted to be higher in the early planning condition than in the late planning condition (see **Figure 2**).

For the Control Experiment, also only critical trials were analysed.[4] In 112 trials, participants indicated that they did not know whether the question was correct or not, and in 94 of the remaining trials, the post-test revealed that participants did not have the necessary factual knowledge to spot the original illusion, leaving 484 trials for analyses. Descriptively, illusions passed in 24 out of 239 trials (1.0%) in the early condition and in 22 out of 245 trials (0.9%) in the late condition. A model parallel to the Main Experiment model was fitted (see Appendix), which revealed that Condition had a very weak effect on the probability of the illusion being accepted (β = -0.4; CI = [-1.33; 0.49]; BF$_{10}$ = 3.77, with the best estimate for the intercept at β = -2.75), showing merely anecdotal evidence for the probability of illusions to differ between the early and late planning conditions.

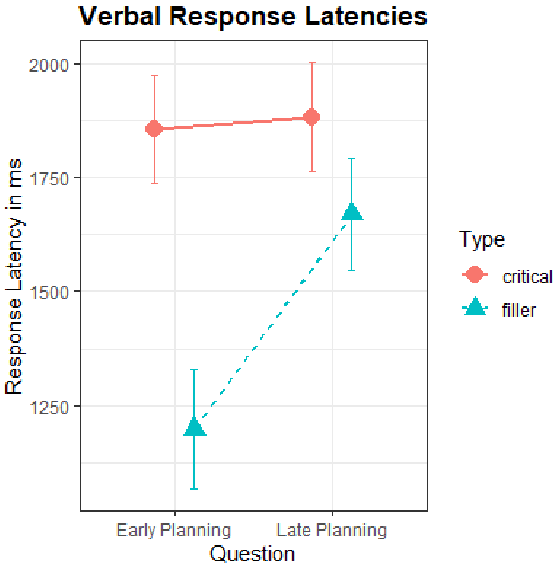

Figure 1: Mean response latencies by condition in critical questions (containing a semantic illusion) and filler questions (not containing an illusion). Error bars indicate 95% confidence intervals.

---

[4] Data from one participant were discarded because task instructions were ignored in the post-test questionnaire and most responses in the main part were 'I don't know' responses.

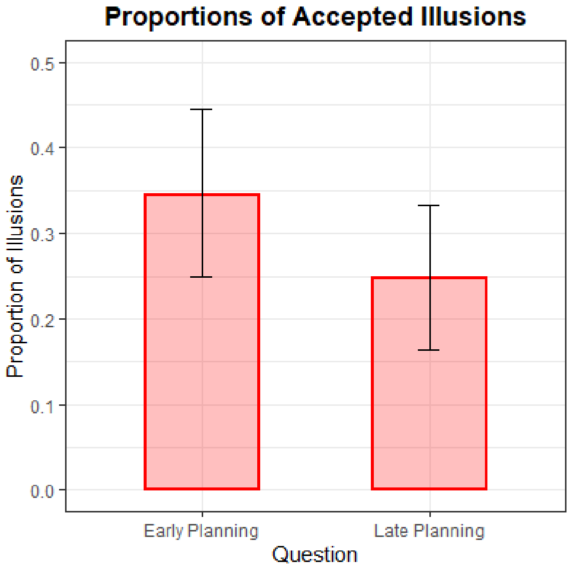

Figure 2: Fitted proportions of illusions that were accepted, i.e., not spotted, by participants in the Main Experiment. Error bars indicate 75% credible intervals. See section 3.2 for model description.

4 Discussion

Conversation is a well-practiced but cognitively demanding dual-task situation, where processes of language comprehension and speech planning can interfere with one another, either due to cross talk between the representations that are relevant for each of the tasks or due to limited resources that need to be shared between the tasks. This study tested two competing hypotheses about the default allocation of processing capacities in moments of increased cognitive load in a dialogic task. In conversational turn taking, cognitive load is especially high in next speakers when they are concurrently planning their upcoming turn and listening to the incoming turn. The *comprehension-prioritized hypothesis* states that when processing capacities are temporarily limited, processing of the incoming speech would be prioritized over response planning because the rate of incoming information to be processed is defined by the speech rate of the incoming turn and is thus not under the control of the listener/next speaker. If the incoming speech is not processed thoroughly at the rate it is coming in, part of the signal would be lost, which might be an undesirable characteristic of any adapted processing strategy. The *turn-timing-prioritized hypothesis*, on the other hand, states that well-timed responses are central in order to convey the intended messages in a conversational situation. And in order to be able to deliver the next turn quickly in response to the incoming turn, response planning needs to start and progress rapidly in overlap with the incoming turn. When this dual-task situation leads to increased processing load, response planning would be prioritized so as to not jeopardize seamless turn-timing. These two hypotheses were tested in a question-answer paradigm with questions containing semantic illusions, such as “How many animals of each kind did *Moses* take on the Ark?”

Previous studies showed that next speakers readily put themselves in the described dual-task situation when they are in conversation, starting to plan their next turn as soon as they can anticipate the message of the incoming turn, even though planning in overlap leads to increased processing load at turn-transitions (Barthel, 2020; Barthel and Sauppe, 2019; Bögels, 2020; Levinson and Torreira, 2015). The response latency results in the present study replicate these previous findings. Here, participants verbally responded to quiz questions. For half of the questions, the response could be planned already in the middle of the question, for the other half of the questions, the response could only be planned at the end of the question. Questions whose responses could be planned already in overlap with the question were responded to faster than questions whose response could only be known at the question’s end, showing that participants started to plan their response already in overlap with the incoming question when this was possible, and thereby achieved shorter response latencies. This attested effect of question format on response latencies needs to be qualified, however. While the effect was observed to be strong in questions that did not contain a semantic illusion, the effect was not attested in questions that did contain a semantic illusion. However, post-hoc analyses did reveal the effect in questions containing a semantic illusion, but only when the illusion was not detected, i.e., when the question was answered as would be expected without the illusion. This pattern of results indicates that also in questions containing an illusion, participants started to plan their response as early as possible but had to abandon the planning process when they detected a mismatch between their anticipation of the upcoming input and the actual continuation of the question. In these cases, participants had to begin planning from scratch, this time to reject the illusion, and therefore did not show any observable gain from early response planning.

Overall, about one third of the questions containing an illusion were accepted and answered as would be expected without the illusion. This acceptance rate was affected by whether planning in overlap with the illusion word was possible or not. Participants failed to detect the illusion more often when they were planning in overlap than when they were not concurrently planning, showing that speech planning in overlap is detrimental to semantic input processing. While the early and late planning questions differed in their format and in the position of the critical word containing the illusion, these differences were not driving the effect. This possibility can be excluded on the basis of the results of the control experiment, in which response planning was not necessary. When participants rated the questions for correctness instead of answering them verbally, the position effect disappeared.[5] This pattern of results supports the *turn-timing-prioritized hypothesis*, which predicted that in phases of high processing load in dialogue situations, dynamic progress in response planning would be prioritized over deep processing of the input, so that the response is ready for articulation shortly after the incoming turn comes to an end. The results do not support the *comprehension-prioritized hypothesis*, as comprehension was found to be less accurate when planning was executed in overlap with the question. Instead, the results indicate that in these phases, participants processed the input more shallowly and based their response planning on their anticipations of the question continuations (Ferreira and Patson, 2007; Ferreira et al., 2002; Song and Schwarz, 2008; Van Oostendorp and De Mul, 1990; van Oostendorp and Kok, 1990). Shallow input processing can be assumed to occur most prominently in situations when processing load increases, which are to be most frequent before turn-transitions (Barthel and Sauppe, 2019).

Prioritizing response planning can indeed be argued to be an efficient strategy in dialogue, even if it might be at the expense of comprehension accuracy. Listeners in conversation have been found to generate predictions about the incoming turn in order to be able to start planning their response early on the basis of their predictions (Corps et al., 2018; Magyari et al., 2014; Gisladottir et al., 2015, 2018). This early planning enables them to take their next turn quickly after the incoming turn ends. The fact that most turn-transitions are fast makes conversation efficient with respect to the utilisation of the available time, and on top of that, it is the basis for turn-timing to be interpreted as meaningful when transitions are slow (Henetz, 2017; Roberts and Francis, 2013; Roberts et al., 2011). In the majority of cases, predictions about the message of the end of the incoming turn are probably correct, as turn endings are often predictable (Magyari and de Ruiter, 2012). In these cases, relying on the predictions is certainly an efficient strategy. In cases where the upcoming input does not match the predictions, two reasons for the mismatch come to mind. Either the input was ‘wrong’, i.e., not as intended by the current speaker, e.g., when they erroneously replaced *Noah* with *Moses* (Fromkin, 1971; Meringer, 1908; Levelt, 1989), in which case the prediction was actually ‘right’ and the conversation can continue smoothly even if the error passes unnoticed. Or the prediction was wrong, which would lead to misunderstanding if the mismatch passes unnoticed. These latter, problematic cases can be considered to be rare enough in natural conversation for the turn taking system to be efficient, and if they do arise, they are commonly detected and dealt with by the interactants immediately in the next turn with the help of repair sequences (Dingemanse et al., 2015; Schegloff, 1992). Prediction and planning strategies can be argued to be readily built upon this safety-net that comes with conversational repair, as repair mechanisms are general purpose tools that are used for any form of misunderstanding, e.g., in problems in acoustic understanding or reference matching, and are not specific to fixing the consequences of prediction errors.

It remains difficult to judge the relevance of prediction for the probability of an illusion to pass unnoticed. Given that interlocutors predict the end of an incoming turn in order to prepare their response (e.g., Corps et al., 2019), the difference in illusion rates could be due to a higher predictability of the target word in the early planning condition than in the late planning condition. This line of thought would assume that comprehension is more shallow when predicting the input. This is indeed

---

[5] While response planning was not prohibited with certainty in the control experiment, the absence of an effect of question type indicates that subjects did not engage in response preparation but rather focused on comprehending the question in the control task.

possible and would underpin the assumption that input prediction as a conversational strategy is efficient because more resources are available for planning earlier before the end of the incoming turn. It is thus conceivable that a higher rate of illusions might not be due to the planning itself, but due to prediction of the target word or concept that was replaced by the illusion word. The absence of an effect of question type in the control experiment (where no response to the question was given) could be argued to refute this idea, since the questions were the same as in the main experiment and the target words were therefore equally predictable. However, due to the different tasks, not only response planning but also input prediction might have been reduced in the control experiment, which could have eliminated the effect of question type. In the context of the tasks of the present study, the two sides of the medal might be too closely coupled to tease apart their contributions. Arguably though, subjects might have engaged less in input prediction in the control experiment, *because* there was no need for response preparation. Consequently, comprehension of the input suffered when planning the next turn as compared to when not planning the next turn, possibly mediated by input prediction.

Contrary to the assumption that language input is processed more shallowly when predictions are maintained, it would also be sensible to argue that word substitutions should be *more* obvious when a prediction to hear a different word has already been generated. In this line of thought, not shallow comprehension but rather processing ease at the time of encounter should follow from prediction, which is corroborated by findings that comprehension is less effortful in high predictability sentences (e.g., Obleser and Kotz, 2011). In that case, a higher rate of illusions would be indicative of shallow comprehension due to concurrent response planning rather than due to prediction. Under these considerations, it should be easier to detect a word substitution when the input is predictable, so that lower illusion rates would be expected in the questions with word substitutions at their ends (i.e., in the early planning questions). The fact that the opposite pattern of results was found thus speaks for the interpretation that concurrent response planning rather than prediction was responsible for the differential illusion effect. Future research will be needed to conclusively disentangle the relative contributions of these two confounded factors.

One final side-note needs to be added about the comparability of the present study with previous studies investigating semantic illusions. Seeing that planning in overlap increases the rate of semantic illusions brings up the question what relevance this effect might have had in previous studies. This question is difficult to answer conclusively, firstly because the point in time when response planning was possible was not included as a control variable in previous studies, and secondly because results on that question are not directly comparable between studies, since, to the best of our knowledge, all previous studies presented the critical questions in print, whereas the present study is the first to illustrate the occurrence of semantic illusions with auditorily presented speech. What can be said with certainty, however, is that semantic illusions do not depend on speech planning in overlap. In the classic experiment by Erickson and Mattson (1981), two of the four critical questions contained the illusion word in a position after the question can be known and two contained the illusion word before the question can be known. In their study, both types of questions were reported to elicit semantic illusions. Moreover, also the present study found evidence that concurrent response planning is certainly not a prerequisite for semantic illusions to arise. Still, semantic processing of the input was found to be less effective during speech planning, and future studies should take this factor into account, either by controlling for the position of the illusion with respect to the point where planning can begin, by balancing their materials, and/or by statistically controlling for the influence of the factor post-hoc.

5 Conclusion

Semantic illusions have been found to be stronger when speech planning is executed while comprehending the illusory input than without concurrent speech planning. Hence, semantic processing of language input in dialogic situations can be assumed to be more shallow during speech planning, even when the planned content is contingent upon the content of the incoming speech. The effect of concurrent planning on semantic processing is possibly due to limited processing resources operating on related linguistic representations, so that next speakers need to strike an efficient balance between comprehension and planning before turn transitions. Nonetheless, as it is a prerequisite for seamless turn-timing, speech planning in overlap with comprehension is a communicatively effective strategy, as it is a cornerstone of the turn taking system that forestalls abundant long gaps and allows turn-timing to be interpreted as meaningful by interlocutors. In sum, planning the next turn in overlap with the incoming turn does not seem to be efficient from a processing perspective, as comprehension accuracy suffers from concurrent speech planning. Still, prioritizing planning the next turn under high processing load before turn transitions could be a very effective strategy for communication, and the present experiment provides evidence that planning is prioritized over accurate comprehension in periods when these processes compete for cognitive resources.

Acknowledgments

This study was funded by the Berlin University Alliance and conducted within the X-Student Research Group *Language Comprehension in Dialogue*, led by the author. This paper has been improved by three valuable anonymous reviews. The author thanks Hannah Ida Hullmeine for recording the experimental stimuli.

References

Abdel Rahman, R. and Melinger, A. (2019). Semantic processing during language production: an update of the swinging lexical network. *Language, Cognition and Neuroscience*, 34(9):1176–1192.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., and Wagenmakers, E.-J. (2015). An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*, 41(2):521–543.

Barthel, M. (2020). *Speech Planning in Dialogue - Psycholinguistic Studies of the Timing of Turn Taking*. PhD Thesis, Radboud University Nijmegen, Nijmegen.

Barthel, M. and Levinson, S. C. (2020). Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*, 35(9):1183–1202.

Barthel, M., Meyer, A. S., and Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final “Go-Signals”. *Frontiers in Psychology*, 8:393.

Barthel, M. and Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, 43(7):e12768.

Barthel, M., Sauppe, S., Levinson, S. C., and Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm. *Frontiers in Psychology*, 7:1858.

Boiteau, T. W., Malone, P. S., Peters, S. A., and Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, 143(1):295–311.

Bond, Z. S. (1999). *Slips of the ear: errors in the perception of casual conversation*. Academic Press, San Diego, Calif.

Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25(5):663–678.

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, 203:104347.

Bögels, S., Magyari, L., and Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5(12881):1–11.

Bürki, A., Elbuy, S., Madec, S., and Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, 114:104125.

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1).

Corps, R. E., Crossley, A., Gambi, C., and Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175:77–95.

Corps, R. E., Pickering, M. J., and Gambi, C. (2019). Predicting turn-ends in discourse context. *Language, Cognition and Neuroscience*, 34(5):615–627.

Daliri, A. and Max, L. (2016). Modulation of Auditory Responses to Speech vs. Nonspeech Stimuli during Speech Movement Planning. *Frontiers in Human Neuroscience*, 10.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., and Enfield, N. J. (2015). Universal Principles in the Repair of Communication Problems. *PLOS ONE*, 10(9):e0136100.

Erickson, T. D. and Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5):540–551.

Fargier, R. and Laganaro, M. (2016). Neurophysiological Modulations of Non-Verbal and Verbal Dual-Tasks Interference during Word Planning. *PLOS ONE*, 11(12):e0168358.

Fargier, R. and Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific Reports*, 9(1):5375.

Ferreira, F., Bailey, K. G., and Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, 11(1):11–15.

Ferreira, F. and Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, 1(1-2):71–83.

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, 47(1):28.

Gisladottir, R. S., Bögels, S., and Levinson, S. C. (2018). Oscillatory Brain Responses Reflect Anticipation during Comprehension of Speech Acts in Spoken Dialog. *Frontiers in Human Neuroscience*, 12.

Gisladottir, R. S., Chwilla, D. J., and Levinson, S. C. (2015). Conversation Electrified: ERP Correlates of Speech Act Recognition in Underspecified Utterances. *PLOS ONE*, 10(3):1–24.

Hagoort, P. and Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, 37(1):347–362.

He, J., Meyer, A. S., and Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, pages 1–23.

Heldner, M. and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Henetz, T. (2017). *Don't hesitate! The length of inter-turn gaps influences observers' interactional attributions*. PhD Thesis, Stanford University, Stanford.

Indefrey, P. and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–144.

Jescheniak, J. D., Matushanskaya, A., Mädebach, A., and Müller, M. M. (2014). Semantic interference from distractor pictures in single-picture naming: evidence for competitive lexical selection. *Psychonomic Bulletin & Review*, 21(5):1294–1300.

Kempen, G., Olsthoorn, N., and Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3):345–380.

Kendrick, K. H. and Torreira, F. (2014). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, 52(4):1–35.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1):143–162.

La Heij, W., Dirkx, J., and Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, 81(4):511–525.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press, London.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., and Pechmann, T. (1991). The Time Course of Lexical Access in Speech Production: A Study of Picture Naming. *Psychological Review*, 98(1):122–142.

Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6(731):10–26.

MacKay, D. G. (1987). *The Organization of Perception and Action*. Cognitive Science Series. Springer New York, New York, NY.

Magyari, L. (2015). *Timing Turns in Conversation: A temporal preparation account*. PhD Thesis, Radboud University Nijmegen, Niljmegen.

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early Anticipation Lies behind the Speed of Response in Conversation. *Journal of Cognitive Neuroscience*, 26(11):2530–2539.

Magyari, L. and de Ruiter, J. P. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, 3(376):1–9.

Menenti, L., Gierhan, S. M. E., Segaert, K., and Hagoort, P. (2011). Shared Language: Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI. *Psychological Science*, 22(9):1173–1182.

Meringer, R. (1908). *Aus dem Leben der Sprache*. B. Behr's, Berlin.

Meyer, A. S. (1996). Lexical Access in Phrase and Sentence Production: Results from Picture–Word Interference Experiments. *Journal of Memory and Language*, 35(4):477–496.

Obleser, J. and Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, 55(2):713–723.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Roberts, F. and Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, 133(6):EL471–EL477.

Roberts, F., Margutti, P., and Takano, S. (2011). Judgments Concerning the Valence of Inter-Turn Silence Across Speakers of American English, Italian, and Japanese. *Discourse Processes*, 48(5):331–354.

Roberts, S. G., Torreira, F., and Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in Psychology*, 6.

Roelofs, A., Özdemir, R., and Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5):900–913.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735.

Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, 97(5):1295–1345.

Schriefers, H., Meyer, A. S., and Levelt, W. (1990). Exploring the Time Course of Lexical Access in Language Production: Picture-Word Interference Studies. *Journal of Memory and Language*, 29:86–102.

Shafto, M. and MacKay, D. G. (2000). The Moses, mega-Moses, and Armstrong illusions: integrating language comprehension and semantic memory. *Psychological Science*, 11(5):372–378.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, 111(43):E4687–E4696.

Song, H. and Schwarz, N. (2008). Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion. *Social Cognition*, 26(6):791–799.

Speckmann, F. and Unkelbach, C. (2020). Moses, money, and multiple-choice: The Moses illusion in a multiple-choice format with high incentives. *Memory & Cognition*, 49(4):843–862.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592.

Van Oostendorp, H. and De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, 74(1):35–46.

van Oostendorp, H. and Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, 5(2):105–113.

A Appendix - Bayesian regression models

Group-Level Effects:				
~itemID (Number of levels: 60)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	439.04	61.03	326.68	567
sd(condition=late)	248.78	97.22	42.79	427.24
cor(Intercept,condition=late)	-0.54	0.27	-0.91	0.11
~subjectID (Number of levels: 24)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	551.03	93.93	397.13	764.51
sd(condition=late)	136.81	87.58	5.97	323.47
cor(Intercept,condition=late)	-0.04	0.44	-0.84	0.86
Population-Level Effects:	Estimate	Est.Error	l-95% CrI	u-95% CrI
Intercept	1743.50	151.24	1450.64	2034.21
condition=late	34.07	92.59	-148.27	215.21
type=filler	-459.65	146.55	-745.62	-172.12
questionDuration_centered	-171.54	52.52	-277.65	-72.27
condition=late:type=filler	496.19	125.70	251.73	744.07
Residual Error:	Estimate	Est.Error	l-95% CrI	u-95% CrI
sigma	998.81	19.91	961.37	1038.92

Table 1: Model output of main reaction times model. Family = gaussian. Link = identity. Formula = responseLatency_inms˜1 + condition * type + questionDuration_c + (1 + condition | subjectID) + (1 + condition | itemID). Number of observations = 1435. Samples = 3 chains, each with iter = 3000; warmup = 1000; thin = 1. Factor reference levels: condition = early; type = critical.

Group-Level Effects:				
∼itemID (Number of levels: 30)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	461.46	124.87	224.50	719.84
∼subjectID (Number of levels: 24)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	414.23	106.40	230.71	650.00
Population-Level Effects:	Estimate	Est.Error	l-95% CrI	u-95% CrI
Intercept	1528.37	161.52	1208.68	1844.13
condition=late	126.01	134.40	-141.33	388.87
questionDuration_centered	-136.46	131.12	-397.47	123.19
Residual Error:	Estimate	Est.Error	l-95% CrI	u-95% CrI
sigma	912.35	53.02	815.92	1022.51

Table 2: Model output of reaction times model on subset of accepted illusions. Family = gaussian. Link = identity. Formula = responseLatency_inms ˜ 1 + condition + questionDuration_c + (1 | subjectID) + (1 | itemID). Number of observations = 208. Samples = 3 chains, each with iter = 3000; warmup = 1000; thin = 1. Factor reference level: condition = early

Group-Level Effects:				
∼itemID (Number of levels: 29)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	0.81	0.24	0.39	1.33
sd(condition=late)	0.37	0.29	0.02	1.06
cor(Intercept,condition=late)	-0.11	0.55	-0.95	0.92
∼subjectID (Number of levels: 24)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	1.48	0.32	0.95	2.19
sd(condition=late)	0.31	0.25	0.01	0.93
cor(Intercept,condition=late)	0.01	0.57	-0.94	0.95
Population-Level Effects:	Estimate	Est.Error	l-95% CrI	u-95% CrI
Intercept	-0.67	0.39	-1.46	0.09
condition=late	-0.49	0.28	-1.06	0.06

Table 3: Model output of model on the rate accepted illusions in the Main Experiment. Family = bernoulli. Link = logit. Formula = responseLatency˜condition + (1 + condition | subjectID) + (1 + condition | itemID). Number of observations = 439. Samples = 3 chains, each with iter = 6000; warmup = 2000; thin = 1. Factor reference level: condition = early

Group-Level Effects:				
∼itemID (Number of levels: 29)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	1.27	0.42	0.57	2.20
sd(condition=late)	0.37	0.29	0.02	0.52
cor(Intercept,condition=late)	-0.16	0.56	-0.97	0.92
∼subjectID (Number of levels: 24)	Estimate	Est.Error	l-95% CrI	u-95% CrI
sd(Intercept)	0.55	0.36	0.02	1.37
sd(condition=late)	1.00	0.56	0.07	2.20
cor(Intercept,condition=late)	-0.10	0.55	-0.94	0.91
Population-Level Effects:	Estimate	Est.Error	l-95% CrI	u-95% CrI
Intercept	-2.75	0.45	-3.77	-1.98
condition=late	-0.40	0.56	-1.55	0.66

Table 4: Model output of model on the rate accepted illusions in the Control Experiment. Family = bernoulli. Link = logit. Formula = responseLatency˜condition + (1 + condition | subjectID) + (1 + condition | itemID). Number of observations = 439. Samples = 3 chains, each with iter = 6000; warmup = 2000; thin = 1. Factor reference level: condition = early