Running head: PRESUPPOSITION PROJECTION IN ONLINE PROCESSING

Presupposition Projection in Online Processing

Florian Schwarz and Sonja Tiemann

Department of Linguistics and IRCS, University of Pennsylvania and

Eberhard Karls Universität Tübingen

Abstract

A central aspect of language comprehension is that hearers integrate incoming linguistic content both with the rest of the current sentence and the larger discourse context. Presuppositions crucially interact with both intra- and inter-sentential context in intricate ways, which makes their study especially useful in this regard. We present a series of experiments investigating the time-course of interpreting presuppositions in online comprehension and the impact that so-called presupposition projection in cases where presuppositions appear in embedded environments has on this. We find immediate delays in eye tracking reading times when the presupposition of German *wieder* ('again') is not supported by the context, but only for unembedded occurrences of *wieder*. Further evidence from a rating experiment and a stops-making-sense study supports our interpretation of this result to the effect that global presuppositions of embedded presupposition triggers are not immediately available in processing. A second reading time experiment explores the effects of embedding further by providing presuppositional support in different locations in contexts with a more complex structure involving conditionals. We find longer reading times when the support is more distant, measured in terms of the number of projection steps posited by Discourse Representation Theory (DRT). Altogether, the results lend support to the notion that presupposition projection is a cognitively effortful process, and are thus consistent with theoretical accounts that reflect this in terms of the complexity of the representations involved in the different types of contexts. On the more general level of discourse processing models, these results suggest that there is even more structure relevant to cognitive processes at a level between the surface representation and the purely semantic level (e.g., the commonly assumed level of a text-base) than previously assumed.

Presupposition Projection in Online Processing

Introduction

One important question in the study of natural language meaning is how comprehenders piece together the overall meaning that they take a given utterance to convey based on the various components of meaning that can be differentiated theoretically in terms of their properties. A substantial body of work growing out of seminal proposals made within philosophy of language and technically more detailed accounts within formal semantics in linguistics has argued, in particular, that we need to distinguish between levels such as the literal, truth-conditional meaning of an expression, presuppositions, and conversational implicatures (as well as conventional implicatures, which some authors see as on par with presuppositions). From a theoretical perspective, many interesting questions arise as to what exactly the classes of meanings to be distinguished are, how these various ingredients of the overall conveyed meaning arise, how they interact, and how they behave in different linguistic environments.

From the perspective of online processing, equally interesting questions concern the when and how of these aspects of meaning being computed in comprehension. One question, which has already received substantial attention in the recent experimental literature on implicatures, is how the different types of meaning relate to one another in online processing. Another question is how aspects of meaning are related to and integrated with both other information in the same sentence as well as the overall discourse context. The psychological literature on discourse processing includes various proposals for how situation models, seen as language-independent representations of the expressed content, are constructed from linguistic input. One of the central upshots from this literature is that there are cognitively real levels of representation that are neither identical to the linguistic surface form nor as independent from the linguistic structure of the surface form as situation models, as they maintain, say, a structured representation of propositions in the form of predicates and their arguments. Understanding precisely how

information based on linguistic input is structured at this intermediate level is central for a cognitive theory of language comprehension in context.

The experiments reported in this paper focus on the online processing of presuppositions, by looking at sentences such as the one in (1), where *again* introduces the presupposition that Tina had been ice-skating before (see below for more detailed discussion of the properties of presuppositions).

(1) On Saturday, Tina went ice-skating again.

Presuppositions are of particular interest in light of the general questions raised above. First, while presuppositions are a formally well-studied aspect of meaning, little is known about their online processing, especially in comparison with other aspects of meaning, such as implicatures and asserted content. Even more importantly, presuppositional phenomena exist, as it were, right at the interface between the intra-sentential computation of meaning (typically seen as the realm of formal semantics in theoretical terms) and the broader integration of sentential meaning into the discourse context (part of a theory of discourse, or pragmatics more generally). The core phenomenon of interest in this regard is that of presupposition projection, which arises when presuppositional expressions are introduced in embedded contexts and involves - loosely speaking - a mismatch between the syntactic location of the expression introducing a presupposition and the level at which the presupposition is interpreted (more details below). Our experiments investigate the time-course of presupposed content in both embedded and unembedded environments.

We begin by introducing the basic theoretical and experimental background for our later discussion. We then discuss two sets of experiments, both of which are concerned with German *wieder* ('again') in unembedded and embedded environments. In the first experiment, we look at eye tracking reading times in contexts that either do or do not support the presupposition of *wieder*, which in turn is introduced either in an unembedded position or in the scope of negation. We find immediate slow-downs in eye tracking reading time measures when the context does not match an unembedded presupposition, indicating that in these cases, the presupposition is

assessed relative to information present in the context right away. Intriguingly, such effects are absent when the presupposition trigger is embedded under negation. Two follow up studies, one involving a simple acceptability rating task and one using a 'stops making sense' task, are presented as well to narrow down the interpretation of the reading time results. Taken together, the experimental results suggest that while unembedded presuppositions are evaluated immediately relative to the context, there is a delay in evaluating those embedded under negation. In the first part of the general discussion, we discuss these findings relative to two prominent semantic theories of presupposition - DRT and dynamic semantics -, and argue them to fit more naturally with the one that assumes a structured representation of the discourse context. We also reflect on the implications of this interpretation of the results for more processing-oriented models of discourse comprehension.

The interpretation of the first set of studies yields straightforward predictions for more complex embedding environments. The second set of studies introduces a larger range of embeddings, both in terms of the embedding expressions involved and the depth of embedding, and varies the level at which the presuppositions is resolved. We find significant increases in reading times based on the hierarchical distance, as measured with respect to structures assumed by Discourse Representation Theory (DRT), between the location where the presupposition is introduced and where it is interpreted. The upshot of our results is that the evaluation of presupposed content in online processing takes place relative to a structured representation of information introduced in the discourse, as proposed by DRT, rather than a purely semantic representation that is unstructured. At the more general level of a cognitively real theory of language comprehension in context, this suggests that an intermediate level between surface form (e.g., viewed as the actual linguistic tokens expressed in a given structure) and semantic content (e.g., in the form of situation models or other abstract and general semantic representations, such as sets of possible worlds) has to be structured in more complex ways than previously assumed. We conclude with a general discussion of the emerging picture of presupposition processing and its implications for accounts of discourse comprehension more generally, as well as a brief

consideration of other theories of presuppositions and related experimental results.

Presuppositions: Theoretical and Experimental Background

The Classical Picture - Basic Properties of Presuppositions

The probably most influential tradition in presupposition theory, following work by Karttunen and Stalnaker (Karttunen, 1973; Stalnaker, 1973, 1974, 1978) sees presupposition triggers as introducing constraints on the contexts in which a given sentence can be uttered felicitously. In particular, they express information that has to be entailed by what counts as established in the discourse. While on some level, this clearly suggests a pragmatic view of presuppositions in that they place restrictions on the use of linguistic expressions, different strains of theories in this tradition vary with respect to whether they see presuppositions as part of what is conventionally encoded in the lexical entries of certain expressions or whether they instead assume that presuppositions are derived in an entirely pragmatic fashion, based on general reasoning much like that involved in generating conversational implicatures. Historically speaking, there originally was a trajectory from pragmatic to semantic accounts as precise formal implementations of presuppositional phenomena where developed in linguistics. In particular, the work of Heim (1983) and Kamp (1981), to be discussed in more detail later, introduced a shift that incorporated contextual aspects of interpretation into the semantic realm. However, the recent surge in new variations of presupposition theories has come with a revival of more purely pragmatic accounts (Simons, 2001; Abusch, 2010; Schlenker, 2008b, 2009; Abrusán, 2011).[1]

Independent of the debate about the semantic vs. pragmatic status of presuppositions, the probably most central characteristic of presuppositions is that they are commonly interpreted outside of the scope of certain operators, even if the expression introducing the presupposition

[1]Note that some of these types of accounts, perhaps most prominently Abusch's, assume that there are different types of presupposition triggers, and may not be intended to apply to triggers like *again*. But since this restriction does not hold for all of these types of accounts, it's fair to consider the class as a whole in light of the present experimental investigation.

appears in the relevant operator's scope. For example, all of the following variations of the simple sentence in (2) have the same presupposition, introduced by *again*, that Tina had been ice-skating before:

(2) Tina went ice-skating again today.

a. It is just not true that Tina went ice-skating again today.

b. If the weather was nice, Tina went ice-skating again today.

c. Did Tina go ice-skating again today?

Unlike the plain declarative version in (2), the asserted content that Tina went ice-skating today is no longer conveyed by these variations due to the effect of the various embedding operators: negation in (2a) leads to the assertion of just the opposite; the conditional in (2b) considers the possibility of the truth of the assertion in particular circumstances characterized by the *if*-clause; and the question in (2c) requests information as to whether this content holds or not. But the presupposition remains unaffected by all of these embedding operators. This phenomenon of presuppositions contributing to the interpretation of an utterance without being affected by various types of embedding operators is generally known as 'presupposition projection' (Karttunen, 1973). Throughout this paper we will use the term *global interpretation* for the kind of presupposition evaluation outlined above, where the presupposition is interpreted with respect to the global utterance context and not within the scope of an operator. This means that the presupposition has to be established as true in the context or else presupposition failure will occur. In most simplistic terms, presupposition failure is a mismatch of what is implicitly assumed in a sentence with what is given in the context. Howeer, at least since (Lewis, 1979), it is commonly assumed that a presupposition which was not in the context before the utterance of the sentence it occurred in, can - with certain limitations - be added to the context after encountering this sentence. This process is called *accommodation*.[2]

[2]Note that global accommodation is crucially different from the standard case of a global interpretation, where the global context supports the presupposition. If the context does not

Things are even more complicated than illustrated in (2), in that presuppositions can interact with other content in their sentence so that the contextual requirements otherwise present at the global level sometimes disappear. For example, the overall conditional sentence in (3), which has the same consequent as (2b), does not seem to come with any presupposition requiring that it counts as established in the context prior to its utterance that Tina had been ice-skating before:

(3) If Tina went ice-skating last week, then she went ice-skating again today.

This obviously has to do with the fact that an occasion of Tina ice-skating is mentioned in the *if*-clause here, but formulating a theory that systematically derives these phenomena turns out to be a hard nut to crack. Indeed, capturing the projection behavior of presuppositions theoretically has been a central challenge for a compositional semantic account of natural language meaning (i.e., an account that derives the meanings of complex expression from the meanings of their parts and the way in which they are hierarchically structured). However, a family of theories originating in the early 1980's proposed to solve this issue by incorporating aspects of contextual interpretation into the semantic representation. These theories, which see the denotation of sentences as including the interaction of the propositional content with the context, often are grouped together as broadly speaking 'dynamic' theories. We will focus on two classical instances of such theories, namely Dynamic Semantics, which started with Heim (1983) (Groenendijk & Stokhof, 1990; Chierchia, 1995, among many others), and DRT (Kamp, 1981; van der Sandt & Geurts, 1991; van der Sandt, 1992; Kamp & Reyle, 1993; Geurts, 1999). While similar in spirit in that they formally incorporate the discourse context into the process of semantic interpretation, they crucially differ in that DRT adds an additional representational

---

support it, the presupposition can either be accommodated (which involves an adjustment of the context resulting in a new context that does support it, (Lewis, 1979)) or the whole sentence ends up as infelicitous (Strawson, 1950). For embedded occurrences of presupposition triggers, there may be the additional option of local accommodation, which is standardly seen as a last resort; see discussion below.

level of so-called 'Discourse Representation Structures' (DRSs), whereas Dynamic Semantics sees the semantic effect of sentences as directly updating the information encoded by the context in a non-representational way. Both theories offer accounts of presupposition projection that overlap substantially in their predictions, and are largely successful in capturing the empirical data. However, given the difference in the representational levels involved, we arguably end up with quite different processing predictions of the two accounts, as will be illustrated below.

The Experimental Investigation of Presuppositions in Context

While presuppositional phenomena have played a role in a number of experimental investigations of language processing for quite some time, e.g., in connection with syntactic parsing decisions (Crain & Steedman, 1985) and reference resolution of definites (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, and a large subsequent literature in the visual world paradigm), they have only become a focus of psycho-linguistic work on meaning more recently. The past few years have seen a growing body of work that tries to assess the interpretive properties of presuppositional content, much of it using off-line behavioral measures. Some of the issues addressed in the literature include the strength of contextual constraints imposed by a variety of presupposition triggers (Jayez & van Tiel 2011, Amaral et al. 2011, Smith & Hall 2011), the effects of presuppositional content on resolving ambiguities (Schwarz, 2007), and the exact nature of presuppositions in conditionals and under quantification (Chemla & Schlenker, 2012; Chemla & Bott, 2013; Romoli, Sudo, & Snedeker, 2011).

There also is a growing body of work trying to understand the online processing of presuppositions at a more detailed level by looking at various measures reflecting the time course of interpreting presuppositions in online processing. Based on the general notion that presuppositions require some form of contextual support, previous experimental studies have explored experimental methods for investigating presuppositions. Schwarz (2007) found reading time effects for the part of the sentence containing *also* when the preceding sentential context did not support its presupposition. Building on this paradigm, Tiemann et al. (2011) investigated a broader range of presupposition triggers and found that unsupported presuppositions gave rise to

decreased acceptability and increased reading times on the presupposition trigger itself in word-by-word self-paced reading. The self-paced reading studies of Schwarz (2007) and Tiemann et al. (2011) found increases in reading times upon seeing the expressions giving rise to the presupposition when the context did not support the presupposition. These increases are attributed to the clash between the context and the presupposition trigger and thus can be seen as indicative of the availability of the presuppositional content, since a mismatch can only be noticed when the presupposition has been fully computed. Finally, Chemla and Bott (2013) investigate reaction times for various interpretive options of presupposition triggers like *realize* under negation and report that global interpretations are faster than local ones (see below for discussion of local accommodation).

Theoretically speaking, one approach to the interpretation of findings about the time-course of interpreting presuppositions, in particular in comparison to asserted content, is to take a perspective broadly parallel to the literature on scalar implicature processing. Many authors (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Huang & Snedeker, 2009, 2011) have argued that the pragmatic strengthening of *some* to *some but not all* (Grice, 1975) is a process that takes place online and comes with a processing delay relative to the literal *some and possibly all* interpretation. From this angle, pragmatic theories of presupposition would seem to suggest that presuppositions, too, are slow and costly to derive in online processing. Semantic presupposition theories, on the other hand, would lead us to expect that presuppositions are processed at least as early as asserted content, if not before.

However, this perspective likely is both too simplistic on conceptual grounds and contentious in empirical terms: the evidence for delayed availability of implicatures is by no means uniform, as various authors have presented evidence for rapid implicature computation (Grodner, Klein, Carbary, & Tanenhaus, 2010; Breheny, Ferguson, & Katsos, 2013); and the assumption that pragmatically generated content is only available with a delay is by no means a necessary one. The latter point is particularly relevant for our finding of immediate reflexes of unembedded presuppositions in processing, which are thus compatible either with a semantic

account or a pragmatic one that assumes rapid pragmatic processing. Importantly, however, the parallels between implicatures and presuppositions only go so far, and the projection phenomena for embedded cases add another angle to this debate. In particular, at least certain pragmatic accounts of presupposition projection do not seem to be compatible with the overall pattern of experimental results (see also Chemla & Bott, 2013, for highly relevant results, which are reviewed in the General Discussion, Part II, see).[3]

But even when restricting our attention to semantic theories theories of presupposition, it would seem that not all accounts are on par in terms of the type of processing model that they are most naturally compatible with and - correspondingly - the predictions about possible processing effects related to projection. To begin with, consider what is involved in projection on a general level. Descriptively speaking, projection could be characterized as a mismatch between the level at which an expression is introduced syntactically and the level at which it is interpreted. In a comprehension system that works incrementally, dealing with this type of mismatch could well involve additional efforts, as the different aspects of interpretation have to be sorted out and integrated at the appropriate level relative to the relevant embedding operators. The two standard theories of projection that we will consider in detail below, DRT and Dynamic Semantics, arguably differ in whether or not they would lead us to expect such additional processing efforts, based on the level at which they deal with projection: DRT assumes a representational level at which projection takes place in the form of a multi-step process involving manipulations of the representation, while Dynamic Semantics operates on a purely semantic level that is non-representational. Consequently, DRT can plausibly be seen as giving rise to the expectation that presupposition projection corresponds to costly cognitive processes in the form

[3]More specifically, most pragmatic accounts of recent fall in the same category as Dynamic Semantics in this regard. While they might lead us to expect that presuppositions in general are costly to compute, they do not seem to be able to differentiate between environments that vary with respect to the complexity of embedding involved. We'll return to this issue briefly in the general discussion.

of operations on representations, while Dynamic Semantics is unable to distinguish unembedded and embedded occurrences of presupposition triggers in terms of their potential processing efforts. A more full-fledged discussion of the relevant theoretical details is presented in the general discussion of the first set of experiments, along with a reflection on implications for more processing-oriented models of discourse comprehension.

On a more general level, the difference between the two theoretical proposals we focus on ties in directly with central questions about the processing of linguistic input in the context of a discourse. In particular, there is the general question of the extent to which information introduced linguistically is stored in a structured form. In the discourse processing literature, a distinction is commonly made between the level of a text-base, where propositions are represented in a format that is more abstract than the linguistic surface form, but still preserves crucial aspects, such as whether pairs of expressions were part of the same proposition expressed by a given clause, and that of a situation model, which encodes the overall expressed information in an even more abstract, language-independent format (Kintsch & Van Dijk, 1978). Furthermore, there is the question of whether linguistic input is processed (at least initially) relative to a fairly narrow local window (e.g., consisting of the relevant clause or sentence), or integrated with the overall discourse context right away. As already mentioned in the introduction, presuppositions constitute an ideal phenomenon for investigating such questions, as they can relate directly both to the more immediate, intra-sentential linguistic context as well as the more general discourse context. The difference between the theoretical proposals of DRT and Dynamic Semantics maps onto these issues more or less directly: the latter operates on an entirely abstract semantic level (roughly on par with situation models, though conceptualized differently) and always keeps the overall discourse content immediately at hand, as it were; in contrast with that, DRT is more procedural and first constructs a representation of the immediate linguistic context that is then integrated with the representation of the discourse as a whole. Assuming that these representations are cognitively real, DRT-style structures would be yet another level of representational structure, e.g. at the level of the text-base.

This characterization of course only represents a first shot at seeing the theoretical proposals for dealing with presupposition projection as an integral part of a more comprehensive and cognitively realistic theory of discourse processing. However, we submit that such an integration is necessary, as such a theory has to be compatible with the robust set of facts concerning presupposition projection, which constitute an intricate and crucial aspect of our linguistic capacities, and to the best of our knowledge, no comprehensive account beyond the existing formal semantic approaches exists. Independent of whether or not the semantic proposals are seen as making any direct claims about actual psychological processes, they cover important empirical ground and a corresponding processing theory will have to yield equivalent results. We see it as part of our present contribution to initiate more in-depth efforts of integrating central aspects of formal semantic theory with a theory of language processing in context.

The experiments discussed here follow the general approach of the reading time studies just mentioned in that they involve target sentences containing a presupposition trigger, German *wieder* ('again'), presented in contexts that either do or do not support the presupposition. However, we include the use of eye tracking during reading in order to allow for a more natural reading experience for the participants and to have a more fine grained temporal resolution, which allows us to capture effects closer to the real time course. In addition to considering presupposition triggers in simple sentences without any embedding, we also consider two types of embedding environments, namely negation (Experiments 1a-c) and conditionals (Experiment 2). This allows us to investigate time course effects related to presupposition projection.

Experiment 1a: Costs of Presupposition Failure in Reading

Design & Materials

The first experiment manipulated whether or not the presupposition trigger *wieder* ('again') is embedded under negation. In implementing this, we took advantage of the syntax of German, where *wieder* ('again') and *nicht* ('not') can appear in adjacent positions in either order. This allowed us to construct target sentences which are minimally different with respect to

whether *again* appears inside or outside the scope of negation. We presented such sentences in two different contexts, each of which supported the presupposition of one of the orders of *wieder* and *nicht* while being inconsistent with the other. In the sample item from our materials below, the context sentence (4i) supports the presupposition of (5a) (that Tina went ice-skating before), while (4ii) is inconsistent with it. Conversely, (4ii) supports the presupposition of (5b) (that there was a preceding occasion where Tina did not go ice-skating), while (4i) is inconsistent with it.[4]

(4) **Contexts**

i. Tina **went** ice skating for the first time with Karl last week. The weather was **beautiful**, and they **had a great time**.

ii. Tina **wanted to go** ice skating for the first time with Karl last week. But the weather was **miserable** and they **gave up on their plan**.

(5) **Target Sentences**

a. *Dieses Wochenende war Tina* ***nicht wieder*** *Schlittschuhlaufen, weil das Wetter so schlecht war.*
This weekend, was Tina not again ice skating because the weather so bad was
'This weekend, Tina didn't [go ice-skating again] because the weather was so bad.'[5]

Presupposition: Tina had been ice-skating before.

Assertion: Tina did not go ice-skating this weekend.

b. *Dieses Wochenende war Tina* ***wieder nicht*** *Schlittschuhlaufen, weil das Wetter so schlecht war.*
This weekend, was Tina again not ice skating because the weather so bad was
'This weekend, Tina again did not go ice-skating because the weather was so bad.'

Presupposition: There's a salient previous time when she did not go ice-skating.

[4]At least on the global interpretation. This generally is taken to be the default, which is supported by our data from Experiment 1b below. See discussion of local accommodation below.

[5]Bracketing in the English paraphrase is indicates the intended interpretation with respect to the scope of *again* and negation. The German sentence is unambiguous.

Sentence	Context	Firstword	Felicitous
(5a)	(4i)	*nicht*	felicitous
(5a)	(4ii)	*nicht*	infelicitous
(5b)	(4i)	*wieder*	infelicitous
(5b)	(4ii)	*wieder*	felicitous

Table 1
Overview of Conditions and Factors (Expt1a)

Assertion: Tina did not go ice-skating this weekend.

The pairing of sentences and contexts yielded a fully counterbalanced 2×2 interaction design with two factors: **Firstword** (whether *wieder* or *nicht* appeared first) and **Felicity** (whether the context supports the presupposition or not). The conditions with the combinations of the levels of the two factors are summarized in Table 1:

Procedure & Participants.

24 sentences with versions for each of the four conditions were created. In addition to the experimental items, there were 48 unrelated filler items. Participants read the sentences on a computer screen while their eyes were being tracked by an EyeLink 1000 eye tracker from SR Research. For half of the items (of both the fillers and experimental sets), participants had to answer yes/no questions, which followed directly after the sentence, to ensure full comprehension of the materials. 32 native speakers of German from the University of Tübingen community participated in the experiment. Participants were split into 4 groups, where each participant saw 6 of the sentences per condition.

Results

The primary focus in our analysis were the reading times on the verb following the {*wieder nicht*} sequence.[6] Since the presupposition of *wieder* crucially relies on the verb of its clause, it

[6]We use this set-notation to talk about the sequence of *nicht* and *wieder* in either order.

only becomes recoverable from explicitly given materials at the point of the verb. Reading times were also examined for the {*wieder nicht*} sequence itself. Standard reading measures were calculated for purposes of analysis. Based on prior self-paced reading experiments using the same general approach (Schwarz, 2007; Tiemann et al., 2011), we expect increases in reading time when sentences are presented in contexts that are inconsistent with the presupposition. The time point at which such increases arise is indicative of the relevant presupposition having been computed at this point.

All analyses used mixed-effect models with participants and items as random effects, using the *lmer* function of the *lme4* package in *R* (Bates, 2005). Given recent arguments by Barr, Levy, Scheepers, and Tily (2013) that maximal random effect structures should be used whenever possible, we generally computed models with the maximal random effect structure that would converge, with random effect slopes for each factor, as well as the interaction where applicable. To assess whether inclusion of a given factor significantly improved the fit of the overall model, likelihood-ratio tests were performed that compared two minimally different models, one with the fixed effects factor in question and one without, while keeping the random effects structure identical (Barr et al., 2013). We report estimates, standard errors, and t-values for all models, as well as the χ^2 and p-value from the likelihood-ratio test for individual factors. To facilitate presentation of results, we will adopt the following shorthand indications for which random effect structure (RES) was used in the *lmer*-syntax in a given case:

- RES-1: $(1 + factor1 * factor2|\ participant) + (1 + factor1 * factor2|\ item)$ (Full model)
- RES-2: $(1 + factor1 * factor2|\ participant) + (1 + factor1 + factor2|\ item)$
- RES-3: $(1 + factor1 + factor2|\ participant) + (1 + factor1 + factor2|\ item)$
- RES-4: $(1 + factor1 + factor2|\ participant) + (1 + factor1/2|\ item)$

For the overall interaction analyses, predictors were centered, so as to render estimates of main effects. Planned comparisons between individual conditions were conducted using the appropriate treatment-coding.

While the comprehension questions did not play a crucial role for the main purpose of the

experiment, they serve as an indicator of the extent to which participants did read the experimental materials for full comprehension. Overall accuracy in question answering was 89%, with no significant differences between conditions.

For analyzing reading times, the following standard reading time measures were computed (Rayner, 1998): first fixation duration, which measures the length of the very first fixation on the region of interest (here the verb); go-past time, which here is taken to measure the sum of all fixations on the region of interest prior to any fixations to the right of this region (but not including the time of regressive fixations); first pass time, which includes all fixations on the region when it is looked at the first time, up until leaving the region (to either the left or right); total duration, which sums all the fixations on the region of interest, no matter when they occur; regression path duration, which measures all fixations from first entering the region to first leaving it to the right (including all potential regressive fixations; this is sometimes also referred to as go past time). In addition to the timing measures, we also computed first pass regression proportion, which is the proportion of regressive eye movements following the first time of entering the region. Prior to computing the reading time measures, we removed all trials where a participant blinked at any point while looking at the $\{$*wieder nicht*$\}$ + *Verb* region. This resulted in the exclusion of 34 trials, which amounted to 4.4% of the data. The distribution of excluded trials across conditions was roughly even.

Means for the reading time measures on the verb are presented in Table 7. The primary result is an interaction between **Firstword** and **Felicity**, as illustrated by the graph for first fixation durations in Figure 1: when *wieder* was first (i.e., not embedded under negation), reading times on the verb were significantly higher in the infelicitous condition. When *nicht* was first, on the other hand (resulting in *wieder* being embedded under negation), there was no such slow-down (and except in total reading time and first pass time, no significant difference between the felicitous and infelicitous context conditions).

The interaction factor contributed significantly to model fit for first fixation duration (RES-2: $\beta = 23.2$, $SE = 10.4$, $t = 2.23$; $\chi^2 = 4.8$, $p < .05$), go-past (RES-2: $\beta = 117.1$, $SE = 54.7$,

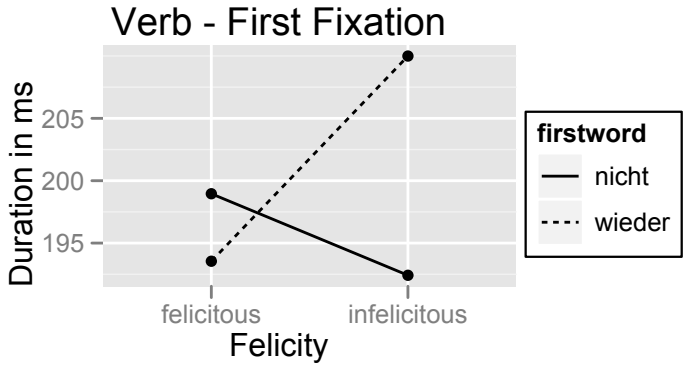


Figure 1. First Fixation Duration (in ms) by condition

$t = 2.14$; $\chi^2 = 4.3$, $p < .05$), total reading times (RES-2: $\beta = 157.6$, $SE = 55.2$, $t = 2.85$; $\chi^2 = 7.4$, $p < .01$), and regression path duration (RES-3: $\beta = 212.2$, $SE = 104.6$, $t = 2.03$; $\chi^2 = 4.09$, $p < .05$), and did so marginally for first pass duration (RES-1: $\beta = 41.8$, $SE = 23.5$, $t = 1.78$; $\chi^2 = 3$, $p < .1$). For first pass regression proportion, the interaction did not contribute significantly when random slopes for the interaction were included, but did so marginally when only random intercepts for participants and items were included (RES-3: $z = -1.79$, $p < .1$).[7] The analysis of first pass regression proportions also revealed a main effect of **Firstword** with lower regression proportions in the *nicht wieder* conditions (RES-3: $z = -2.33$, $p < .05$). In addition, there was a main effect of **Felicity** for first pass regression proportion (RES-3: $z = -2.67$, $p < .01$) and regression duration (RES-3: $\beta = 144.9$, $SE = 69.9$, $t = 2.07$; $\chi^2 = 4.08$, $p < .05$). No other main effects contributed to model fit significantly for any of the measures.

The interaction was primarily driven by a simple effect of **Felicity** for the *wieder nicht* conditions, with increases in reading measures for the infelicitous context. This contributed significantly to model fit for first fixation duration (RES-2: $\beta = 17.9$, $SE = 7.5$, $t = 2.4$; $\chi^2 = 5.73$, $p < .05$), regression path duration (RES-3: $\beta = 253.7$, $SE = 89.5$, $t = 2.84$; $\chi^2 = 5.89$,

[7]For binomial dependent variables, p-values from the *lmer* output are reported here and below.

	wieder nicht		*nicht wieder*	
Reading Measure	felicitous	infelicitous	felicitous	infelicitous
First Fixation	194	210	199	192
Go-Past	292	359	324	285
First Pass	270	281	275	247
Total	309	405	370	307
Reg. Duration	395	619	438	479
Reg. Proportion	17.0%	33.5%	17.4%	20.6%

Table 2
Reading time measures (in ms) and First Pass Regression Proportion (in %) on the verb

$p < .05$), total time (RES-2: $\beta = 93.7$, $SE = 39.4$, $t = 2.38$; $\chi^2 = 5.25$, $p < .05$), and first pass regression proportion (RES-2: $z = 3.11$, $p < .01$), and did so marginally for first pass time (RES-3: $\beta = 27.6$, $SE = 15.8$, $t = 1.75$; $\chi^2 = 3.03$, $p < .1$) and go-past time (RES-1: $\beta = 37.0$, $SE = 30.1$, $t = 1.23$; $\chi^2 = 3.22$, $p < .1$).

For the *nicht wieder* conditions, the only simple effect of **Felicity** appeared in the total reading time (RES-3: $\beta = 63.7$, $SE = 30.1$, $t = 2.11$; $\chi^2 = 4.41$, $p < .05$), and it was in the opposite direction, with a decreased reading time in the infelicitous condition. Regarding simple effects of **Firstword**, the only effect for the felicitous conditions was for total reading time (RES-3: $\beta = 63.7$, $SE = 30.1$, $t = 2.11$; $\chi^2 = 4.41$, $p < .05$), where reading times were faster in the *wieder nicht* condition than in the *nicht wieder* condition. In the infelicitous conditions, the *nicht wieder* condition displayed faster reading times on the verb than the *wieder nicht* condition for first fixation duration (RES-3: $\beta = 17.5$, $SE = 7.8$, $t = 2.24$; $\chi^2 = 4.85$, $p < .05$), go-past time (RES-2: $\beta = 81.8$, $SE = 37.0$, $t = 2.21$; $\chi^2 = 4.62$, $p < .05$), total time (RES-3: $\beta = 91.6$, $SE = 43.2$, $t = 3.23$; $\chi^2 = 10.32$, $p < .01$) and first pass regression proportion (RES-1: $z = 3.01$, $p < .01$), as well as marginally for first pass time (RES-2: $\beta = 33.6$, $SE = 17.6$, $t = 1.91$; $\chi^2 = 3.49$, $p < .1$).

With respect to reading times on {*wieder nicht*} (taken as a unit, in either order), a parallel interaction effect showed up in the total reading times, with corresponding simple effects of

401 **Felicity** for the *wieder nicht* condition and of **Firstword** for the infelicitous condition. There
402 were no other significant effects in this region.

403 Given the lack of an effect of **Felicity** in the *nicht wieder* condition on the verb, follow-up
404 analyses on later and larger regions were carried out. No increases in reading times were found for
405 regions consisting of the verb plus 2 following words, the 3 words following the verb, the entire
406 section of the sentence from *nicht wieder* to the end, or, for that matter, for the entire trial
407 duration (i.e., total reading time for the entire paragraph).

408 *Discussion*

409 There are two main points to discuss with respect to the experimental results. The
410 interaction shows that the effect of encountering a presupposition in a context that is inconsistent
411 with it differs based on whether we are dealing with an embedded or an unembedded trigger.
412 Furthermore, the presupposition of unembedded *wieder* gives rise to fairly immediate effects of
413 inconsistency that are reflected throughout a variety of reading time measures. Of particular
414 interest with regards to the latter point are the simple effects for first fixation duration and first
415 pass regression proportion. Already during the first fixation of the verb (which last less than
416 200 ms), the beginning of which arguably is the logically earliest point possible to fully compute
417 the presupposition of *wieder* based on what has been explicitly provided, a 12 ms effect emerges.
418 Based on the experimental design, the increase in the infelicitous condition can be attributed to
419 the inconsistency between the expressed presupposition and the provided context. But for such
420 an inconsistency to arise, the relevant presupposition must of course have been computed.
421 Similarly, the increase in first pass regression proportions indicates that upon first looking at the
422 verb, there is an increased likelihood of returning to look at the preceding context, which is
423 presumably triggered by noticing the same inconsistency. The experiment thus provides evidence
424 that the presupposition of *again* is computed rapidly online. As discussed above, such a result
425 could be seen as theoretically compatible either with a semantic view of presuppositions or with a

pragmatic view that assumes rapid pragmatic processing.[8]

Returning to the first point, the picture is rather different for cases where *wieder* is embedded under negation. Assuming the standard global interpretation, the two contexts vary in precisely the same way as was the case for the unembedded occurrence of *wieder*, with one context supporting the presupposition, while the other is inconsistent with it. If the global interpretation of the presupposition were available while reading the verb, we would expect to see an effect on reading times similar to the unembedded condition. However, on none of the reading measures was there a significant increase for the infelicitous condition. In fact, the only significant simple effect (for total reading times) went in the opposite direction. The lack of such an increase thus can be taken as an indication that the global interpretation is not available while the verb is being read.[9]

In principle, there are two possible explanations for why this might be the case. First, it could be that more is involved in deriving a global interpretation in the context of an embedding operator like negation, compared to simply recognizing the presupposition of an unembedded trigger. Thus, the lack of an effect might be due to a lag in generating the global presupposition in this more complex sentential context. An alternative exists, however, based on the possibility

[8]The extent to which these results generalize to other presupposition triggers remains to be explored. Triggers very well could vary in their behavior, based on proposed theoretical distinctions in the literature (Simons, 2001).

[9]Note that the pattern here cannot be attributed to differences in ease or speed of accommodation, as accommodation is of no help in any of the conditions: in the felicitous conditions, the (global) presupposition is supported by the context; in the infelicitous ones, it is inconsistent with it, leading to a clash. Even if local accommodation were involved in the *nicht wieder*-condition - see discussion below -, we would expect differences in outcome for the felicitous vs. infelicitous conditions if we assume differences in ease of accommodation (presumably with delays for local interpretations, based on the results by Chemla & Bott, 2013; Romoli & Schwarz, submitted).

(glossed over in our discussion so far) of so-called *local* accommodation of presuppositions in the scope of negation. Perhaps the most well-known case of this involves the existence presupposition of the definite article, as in (6). While definite descriptions such as *the King of France* presuppose that there is an individual with the relevant property, such a global interpretation of the presupposition is inconsistent with the continuation in this case. Nonetheless, there is a consistent interpretation of the entire text, which shows that it is possible to interpret the presupposition locally, so that the existence of a King of France is part of what is denied by the negation.

(6) The King of France is not bald - because there is no King of France!

Similarly, in (7) it seems possible to negate the presupposition that Tina had been ice-skating before, rather than the asserted content (note that these interpretations may require a special intonation).

(7) Tina didn't go ice-skating again last weekend - this was the first time!

A simple way of modeling such a local accommodation is to simply assume that both the presupposed and the asserted content remain in the scope of negation, so that the overall interpretation of the sentence can be paraphrased as follows:

(8) NOT [Tina went ice-skating before AND went ice-skating this weekend]

While in principle, the falsity of either conjunct in the scope of negation would suffice to make this true, the fact that one could express the negation of the second conjunct more straightforwardly (by simply leaving out the presupposition trigger altogether) seems to bias this towards an interpretation where it is indeed the falsity of the conjunct contributed by the presupposition trigger that is conveyed by an utterance of this sentence. Thus, local accommodation gives rise to the claim that Tina had not been ice-skating before.

Without going into further details about local accommodation, it is clear that given a paraphrase along the lines of (8), the existence of local accommodation for the target sentence in the experimental materials would make the *nicht wieder* sentences perfectly consistent with either

context. If the context states that Tina had been ice-skating some time recently, then the global presupposition of course remains consistent with that (note that the local accommodation interpretation is not strictly speaking inconsistent with this either, if the paraphrase above is correct). And if the context states that she did not go ice-skating (and had never done it before, either), then the local accommodation interpretation is perfectly consistent with that. Thus, if both global and local accommodation interpretations for the presupposition of *wieder* in the scope of negation were equally available, we would not expect to see any reflexes of inconsistency in the reading times, since at least one of the readings always is consistent with the given contexts.[10] In order to test whether local accommodation is indeed available for the experimental materials, a follow-up rating experiment was carried out.

Experiment 1b: Acceptability Rating Study of Presupposition Failure in Embedded and Unembedded Contexts

Design & Materials

If local accommodation is indeed available for the presupposition of *wieder* when it appears in the scope of negation, we would expect this to affect speakers' acceptability judgments of these sentences in the two contexts, since infelicity should have a direct impact on acceptability.[11] In particular, the type of interaction that we saw in the reading times should also be present in the

[10] As a reviewer points out, we might expect corresponding differences in reading times if the two readings are both available in principle, but one is strongly preferred and more easily accessible. However, our results don't bear on this issue, since they don't yield any general difference between the two *nicht wieder* conditions.

[11] Acceptability is clearly a gradient notion that is affected by a host of factors, including non-linguistic ones. Whether felicity is categorical or gradient (and whether it is distinct from acceptability in the first place) is not a straightforward matter. For present purposes, all that we need to assume, however, is that felicity and acceptability are correlated, so that a decrease in felicity is reflected in a decrease in acceptability.

judgments. If local accommodation is not available (or only available to a very limited extent), on the other hand, the *nicht wieder* sentences in what we have labeled as the infelicitous context above should be judged to be less acceptable than in the felicitous context.

Procedure & Participants

A rating questionnaire was conducted via the web using the WebExp2 software (`http://www.webexp.info`). The materials were exactly the same as those used in the eye tracking experiment, including all the fillers. Participants were asked to rate the appropriateness of a given discourse on a scale from 1 (least appropriate) to 5 (most appropriate). Data from 24 participants was collected.

Results

The results in form of the mean ratings by condition are summarized in table 3. While there was a marginally significant interaction between **Firstword** and **Felicity** (RES-3: $\beta = .42$, $SE = .22$, $t = 1.90$; $\chi^2 = 3.57$, $p < .1$), as well as a main effect of **Firstword** (RES-3: $\beta = .52$, $SE = .14$, $t = 3.74$; $\chi^2 = 11.5$, $p < .001$), more importantly for our purposes there was a clearly significant main effect of of **Felicity** (RES-3: $\beta = 1.13$, $SE = .13$, $t = 8.94$; $\chi^2 = 35.71$, $p < .001$), with items containing felicitous contexts getting higher (= better) ratings than those containing infelicitous contexts. While this effect was slightly more pronounced in the *wieder nicht* items (as reflected in the marginal interaction), there nonetheless is a significant simple effect for *nicht wieder* in the same direction (RES-3: $\beta = .92$, $SE = .17$, $t = 5.51$; $\chi^2 = 24.18$, $p < .001$), just as there is for *wieder nicht* (RES-3: $\beta = 1.34$, $SE = .17$, $t = 7.98$; $\chi^2 = 39.2$, $p < .001$). With regards to simple effects of **Firstword**, the *wieder nicht* order was judged significantly better in the felicitous condition (RES-1: $\beta = .73$, $SE = .18$, $t = 4.12$; $\chi^2 = 13.10$, $p < .001$), and marginally better in the infelicitous condition (RES-3: $\beta = .32$, $SE = .18$, $t = 1.76$; $\chi^2 = 3.06$, $p < .1$).

Discussion

The rating results show that for both embedded and unembedded *wieder*, the **Felicity** manipulation had a clear effect and resulted in decreased acceptability when the context sentence

	wieder nicht		*nicht wieder*	
	felicitous	infelicitous	felicitous	infelicitous
Mean Rating	3.94	2.63	3.23	2.34

Table 3
Results of the rating experiment

was inconsistent with the (global) presupposition of *wieder*. This would be unexpected if local accommodation of *wieder* under negation was readily available. A possible explanation of the reading time results in terms of the availability of such an interpretation thus is undermined by the rating results. Even if it is temporarily considered, it does not seem to be the one that is ultimately adopted in the general case.

While the common assumption in the theoretical literature that global interpretations of presuppositions constitute the default is supported by the data, it is consistent with the results considered so far that such an interpretation is not available immediately. However, it remains somewhat puzzling that the reading time study found no effect for the *nicht wieder* conditions at all, rather than a delayed effect. If we want to pursue an interpretation of the data in terms of a delay in the availability of global interpretations of presupposition triggers introduced in embedded environments, it would be helpful to have a more direct link between time-course effects and evidence for the presence of the global interpretation. The rating experiment actually provides a potential first hint in this regard, as there is a numerical interaction in the reaction times for the different conditions, with responses being slower in the *nicht wieder* infelicitous condition than in all others. However, this pattern did not reach significance, which is likely due to the course-grained nature of the response time collection method and the various uncertainties of web-based data collection. The experiment reported in the following section was designed to establish a link between the time-course of interpretation and the availability of global interpretations.

Experiment 1c: The Time-Course of Infelicity-Judgments

Design & Materials

In the reading time study, we have no direct evidence on what interpretation participants are pursuing at any given point. In the rating study, we do not have any detailed time-course information. The present experiment attempts to tie the two dimensions together by employing a stops-making-sense task, where participants read the same materials in a self-paced fashion, with a word-by-word presentation of the target sentence. In addition to the self-paced reading task, participants were instructed to abort trials by pushing another button as soon as they felt that the text no longer made sense.

The stops-making-sense design allows us to collect information on two dependent variables. One is the rate of stops-making-sense judgments in the conditions where the presuppositions and the context are inconsistent, which further supplements the results from the rating experiment. Secondly, we get fine-grained response-time data that is different from the reading time data in the first experiment in interesting ways: given the presence of an ultimate judgment on the felicity of the stimulus at hand, we are able to assess the time course of the rejection decisions more directly. For cases where participants continue reading to the end in the felicitous conditions, we get self-paced reading data for the target sentence on a word-by-word basis.

The materials used in this experiment were the same as in the previous two experiments, including fillers.

Procedure & Participants

Participants were seated in front of a computer, and were instructed to read through sentences on the screen, which would be presented bit by bit, with advances to the next part being triggered by pressing a key on the keyboard. The sentences were presented in random order using the Linger software (`http://tedlab.mit.edu/ dr/Linger/`). Context sentences were presented sentence by sentence, and the target sentence was presented word by word. Participants were instructed that if the text on the screen stopped to make sense, they should press another

condition / word	WN1	WN2	Verb	Verb+1	Verb+2	Verb+3
wieder nicht	2	94	85	12	2	4
nicht wieder	9	70	83	17	9	2

Table 4
Number of aborted trials per word

button. Data from 48 participants was collected, all of whom were members of the University of Tübingen community.

Results

In this experiment, our main focus of analysis are the two infelicitous conditions. As far as the frequency of stops-making-sense judgments is concerned the *nicht wieder* and *wieder nicht* conditions are essentially on par at 79.1% and 82.9% respectively, without any significant difference between them. The felicitous conditions were judged to not make sense about 20% of the time (the relatively high number of false alarms presumably is due to the nature of the task; comparable rates were found for filler items for which there is no obvious issue with respect to whether they make sense). This supplements the results from the rating study in that the infelicity is clearly detectable in the materials, and it seems to be noticed by participants at the same rate for the embedded and unembedded *wieder* conditions. Thus, there is further evidence against a general availability of local accommodation of the presupposition in the former.

Turning to the response time-course in the infelicitous conditions, there are a number of ways of looking at the data. To begin with, the distribution across words during which the trials were aborted was comparable in the *wieder nicht* and *nicht wieder* conditions, with perhaps a slight shift towards the right in the latter, as can be seen in Table 4. In order to evaluate the time-course of stops-making-sense judgments in more detail, what we are after is the time span lasting from initial access to the presupposition to the stops-making-sense button press. Given that more than a third of the overall stops-making-sense judgments are already made while

looking at the second word in the {*nicht wieder*} sequence (presumably because the verb is relatively predictable based on the stimuli) we decided to start the clock with the display of that second word (*nicht* in the *wieder nicht* condition and *wieder* in the *nicht wieder* condition), and then sum all following per-word response times up to the point of the stops-making-sense judgment.[12] The mean rejection times calculated in this way for all trials ending in a stops-making-sense judgment are presented in the first results column in Table 5. (For the purposes of aggregation and analysis, outliers that were more than 3 standard deviations from their condition mean were removed.) Rejections were significantly slower in the *nicht wieder* condition (2799 ms) than in the *wieder nicht* condition (2258 ms (RES-1: $\beta = 538.0$, $SE = 202.6$, $t = 2.66$; $\chi^2 = 6.49$, $p < .05$).

In light of potential concerns about the fact that the reading times on *wieder* itself are not included in the calculation of the rejection times based on the method just laid out, we also considered another perspective. In particular, we computed rejection times by summing all reading times starting with *wieder* in both conditions. This, however, results in an imbalance, since there is an additional word that has to be read in the *wieder nicht* condition (namely *nicht*). In order to take this into account, we factored out the reading time for the word *nicht* in the most conservative way, namely by calculating average reading times for *nicht* in the *nicht wieder* condition in the felicitous context for each participant. These mean *nicht* reading times were subtracted from the rejection time in each aborted *wieder nicht* trial in the infelicitous context,

[12] A related point of interest further supporting our choice is that we see a relative slow-down on the second word in both of the felicitous conditions: In both cases, the second word in the sequence is read more slowly than when it appears in the first position in the other condition: *wieder* is significantly faster in the *wieder nicht* condition: 449 ms vs. 610 ms (RES-1: $\beta = 161.7$, $SE = 43.29$, $t = 3.74$; $\chi^2 = 12.82$, $p < .001$); and *nicht* is significantly faster in the *nicht wieder* condition: 514 ms vs. 587 ms (RES-1: $\beta = 77.87$, $SE = 37.54$, $t = 2.07$; $\chi^2 = 4.19$, $p < .05$). This suggests that the main burden of processing of the {*wieder nicht*} sequence arises on the second word, regardless of their order.

condition	from WN2
wieder nicht	2258
nicht wieder	2799

Table 5
Rejection Times (in ms) from second word in wieder nicht sequence in aborted trials in infelicitous contexts

on a by-participant basis. The resulting rejection times yielded an even greater difference between the conditions, which also was significant (RES-1: $\beta = 641.4$, $SE = 186.2$, $t = 3.445$; $\chi^2 = 10.10$, $p < .001$).

As a final measure, we also looked at the self-paced reading times in the felicitous conditions. Reading times in the *nicht wieder* condition (1124 ms) were marginally higher than in the *wieder nicht* condition (1036 ms) when looking at both words in the {*wieder nicht*} sequence (RES-1: $\beta = 81.26$, $SE = 45.72$, $t = 1.78$; $\chi^2 = 3.15$, $p < .1$) as well as pooling these together with the verb (1625 ms vs. 1788 ms) (RES-1: $\beta = 133.97$, $SE = 78.42$, $t = 1.708$; $\chi^2 = 2.90$, $p < .1$).[13] This suggests that the even in the felicitous conditions, greater effort was involved in reading when *wieder* was embedded under *nicht*.

Discussion

The results from the stops-making-sense experiment provide a perspective on the time-course with which the judgments caused by the infelicity of the presupposition of *wieder* arise. Consistent with our proposed interpretation of the eye tracking reading study, it takes longer to reject the materials based on the presupposition of *wieder* when *wieder* is embedded under negation. Furthermore, the reading times from the felicitous conditions provide at least

[13]Looking at just the verb, there was a numerical difference in the same direction (664 ms vs. 589 ms) which did not reach significance (RES-1: $\beta = 77.01$, $SE = 51.32$, $t = 1.50$; $\chi^2 = 2.24$, $p < .15$).

suggestive evidence that the embedded *wieder* condition is more effortful, with somewhat elevated reading times on the *nicht wieder (+Verb)* sequence (compared to *wieder nicht (+Verb)*). Finally, the rejection rates for the *wieder nicht* and *nicht wieder* conditions did not differ from one another significantly, which further corroborates the results from our rating study (Experiment 1b). If local accommodation was widely available in the *nicht wieder* condition, we would expect a lower rejection rate there.

General Discussion - Part I

The results from the series of experiments reported above suggest that presupposition projection requires extra time in processing. In particular, we find evidence for rapid availability of unembedded presuppositions, while global interpretations of presupposition triggers introduced in embedded environments are only available with a delay in comparison. The three experiments reported each contribute a unique piece to the overall story:

In Experiment 1a, we found immediate increases in reading times when the presupposition of unembedded *wieder* was inconsistent with the context. However, no such effects were found when *wieder* was embedded under negation. The result for the embedded condition provides the most detailed evidence to date on the online time course of presupposition interpretation (along with recent results from the visual world paradigm; see Romoli, Khan, Snedeker, & Sudo, submitted; Schwarz, submitted).

Responses by participants in the rating task in Experiments 1b showed that the ultimate interpretation of *wieder* does indeed involve a global interpretation of the presupposition, which requires projection: contexts that are inconsistent with such an interpretation, but not with a local accommodation interpretation, were judged to be less acceptable than contexts that are consistent with it. This result is in line with the assumption standardly made in the theoretical literature that global interpretations of presupposition triggers in such embedded environments are the default. However, ratings for the *nicht wieder* condition were somewhat lower than for *wieder nicht*, which could be seen as an indication that the former are more difficult to process.

Experiment 1c involved categorical responses by participants, and materials with contexts

inconsistent with the presupposition were rejected as frequently for the embedded condition as for the unembedded one, providing further support for global interpretations being the dominant ones for embedded *wieder*. More crucially, while the eye tracking results on their own only support the notion of a delay for presupposition projection indirectly (since there was no delayed effect of inconsistent contexts, but rather no effect at all), the rejection time results from the stops-making-sense task in Experiment 1c provide more direct evidence along these lines, as it took longer to reject a text in inconsistent contexts in the *nicht wieder* condition than in the *wieder nicht* condition. Finally, the reading times from the felicitous context conditions provided further evidence that the *nicht wieder* condition involved additional effort, based on marginally increased reading times on the combined *nicht* and *wieder* regions (both including and excluding the verb).

The immediate effects in the unembedded condition replicate and enhance the findings in (Schwarz, 2007) and (Tiemann et al., 2011), where unembedded presuppositions which were inconsistent with the context also resulted in processing difficulties, as reflected in self-paced reading times. They provide further evidence that presuppositions are evaluated as soon as possible, and with the more fine-grained temporal resolution provided by eye tracking. The results of experiment (1a) and (1c) show that the processing of an embedded presupposition is more intricate.

Given that the contexts and target sentences were minimally varied, and based on the descriptive notion that projection involves a mismatch in the syntactic location of a trigger and the level at which it is interpreted that could lead to increased processing efforts, it is natural to attribute the observed effects to the process of presupposition projection. In order to flesh out this explanation in more detail, we review two of the most prominent linguistic theories that successfully capture the projection behavior of presuppositions below. Despite their conceptual similarities and a large amount of overlap in their analysis of projection, we conclude that they differ when it comes to integrating them into a processing framework, as one involves an additional layer of complexity in the assumed representations relevant for projection. In the final

part of this section, we consider the implications from this discussion for cognitive processing theories of discourse in more general terms.

Presuppositions and Projection in Two Classical Theories

Discourse Representation Theory: Projection as operations on representations

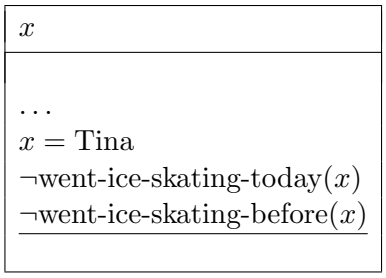
The key feature of DRT is that it provides a mechanism for incorporating the meaning of new sentences into a representation of the discourse context that encodes previously expressed information. It does so by translating individual sentences into discourse representations that can be combined with the representation of the prior context. Discourse representations can be conveniently presented using a box format, where the boxes consist of a 'header' and a main body. The header contains a list of discourse referents or reference markers, which are variables. The main body contains descriptive conditions, i.e., it lists predicates that are said to hold of certain of these variables. Presuppositions are seen as essentially anaphoric expressions in DRT, in that they have to be identified with an antecedent somewhere in the discourse structure (van der Sandt & Geurts, 1991; van der Sandt, 1992). They are initially introduced in the DRS corresponding to the clause they are part of syntactically, but a subsequent procedure manipulates the discourse structure based on a search along a well-defined search path for a suitable antecedent. Discourse representations for the schematically represented target sentences in (9) are presented in (10) and (11).[14]

(9) a. Tina AGAIN NOT went ice-skating today.

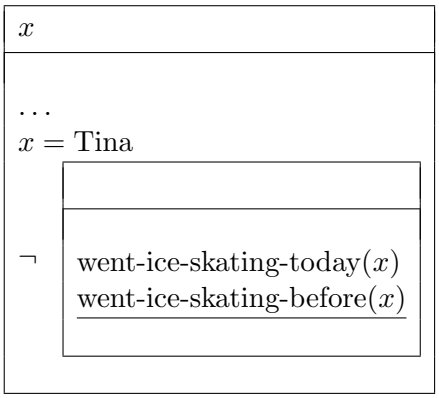
b. Tina NOT AGAIN went ice-skating today.

[14]To avoid the complexities involved in representing events, we simply use complex predicates that can include adverbs like *today*. '...' is used as a place holder for conditions introduced by earlier parts of the discourse. Presupposed material is marked by underlining; in the case of simple negated conditions, it is sometimes convenient to simply write the condition preceded by negation, without including an extra box.

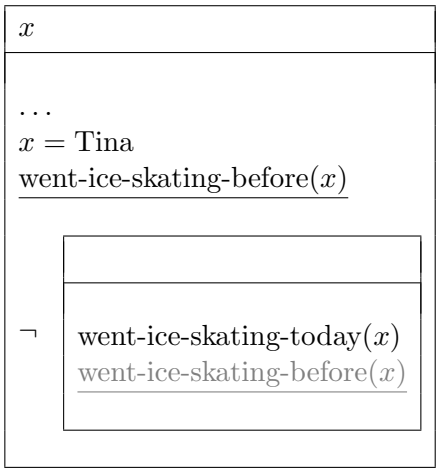
(10)



(11) a.



b.



When *again* is unembedded, its presupposition is introduced at the top level of the discourse representation from the start, i.e., there is no projection. The presupposition then simply has to be resolved by finding a condition introduced by the preceding discourse that matches the presupposition. In other words, it must have been established before the utterance of the sentence that there is some (salient) occasion on which Tina did not go ice-skating. If this is not the case, the presupposition may be accommodated (at least in certain circumstances).[15]

When *again* is embedded under negation (represented by an embedded box in DRT terms), the process is more complicated. The presupposition of *again* (here that Tina had been ice-skating before) is initially introduced in the embedded DRS, since that corresponds to the clause it is part of syntactically (11a). However, the mechanism responsible for handling projection places the underlined condition in the global DRS, yielding (11b). (To keep track of its origin, we will use gray renderings of the condition in its previous location(s).) Now that the presupposition is

[15]This may or may not be an option for *again* - Tiemann (2013) presents experimental evidence that the presupposition of *wieder* is generally not accommodated. In any case, accommodation is not an option in our experimental materials as the discourse context is incompatible with the presupposition in the infelicitous condition.

part of the global DRS, its resolution proceeds in the same fashion as for the unembedded case .

Dynamic Semantics: Projection via constraints on context updates The second theory which we will consider in connection with the processing of embedded presuppositions is that of Dynamic Semantics (Heim, 1982, 1983; Groenendijk & Stokhof, 1990; Chierchia, 1995). This theory takes the perspective suggested by Stalnaker (1973) and Karttunen (1973) that presuppositions are conditions on the *context* of use and incorporates these contextual conditions directly into the semantics. It does this by rethinking the meanings of sentences in terms of the impact they have on any given context - their *context change potential*. Formally speaking, the context is a set of possible worlds in which everything that counts as established by the discourse participants holds to be true. The context change potential of a sentence is understood as the impact it has on the context in terms of further restricting the set of worlds to those where the expressed proposition holds, rendering a new context c'. Negation removes worlds in which the negated proposition holds from the context. An *update* of the context c with our unembedded *wieder* target sentence is illustrated in (12c).[16] The effect of the negated proposition that is asserted (12a) is to remove those context worlds in which Tina went ice-skating today.

(9a) Tina AGAIN NOT went ice-skating today.

(12) a. $S' = NOT\ S$, where S: Tina went ice-skating today.

b. PSP: There's a (salient) earlier time where Tina did not go ice-skating.

c. $c' = c + S' = c + NOT\ S = c - (c + S)$

(defined iff $c + PSP = c$)

The role of presuppositions can now be seen as placing conditions on the definedness of context updates. In particular, an update of c with S' can only take place if its presupposition is already entailed (or *satisfied*) by the context (in which case adding the presupposed proposition to the context would not remove any worlds). If this is the case, the context *admits* the sentence and can be updated as in (12a).

[16]The '$-$' symbol here is essentially equivalent to forming the set-theoretic difference, i.e., $c - X$ will yield a set containing all those c-worlds that are not part of X.

One major goal of Dynamic Semantics is to account for the behavior of presuppositions in embedded contexts, e.g. in the scope of negation, conditionals, or quantifiers. Our example target sentence with embedded *wieder* illustrates the case of negation.

(9b) Tina NOT AGAIN went ice-skating today.

(13) a. $S' = NOT\ S$, where S: Tina went ice-skating today.

b. PSP: There's a (salient) earlier time where Tina went ice-skating.

c. $c - (c + S)$

(defined iff $c + PSP = c$)

Note that even though the presupposition trigger *again* is part of the negated clause, the definition of the context update for negation ends up requiring the presupposition to hold in the global context c, because it involves the update $c + S$ as its initial step. This is because negation formally involves first updating c with the clause it negates, and then subtracting the result from c. This is precisely what captures the projection effect, in that the presupposition (that Tina has been ice-skating before) is evaluated relative to the original context c, just like it would if negation were not present.

Presupposition Projection from a Processing Perspective

We now turn to the question of how these theories might relate to online processing, and more specifically, to our experimental results. In the final part of this section, we take a broader perspective by relating this discussion to more general processing models of discourse.

Processing Implications of the Two Theories The theories we considered are not processing models themselves, of course. However, we believe that they come with implications for processing, based on their general properties. Furthermore, given that any comprehensive theory of language comprehension in context will have to tackle the projection problem, some version of a semantic account of these phenomena ultimately has to be incorporated into a broader cognitive theory.

Starting our discussion with DRT, it is worth noting that the structural representations of this theory are commonly intended to correspond to (or at least model parts of) cognitively real

representations of the discourse. To the extent that we assume the process of presupposition projection that we illustrated in (11) to correspond to actual cognitive processes involved in deriving the appropriate interpretations for embedded occurrences of presupposition triggers, this could plausibly be expected to involve additional processing effort and time, compared to unembedded occurrences. The delayed availability of the presupposition introduced by *again* in the scope of negation in our experiments thus can be seen as being a reflex of the added theoretical complexity of projection posited by DRT.

Dynamic Semantics, on the other hand, does not give rise to any expectation of added complexity due to projection.[17] Crucially, the process of evaluating the presupposition in the two cases (9a vs. 9b) is of exactly the same nature. It is part and parcel of the definition of negation in dynamic semantics that we evaluate the presuppositions of a negated sentence relative to the global preceding context c. Therefore, checking whether the context update process is defined involves the same step in both of our examples, and no difference in processing complexity concerning the evaluation of the respective presuppositions is expected.[18]

Despite being largely on par in terms of capturing the central facts about presupposition projection, the two semantic theories considered thus differ in how well their properties translate into a processing account of our findings. The representations by DRT and the increased complexity introduced by projection are the key factor behind these differences. Dynamic Semantics lacks a structured level of meaning representation where there is no immediate access to the global context.

Regarding the issue of the general time course of presupposition evaluation in unembedded cases, note that both of the accounts (at least in their standard form) are compatible with the

[17]This leaves open, of course, the option of alternative explanations of the results, e.g., in terms of processing alone. We consider some potential alternative explanations in the final section of the paper.

[18]The only difference is that the presupposition in (9a) is itself 'negated' (loosely speaking; it's just a proposition at the semantic level, of course). But if anything, that might lead us to expect that case to be harder, if there were any extra efforts involved in dealing with negation (e.g., because the update procedure is more complex), contrary to our findings.

notion that presupposed content is available at least as early as literal, truth-conditional content. Dynamic Semantics could perhaps be seen to make an even stronger prediction in this regard, namely that presuppositions are evaluated *before* literal, truth-conditional content, since presuppositions have to be checked prior to updating the context with the asserted information (cf. Beaver, 2001). Whether or not this aspect of the theory translates into any processing effects remains an open question here, and our results do not speak to this directly.

Relation to Discourse Processing Models While the contrast between DRT and Dynamic Semantics provides a very precise illustration of the issues arising from the present experimental results, it is also worth relating these considerations to a broader perspective on language processing. Many processing theories of discourse comprehension crucially involve the notion of a situation model (e.g., Zwaan & Radvansky, 1998),[19] as well as an intermediate, gist-like level of a text-base (Kintsch & Van Dijk, 1978), in addition to the surface structure representing the precise linguistic form. A situation model represents the accumulated representation of the discourse, potentially including background knowledge and additional inferences, perhaps along central dimension such as time, space, and causation (among others). Crucially for our purposes, however, the situation model does not reflect any structural aspects of the original linguistic form, but merely represents the information expressed. At least in this regard, it is not unlike the Stalnakerian notion of context utilized by Dynamic Semantics, which consists of a set of propositions, construed as possible worlds. In contrast, the level of the text-base is generally construed as a an intermediate level in that it preserves some of the major structural aspects of the linguistic surface form (e.g., whether two elements were part of one proposition expressed by a single clause).

While we are not aware of any theory in this realm that addresses presuppositions in general, let alone their projection behavior, it is worth considering what the options for such an account might look like. To begin with, consider the possibility that presuppositions are evaluated at the level of situation models. A distinction is typically made between (a) the *current* model under construction at Time t_n, (b) the *integrated* model of the situations at Times t_i through t_{n-1}, and (c) the *complete* model of the situations at Times t_1 through t_x (Zwaan & Radvansky,

[19] Or related notions such as a mental model (Johnson-Laird, 1983).

1998, p.165). For the issue of presupposition processing, the interplay of the *current* and the *integrated* model would seem to be of central interest. Presuppositions require contextual support and thus, when a presupposition comes up, a link to the situation model prior to the present utterance has to be made. It seems perfectly plausible to suppose that access to the integrated situation model for purposes of presupposition evaluation could involve more effort than just working on the current situation model, e.g., because the former requires retrieval from long term working memory, whereas the former is readily available in short term working memory. However, this would not result in any predictions about differences between presupposition evaluation for embedded vs. unembedded *wieder*, since both cases involve access to the integrated model (in the form of the information introduced in the context sentence in our materials). Given the rough parallel between the role of the situation model and the Stalnakerian notion of context, the problem here is the same as for the Dynamic Semantic account above, in that there is no way of differentiating the processes of relating the presupposition to the global context in the two cases.

A DRT-style account, which crucially relies on structures that preserve (part of the) information encoded in the linguistic form, on the other hand, would seem to fit quite naturally with the idea that presupposition evaluation takes place at the level of the text-base as an intermediate level of representation. Of course, this would require the assumption that information such as that encoded by the hierarchically organized discourse structure of DRT is present at this level. Assuming that the semantically crucial information relating propositions to one another that is expressed by connectives and clausal operators of various sorts needs to be represented at this level in any case, such an extension would seem both welcome and necessary. While we are not in a position to spell out an integrated theory of a text-base with DRT-style discourse representations, we think that pursuing such an integration seems worthwhile and timely. The more general upshot of our present discussion is that our results support the cognitive reality of a structured representation of discourse that is relevant for the processing of presupposed content. The natural level of incorporating such structured information in a comprehensive theory of discourse processing would seem to be one akin to a text-base in the sense alluded to above.

Interim Conclusion

While we think that the evidence from the studies reported so far provides compelling evidence for the notion that presupposition projection involves additional effort in processing, with important implications for the levels of structured representation at play in discourse comprehension, there still are various limitations on the generalizability of the results. The extent to which projection was required was limited to one level of embedding, and only involved one type of embedding expression, namely negation. Furthermore, we only were able to detect effects based on the representation of the presupposed meaning indirectly, by presenting presuppositions in infelicitous contexts. The second experiment, discussed in the next section, addresses these issues.

Experiment 2: Presupposition Projection in Conditionals

To further test the hypothesis that presupposition projection is an effortful process, due to the complexity introduced by the required operations on discourse representations, the materials from Experiment 1a were modified so as to include a conditional, which provides an additional layer of embedding. Another change from the previous experiment was that all of the discourses had a felicitous interpretation, though with variation in the location where support for the presupposition was introduced. This provides a broader perspective on how projection is realized in processing, and allows us to compare different processing accounts of projection-related effects in greater detail. The effects in Experiments 1a-c were based on a single step of projection, and detected in light of clashes with the context. But if projection in general, and perhaps the length of the projection path in particular, are what is behind the results of those experiments, then we should also find projection effects in felicitous contexts when we vary the distance (in terms of the hierarchical discourse structure) between the presupposition trigger and the clause supporting it. Experiment 2 thus provides a more general test of our proposed explanation of the results from Experiments 1a-c.

Methods

Design & Materials Our items were created according to the following pattern: the presupposition of the target sentence was either supported by the global context sentence or by

Condition	Context	Location	Firstword
a	I	local	*wieder*
b	I	global	*nicht*
c	II	global	*wieder*
d	II	local	*nicht*

Table 6
Overview of Conditions and Factors

the local *if*-clause. This corresponded to a manipulation of whether negation was present in the context sentence or in the antecedent of the conditional. As before, we manipulated the order of *wieder* and negation in the target sentence, which in this case was the consequent of the conditional. This yielded a 2 × 2 design, which could be characterized by pairs of factors in various ways. Table 1 provides an overview of the different possible groupings based on what pairs of factors we consider.

(14) *Tina war letzte Woche* {***(I)*** *∅ / **(II)** nicht*} *Schlittschuhlaufen. Wenn sie gestern* {***(I)*** *nicht / **(II)*** *∅* } *Schlittschuhlaufen war, dann...*
Tina was last week ∅ / not} ice-skating. If she yesterday not / ∅ ice-skating was, then...

(15) *...geht sie heute bestimmt* {***(WN)*** *wieder nicht / **(NW)** nicht wieder*}
...goes she today certainly *not again / again not*
Schlittschuhlaufen.
ice-skating.

Participants and Procedure 24 items with variations of the illustrated pattern were created, each with 4 versions for the four conditions. 32 native speakers of German from the University of Tübingen community participated in the experiment. Participants were split into 4 groups, where each participant saw 6 of the sentences per condition, providing us with a balanced number of data points from all conditions for each item and participant. There were 50 filler sentences from other, unrelated experiments. Participants read all sentences on a computer screen while we recorded their eye movements with an EyeLink 1000 eye tracker. Half of the items were followed by a simple yes/no question to ensure that participants were reading the materials for

876 comprehension.

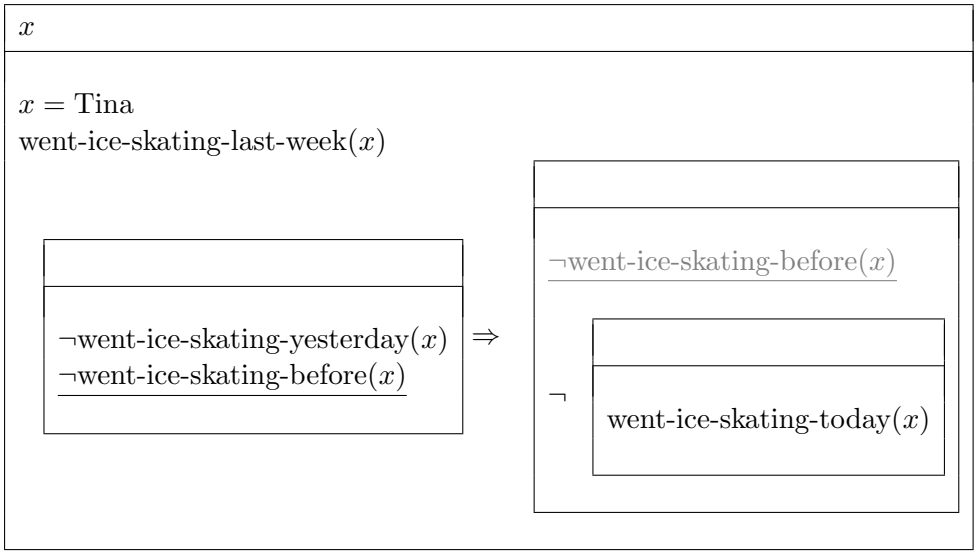
877 *Predictions*

878 Before turning to the results, let us spell out the predictions of the theoretical accounts
879 under consideration in detail. To do so, we have to introduce the way they handle conditionals in
880 general, and presupposition projection in conditionals in particular.

881 **DRT Predictions** Operators such as conditionals introduce new embedded DRSs into
882 the structure in DRT. Depending on where the presupposition is introduced and where a suitable
883 antecedent can be found, different projection path lengths come about. Conditionals introduce
884 two sub-DRSs, connected by an arrow. For presuppositions introduced in the consequent, the first
885 place to look for support for the presupposition is the antecedent of the conditional, followed by
886 higher levels. Let us take a look at the DRSs for our four conditions. In condition (a), illustrated
887 in Figure 2, the presupposition that there is an earlier time where Tina did not go ice-skating is
888 introduced in the top-level box of the consequent and satisfied locally in the antecedent of the
889 conditional.[20] Resolution of the presupposition then only involves one step, yielding a projection
890 path length of 1. In condition (b), illustrated in Figure 3, on the other hand, the presupposition
891 that Tina did go ice-skating before is introduced inside of the negated box in the consequent and
892 resolved at the top-most level, based on the information introduced through the context sentence.
893 Given standard DRT-assumptions about the path followed in searching for an antecedent, this
894 involves three steps. Turning to condition (c) (see Figure 4), we again have the presupposition of
895 the *wieder nicht* order, as in (a), but this time, given Context II, it is resolved by the context
896 sentence at the top-most level, thus rendering a projection path length of 2. Finally, condition (d)
897 (see Figure 5) has the same presupposition as (b), introduced in the scope of negation inside of
898 the consequent, but this time it is resolved locally in the antecedent of the conditional. This also
899 yields a projection path length of 2.

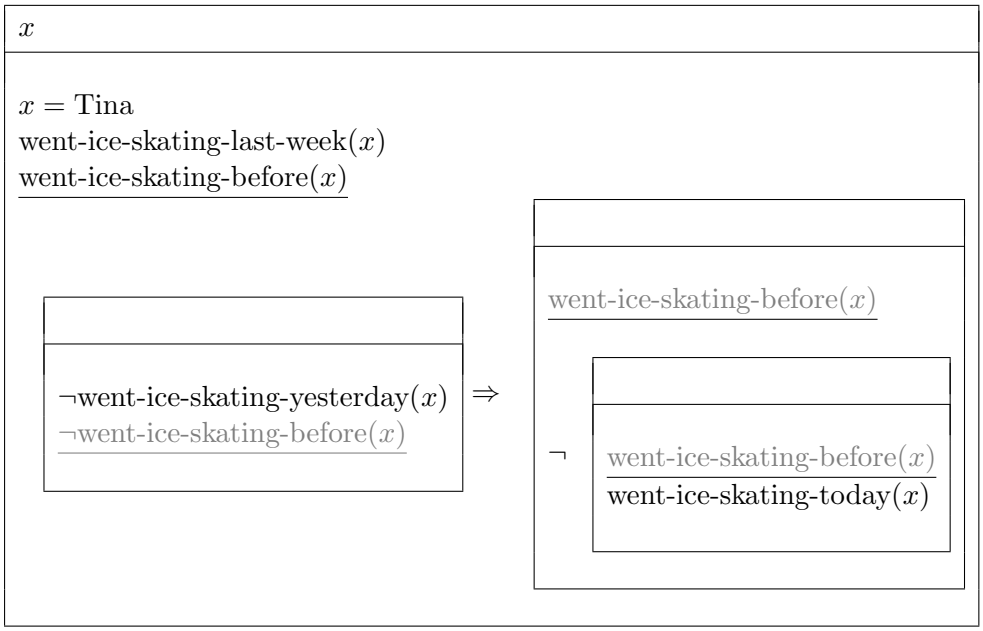
900 Looking at the distribution of the projection path lengths, we see that an interesting

[20]Note that there is a crucial difference between this type of local resolution on the hierarchically closest level and local accommodation in the position where the presupposition was originally introduce; see the general discussion for further details.



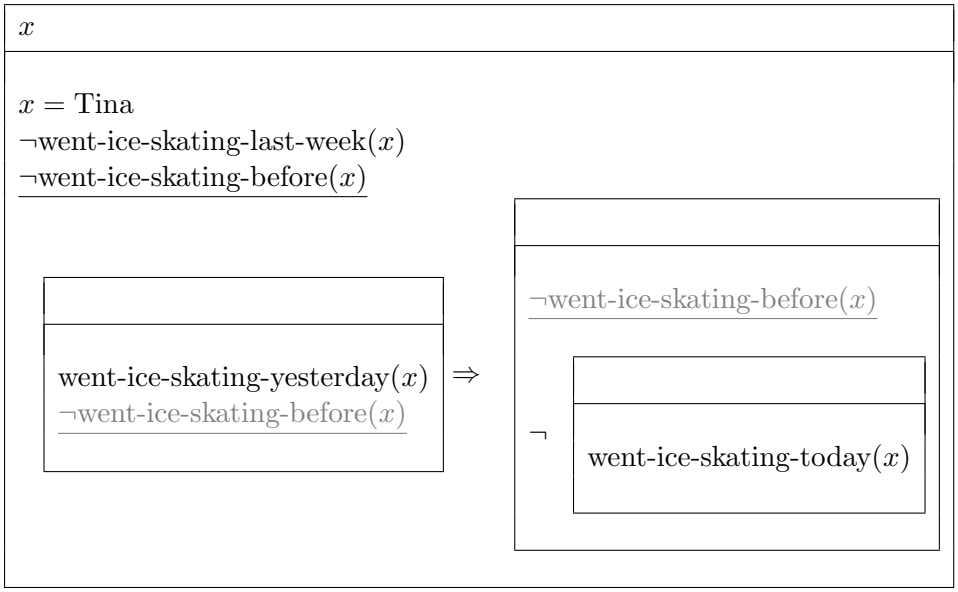
Projection Path Length = 1

Figure 2. **Condition a:** Context: I, Location: local, Firstword: *wieder*



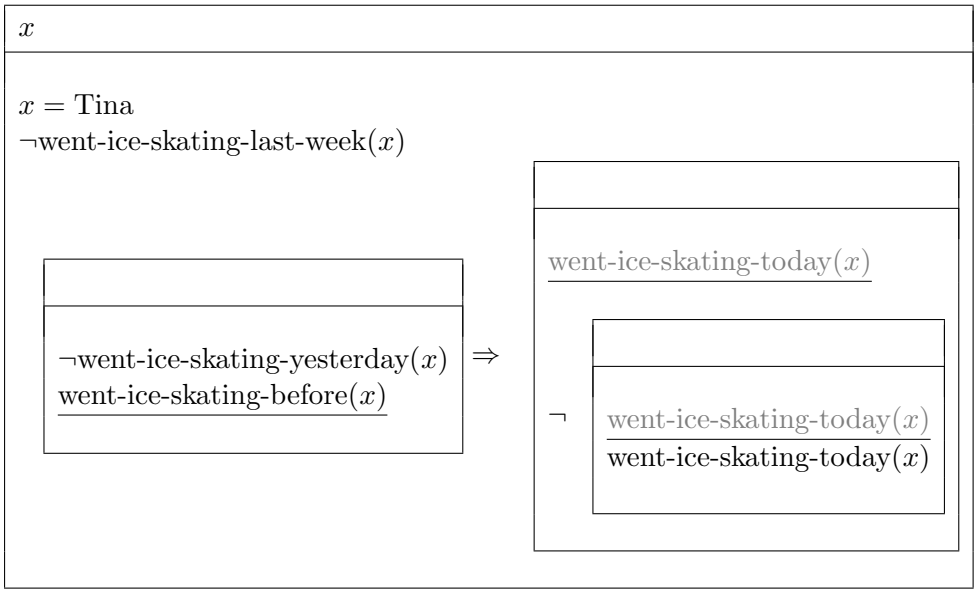
Projection Path Length = 3

Figure 3. **Condition b:** Context: I, Location: global, Firstword *nicht*



Projection Path Length = 2

Figure 4. **Condition c:** Context: II, Location: global, Firstword: *wieder*



Projection Path Length = 2

Figure 5. **Condition d**: Context: II, Location: local, Firstword: *nicht*

pattern emerges. For Context I, we get a difference based on the location where support for the presupposition is introduced (corresponding to a difference between the *nicht wieder* and *wieder nicht* conditions), with a longer projection path for the global condition (3) than for the local condition (1). Context II, in contrast, does not give rise to such a difference, as both the global and local conditions (corresponding to *wieder nicht* and *nicht wieder* respectively) have a projection path length of 2. This comes about because of the additional embedding introduced by negation. For the *nicht wieder* condition, this results in two steps, first out of the negation box into the main consequent box, then to the box for the antecedent of the conditional. For the *wieder nicht* condition, on the other hand, there also are two steps, but different ones: from starting out in the top-most box of the consequent, there's one step to the antecedent, and another to the global level. Hypothesizing that longer projection path lengths correspond to cognitive efforts that are reflected in increased reading times, this projection path distribution of the DRT account leads us to expect an interaction between **Context** and **Location**, with a difference between the (a) and (b) conditions, but no difference between (c) and (d), as illustrated in the left panel of Figure 6. (a main effect of **Location** may also arise based on these predictions, though it would be dominated by the interaction).

Dynamic Semantics Predictions A simple view of conditionals in Dynamic Semantics assumes the equivalent of logical implication as their meaning contribution, i.e., *if p, then q* is assumed to be true unless p is true and q is false. To do this dynamically, and to account for presupposition projection, this is done in a slightly more round-about way, namely by removing those worlds from c where p is true and q is false. The update procedure does this by first removing the combination of c, p, and q from the combination of c and p, and then removing the resulting set from the initial context c, as can be seen in the formal formulation of the context change potential of a conditional in 16a:

(16) a. $c+$ If $p, q = c - ((c + p) - ((c + p) + q))$

b. defined iff $(c + p) + PSP_q = (c + p)$

As far as presuppositions are concerned, and specifically the ones introduced in the consequent as in our case, what is important is that as in the case of negation, the

presuppositions of q have to be met in the context that it is added to, but in this case, this context turns out to be the combination of $c + p$. That is, the context update for a conditional can only be performed if the global context updated with the antecedent of the conditional entails the presuppositions introduced in the consequent of the conditional. This means, however, that it should not make a difference whether the presupposition is supported by the global context (the preceding sentence) or the local context (here: antecedent of the conditional). The fact that this holds indeed is a crucial piece of the Dynamic Semantic account of presupposition projection, as this allows presuppositional support to be introduced a) in the global context, b) in the antecedent of the conditional, or c) by both combined. As far as our experimental manipulation is concerned, however, it also means that on the basis of the semantics alone, we can not differentiate between these different possible sources of support, and thus cannot come up with different processing predictions w.r.t. relative processing effort involved in these cases.

If pressed to squeeze some kind of potential prediction out of the dynamic account, the most plausible factor at play might be negation, e.g., by hypothesizing that a negated presupposition is harder to process than an unnegated presupposition, both on intuitive grounds, and based on the more complicated update steps involved. If we take r in (17) to be the unnegated version of the presupposition in our materials, the underlying mechanisms to check the presuppositions of the *again not* and *not again* variants would involve (18a) and (18b) respectively.

(17) r = Tina had been ice-skating before

(18) a. $PSP_q = \neg r$: $c + \neg r = c?$ $\approx$ $c - (c + r) = c?$

b. $PSP_q = r$: $c + r = c?$

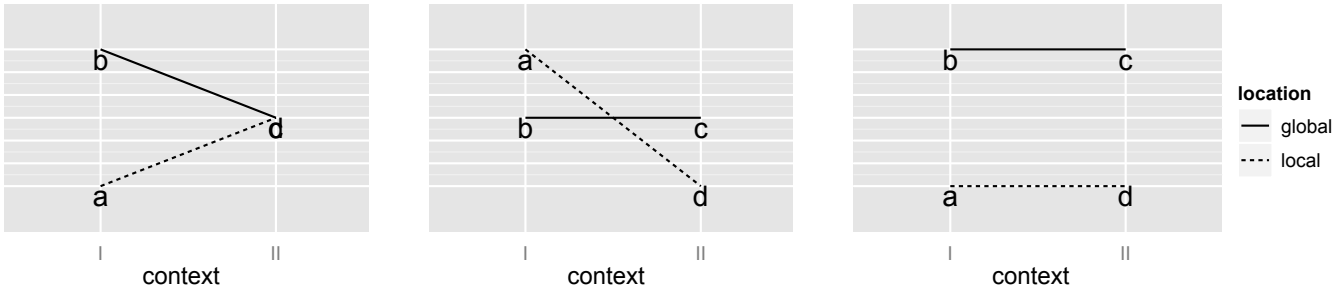
Hypothesizing that negation adds extra complexity would then suggest that the *again not* sentences should be harder to process than the *not again* sentences, based on the presupposition evaluation process. However, the point about the evaluation of negation requiring extra resources presumably should be applied to the processing of non-presupposed content as well. In our case, this would suggest that negation in the antecedent is relevant as well, which would increase the the relative difficulty of the sentences in Context I, but not Context II, as illustrated in the update process involved in assessing the definedness of the overall context update for our four conditions below.

		Context I		Context II
(19)	a)	$(c' + \neg q) + \neg \mathbf{r} = c' + \neg \mathbf{q}$	c)	$(c' + q) + \neg \mathbf{r} = c' + q$
	b)	$(c' + \neg q) \ + \ r = c' + \neg \mathbf{q}$	d)	$(c' + q) \ + \ r = c' + q$

Factoring negation into the equation along these lines, a dynamic account thus also gives rise to the prediction of a **Location**×**Context** interaction. However, this interaction is crucially different from the one predicted by a DRT account in that the expected simple effect of **Location** in Context I is in the opposite direction and a simple effect of **Location** is predicted in Context II (see middle panel of Figure 6).

A Linear Distance Hypothesis However, there is at least one further perspective on the experimental materials that one could consider, namely that what matters is the linear distance between the presupposition trigger and the expression(s) supporting the presupposition. In particular, it could plausibly be assumed that an increase in distance corresponds to an increase in processing cost. This could be motivated in various ways, e.g., by alluding to distance effects for anaphora resolution, since *again* commonly is seen as an 'anaphoric' presupposition trigger. Alternatively, from the perspective of accounts based on situation models, one could allude to the difference between the current model and the integrated model, which are stored in different levels of working memory (short term vs. long term). What we expect to matter, on such a view, then, is simply whether the antecedent for the presupposition is introduced in the global context sentence or in the local antecedent of the conditional. This gives rise to a straightforward processing prediction, namely a main effect of **Location**, with no differences based on **Context** (or **Firstword**). This prediction is illustrated on the right side of Figure 6.[21]

[21] As a reviewer points out, it's also possible to consider a distance measure in terms of number of intervening words, which would be slightly more differentiated in that the order of *nicht* and *wieder* would end up making a difference in distance. However, it's not clear to what theoretical motivation there would be for such a view. In any case, such a hypothesis could be straightforwardly investigated by adding further intervening materials without introducing the clausal complexity. We leave such investigations to future research.



DRT Dynamic Semantics (Negation) Processing Hypothesis

Figure 6. Processing Predictions

977 *Results*

978 As in Experiment 1a, the primary focus in our analysis were the reading times on the verb
979 following the {*wieder nicht*} sequence in the consequent of the conditional, since this is the point
980 at which the presupposition becomes fully explicit, but we also analyzed reading times on the
981 {*wieder nicht*} sequence itself. The same standard reading measures as in Experiment 1a were
982 calculated, and the analyses proceeded as before, using linear mixed effect models with the
983 maximal random effect structure that would converge, and model comparisons to assess the
984 impact of a given factor on the overall fit of the model. Significant effects were found for regression
985 path duration and total reading times. Re-reading times also gave rise to some significant effects,
986 but the number of data points was very small for these, so we focus on the former two in the
987 presentation of our results. Trials where any blinks occurred while looking at the region of
988 interest were removed prior to analysis. Data points that deviated by more than three standard
989 deviations from the mean of their condition were excluded from the analysis as well.

990 The mean reading times for the two measures to be discussed here are summarized in Table
991 7, and the general pattern of results is illustrated in Figure 7. The main set of statistical analyses
992 were carried out using **Location** and **Context** as factors to assess the predictions of the different
993 accounts laid out above. The central result is that there is a significant difference between
994 conditions a and b, but no reliable difference between c and d, i.e., the local vs. global distinction

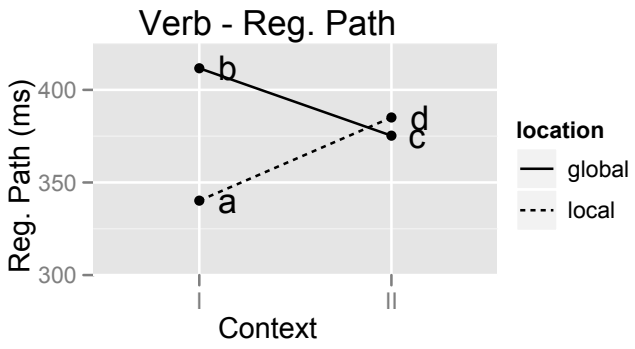


Figure 7. Summary of Regression Path Duration for the verb (in ms)

mattered in Context I (where *wieder nicht* was paired with local and *nicht wieder* with global), but not in Context II (with the reverse pairing). Statistically, this was supported by a significant interaction between context and location.

For Regression Path Duration, the interaction between Context and Location significantly improved overall model fit (RES-2: $\beta = 78.39$, $SE = 35.92$, $t = 2.18$; $\chi^2 = 4.6112$, $p < .05$). The same held for total reading time (RES-4: $\beta = 85.11$, $SE = 35.7$, $t = 2.38$; $\chi^2 = 5.65$, $p < .05$). For the latter, there also was a significant main effect of Location (RES-2: $\beta = 53.86$, $SE = 19.18$, $t = 2.81$; $\chi^2 = 6.67$, $p < .01$), though note that this was dominated by the cross-over interaction. For regression path duration, this effect approached, but did not reach, marginal significance (RES-2: $\beta = 31.96$, $SE = 19.35$, $t = 1.65$; $\chi^2 = 2.69$, $p < .11$). The interactions were driven by the difference between conditions (a) and (b), both for regression path duration (RES-1: $\beta = 70.94$, $SE = 28.97$, $t = 2.45$; $\chi^2 = 5.69$, $p < .05$) and total duration (RES-2: $\beta = 86.01$, $SE = 37.78$, $t = 2.28$; $\chi^2 = 11.29$, $p < .001$). No significant differences emerged between conditions (c) and (d).

Comparable results emerged for these two reading time measures when looking at the {*wieder nicht*} region: there was an interaction for both regression path duration (RES-3: $\beta = 71.06$, $SE = 31.19$, $t = 2.28$; $\chi^2 = 5.18$, $p < .05$) and total reading time (RES-3: $\beta = 77.46$, $SE = 33.80$, $t = 2.29$; $\chi^2 = 5.23$, $p < .05$), as well as a main effect of Location for the latter

		Context I		Context II	
	firstword	*wieder*	*nicht*	*wieder*	*nicht*
	Location	local	global	global	local
	Condition	a	b	c	d
verb	Reg. Duration	340	411	375	385
	Total	396	493	432	427
{*wieder nicht*}	Reg. Duration	341	413	381	389
	Total	400	489	435	429

Table 7
Reading times on Verb and wieder nicht sequence (in ms)

(RES-2: $\beta = 50.27$, $SE = 19.39$, $t = 2.59$; $\chi^2 = 5.91$, $p < .05$) and a marginal one for the former (RES-3: $\beta = 32.00$, $SE = 16.24$, $t = 1.97$; $\chi^2 = 3.79$, $p < .1$). As with the verb-region, the interaction was driven by simple effects between conditions (a) and (b), both for regression path duration (RES-1: $\beta = 68.09$, $SE = 23.87$, $t = 2.85$; $\chi^2 = 7.79$, $p < .01$) and total duration (RES-1: $\beta = 88.94$, $SE = 24.57$, $t = 3.62$; $\chi^2 = 12.52$, $p < .001$).

Given the nature of our factors, we conducted statistical analyses for several other pairwise comparisons. Within the global conditions, we found a significant difference for total times both on the {*nicht wieder*} (RES-3: $\beta = 52.68$, $SE = 24.22$, $t = 2.18$; $\chi^2 = 4.73$, $p < .05$) and verb regions (RES-2: $\beta = 61.62$, $SE = 26.81$, $t = 2.30$; $\chi^2 = 5.189$, $p < .05$). A marginally significant difference in regression path duration was found between the local conditions on the {*nicht wieder*} region (RES-1: $\beta = 40.43$, $SE = 21.90$, $t = 1.85$; $\chi^2 = 3.41$, $p < .1$). Furthermore, there were differences between the two *nicht wieder* conditions for total duration, both for the *nicht wieder* region (RES-2: $\beta = 65$, $SE = 25.27$, $t = 2.57$; $\chi^2 = 61.6$, $p < .05$) and for the verb region (RES-4: $\beta = 72.45$, $SE = 25.80$, $t = 2.81$; $\chi^2 = 7.66$, $p < .01$). Finally, the *wieder nicht* conditions differed marginally from one another on the *wieder nicht* region for regression path duration (RES-1: $\beta = 36.99$, $SE = 22.13$, $t = 1.67$; $\chi^2 = 2.79$, $p < .1$).

Altogether, the pattern of the results very closely resembles the predictions of the DRT account as laid out above. The greatest and clearest difference in reading times is found for the

two conditions with the greatest difference in projection path length, namely (a) and (b), whereas no difference is found between (c) and (d), counter to what we would expect based on the processing hypothesis. Furthermore, for total reading times, we also find differences based on the level of the **Location** factor for both *wieder nicht* and *nicht wieder*, with longer reading times in the global conditions. There also are some significant differences between the local and global contexts for the *nicht wieder* condition, as well as a marginal one for the *wieder nicht* condition. These results correspond surprisingly well to the pattern expected based on DRT projection path length. Based on this observation, we conducted an additional follow-up analysis, in which we tried to model the data using only projection path length as a (numerical) predictor. Both for regression path duration and total time, this yielded a significant effect of projection path length ($t's > 2.8$). Based on model comparisons with the interaction analysis above, the resulting fit was not significantly different for the two analyses. In other words, projection path length alone was as good a predictor of reading times as the **Location**×**Context** interaction.

Discussion

The results for total reading time and regression path duration mirror the predictions of the DRT hypothesis based on projection path length remarkably well, and indeed the data can be modeled adequately by just considering projection path length as a predictor.[22] This provides strong support for our proposed interpretation of the results from the first set of experiments, based on the idea that presupposition projection involves representational complexity of some kind, as on the DRT account, and that representational complexity has real cognitive correlates. As in the first set of experiments, a Dynamic Semantic account either doesn't predict any differences at all, or it predicts an interaction of a very different nature, if we assume that negation introduces processing complexity. Similarly, the predictions of an account based on situation models (or other accounts motivating an effect of linear distance) are not borne out in the data either, as they only involve a main effect of location. However, whether presuppositional

[22]It's worth noting that there's a difference in terms of what reading time measures exhibit the effects in the two eye tracking experiments. Our tentative explanation for this is that the types of sentences in experiment 2 are somewhat more complex, and that we are only dealing with felicitous texts.

support was introduced at the level of the global context sentence or in the antecedent of the conditional did not matter if one kept the overall projection path length constant by switching from the *wieder nicht* (c) condition to the *nicht wieder* condition (d). What rather seems to matter in terms of processing effort is the length of the projection path overall, not the absolute location at which support for a presupposition is introduced.

General Discussion - Part II

Summary

In the four experiments reported here, we have considered the impact of embedding on the online interpretation of the presupposition of *wieder*. In Experiment 1a, we found immediate reflexes in reading time measures when unembedded *wieder* did not match the context, but no corresponding delays were found for embedded *wieder*. Our proposed explanation for this was that in the latter case, the global presupposition is not immediately available. The results from the rating task in Experiment 1b confirmed that participants generally do perceive the infelicity for both the *wieder nicht* and *nicht wieder* versions in the respective contexts, and the rate of stops-making-sense judgments in Experiment 1c furthermore was comparable for both types of conditions. It thus does not seem possible to explain the reading time results in terms of a general availability of local accommodation. Tying together these judgments with the time-course of interpretation, we also looked at the rejection times in the stops-making-sense task. Consistent with the hypothesis that global presuppositions of embedded presupposition triggers only become available with a delay compared to unembedded ones, we found longer rejection times for the *nicht wieder* condition. Looking at word-by-word reading times in the felicitous conditions, we also found tentative evidence for slow-downs in the *nicht wieder* condition, even when the presupposition was supported by the context. Experiment 2 attempted to broaden the evidence for our interpretation of the first set of studies by looking at additional embeddings, and further variation in the location of support of the presupposition. The results neatly match the predictions of a representational account like DRT, which fits naturally with a correspondence between processing effort and the level of representational steps that have to be taken to arrive at the final interpretation.

Another aspect of the reading time results from Experiment 1a was that for unembedded *wieder*, we observed slow-downs due to infelicity as soon as all the content going into the presupposition is made explicit. In particular, we found increases in first fixation duration on the verb. Clear experimental support at a detailed temporal level for the immediate availability of (unembedded) presuppositions constitutes an important advance of the empirical picture on the processing of different aspects of meaning. Furthermore, note that the delays for embedded *wieder* are highly relevant for the evaluation of pragmatic accounts, as these generally assume all occurrences of presupposition triggers - embedded or unembedded - to involve the same pragmatic reasoning processes. They therefore do not have a general explanation for differences between these contexts at the ready. This is particularly so for accounts such as that proposed by Schlenker (2009), as it is entirely on par with Dynamic Semantics with regard to the inability to differentiate between locations of support for presuppositions, since it evaluates presuppositions relative to all preceding text, without differentiating between, say, the (local) antecedent of a conditional and preceding sentences.

While we have highlighted the compatibility of a representational account such as DRT with our results, it may of course be possible to come up with another theoretical account of presupposition projection, paired with suitable processing assumptions, that result in similar predictions. In that sense, we by no means claim to have presented final evidence for DRT, and we also duly note that DRT may very well face its own problems on the theoretical side. However, with respect to the crucial integration of representational levels that allow us to differentiate between different locations of support for presuppositions in the way spelled out above, DRT very much seems to be on the mark. We are not aware of any such alternative proposals, and whatever they might turn out to be, they would need to model exactly these properties to fit with the patterns in our data (in addition to the standard set of projection phenomena discussed in the theoretical literature). The current paper is not the place to explore such alternative possibilities in any depth, but we hope to at least provide a new starting point for theoretical discussions that incorporate more fine-grained empirical evidence. In particular, an integration of appropriate formal semantic theories with more general processing models of language comprehension in context seems to be an order. The upshot of our discussion is that such an integration should include structured representations at a level such as the text-base that go beyond what has so far

been assumed in the relevant literature in that they require a hierarchical representation of discourse, combined with an account of how presuppositions relate to that level of discourse structure.

Potential Alternative Interpretations and Related Results from The Literature

In this section, we consider some potential alternative explanations and related results from the literature that might seem to conflict with our findings. We begin with an alternative interpretation of the eye tracking results from Expt 1a, suggested to us by Judith Degen (p.c.): it seems to be the case that the form *nicht wieder* is more frequent than *wieder nicht*.[23] If higher frequency forms give rise to decreased reading times, which furthermore might extend to a following word (in our case the verb), then a possible explanation of the lack of felicity effects in reading times in the *nicht wieder* condition could be based on a floor effect: Fixations can only be so short, and a strong frequency effect might bring the infelicitous condition down to a minimum, erasing any further advantage based on felicity. While such an account is entirely consistent with the basic results from Expt 1a, we do not believe that the experimental data on the whole are consistent with it.

The floor-effect account crucially rests on the assumption that *nicht wieder (+Verb)* is easier and faster to process than *wieder nicht (+Verb)*, which makes a straightforward prediction for what should happen in the absence of infelicity: reading times for the latter should be greater than for the former. This prediction is not borne out in our data. In Experiment 1a, total reading times for the felicitous *nicht wieder* condition are greater than for the *wieder nicht* one. The same holds for combined self-paced reading times on {*nicht wieder*} and {*nicht wieder*} *+Verb* in the felicitous conditions of Experiment 1c. Similarly, there were simple effects based on the order of *nicht* and *wieder* for total times and regression path duration on {*wieder nicht*} and the verb in Experiment 2, both in the local and global contexts. Relatedly, also note that the ratings for the

[23]According to a quick search in the Guttenberg corpus performed by Degen, the sequence 'NOUN nicht wieder' yields 1395 occurrences, as opposed to 95 for the sequence ''NOUN wieder nicht'. Google searches for the bigrams 'nicht wieder' and 'wieder nicht' yield about twice as many hits for the former. While none of these measures may be perfect, the general picture of a frequency asymmetry seems relatively clear.

nicht wieder conditions in Experiment 1b were lower across the board, which could be seen as a further indication of increased complexity and corresponding processing difficulty. Finally, a floor-effect account of the results from Experiment 1a would have to explain why there is no comparable floor-effect in Experiment 1c, where stops-making-sense judgments in the *nicht wieder* condition were slower than in the *wieder nicht* condition. Apart from these aspects of the overall data that pose empirical challenges to a floor-effect account, it is not clear whether the conceptual motivation of such an account survives closer scrutiny. In particular, the effect that we do find in the unembedded condition of Experiment 1a is due to the inconsistency of the presupposition with the preceding context. We have interpreted this as an indication that the presupposition is immediately available and evaluated relative to the context. If the *nicht wieder* condition were easier to process due to its higher frequency, as posited by the floor-effect account, then the relevant presupposition should be available even more easily and quickly. But then the inconsistency should be immediately detected, as in the unembedded case, and give rise to the same perception of infelicity, with a corresponding slow-down in this condition. The trade-off between a speed-up due to frequency and a lack of increase due to infelicity seems like a peculiar one, since the infelicity should be detectable even more easily and quickly in the variant that is easier to process. Be this as it may, given the relative slowness of the *nicht wieder* felicitous conditions, we do not find this alternative account convincing when considering the full set of data from our experiments.

There is a potential further alternative explanation for the eye tracking results from Experiment 1a based on order effects, pointed out by Jesse Snedeker (p.c.) at CUNY 2012, which also is not supported by the data. It might be the case (perhaps also due to the frequency asymmetry) that participants get used to dealing with the infelicity more quickly in the *nicht wieder* condition. The reading times in the infelicitous conditions then would be expected to start out at comparable levels on the first couple of trials, but decrease more quickly for the *nicht wieder* condition. Including order as a factor in a linear regression analysis does not support this possibility. While the only significant effect is the Firstword*Felicity interaction, reading times in the infelicitous *nicht wieder* condition remain at a stable low level throughout, whereas those for

the infelicitous *wieder nicht* condition decrease numerically over the course of the experiment.[24] From the perspective of our interpretation of the data, this suggests that participants initially stumble badly over the infelicity in the unembedded condition, but get somewhat used to it over the course of the experiment.

We also would like to consider a potential alternative approach to the results from Experiment 2.[25] If we provided an alternative explanation of why *wieder nicht* has advantages over *nicht wieder* that is independent of presupposition projection per se, and combined that with the linear distance hypothesis above, we would end up with a close to equivalent set of predictions. While we're not in a position to fully spell out such an alternative explanation, consideration of its general logic may nonetheless be valuable. This sort of combined processing account would predict a comparable interaction of **Context** and **Location**, since the two conditions for which we found large reading time differences involve both a switch from *nicht wieder* to *wieder nicht* and from a local to a global context. So if we looked at the data from the perspective of the factors of **Firstword** and **Location**, we might simply see additive main effects of these factors. However, we don't think that the evidence overall provides compelling support for this view. In particular, the results from Experiment 1a do not seem to fit with it. If *nicht wieder* in general were at a disadvantage with respect to *wieder nicht*, we would expect corresponding differences to come into play when comparing the infelicitous conditions in Experiment 1a to one another. Depending on whether the alternative account assumed a general effect of infelicity or just for the *wieder nicht* conditions, it would lead us to expect that the infelicitous *nicht wieder* condition would be even worse than the infelicitous *wieder nicht* condition, or at least roughly as bad (assuming the two factors operate on a comparable level in terms of the magnitude of their effects). Furthermore, under the former assumption, we'd expect

[24] An anonymous reviewer points out that this pattern might lend further support to the floor-effect account, with a decrease seen over time in the less frequent form. However, note that in the felicitous conditions both orders are roughly on par and more or less flat. This is unexpected from the floor-effect perspective, which would lead one to expect a speed-up for the less frequent *wieder nicht* case in the felicitous context as well.

[25] Thanks to Colin Phillips and Jeff Lidz for discussion of this point.

an effect of infelicity for *nicht wieder* after all. This prediction is not borne out in the results from Experiment 1a above. While a more detailed evaluation of this alternative account may be warranted, in particular if a fully motivated theoretical proposal for a disadvantage of *nicht wieder* over *wieder nicht* can be spelled out, we thus do not see a strong case in favor of it at present.[26]

As a final point, we'd like to put our results in perspective with another recent study that looks at local accommodation vs. global interpretation of presuppositions:[27] Chemla and Bott (2013) report results from a reaction time study involving truth-value judgments, where 'false'-judgments correspond to global interpretations and 'true'-judgments to local accommodation, which show the former to be faster than the latter. While this may initially seem as if it were the opposite of our findings, we have to take into account the differences between the studies to appreciate that the results may indeed be entirely compatible. Chemla and Bott used factive verbs such as *realize* to construct sentences such as (20) in the context of a story about alien zoologists:

(20) Zoologists do not realize that elephants are birds.

The verb *realize* introduces a factive presupposition to the effect that its complement clause is true. If this presupposition is interpreted globally relative to the matrix negation, the sentence will naturally be judged to be false. However, if interpreted locally, the sentence essentially conveys that it is false that elephants are birds and zoologists believe so. Such a negated conjunction is, of course, true, since the first conjunct on its own is obviously false. Chemla and Bott compare semantic and pragmatic accounts of presuppositions (namely a dynamic account a lá Heim 1983 as compared to a pragmatic account along the lines of Schlenker, 2008a) and argue that the former predict global interpretations to be faster, while the latter predict the reverse. Their finding of slower response times for 'true' answers then seem to support semantic accounts. How does this relate to our results that suggest that longer projection paths lead to increased processing efforts? It fits into the larger picture if we assume that in the Chemla and Bott study,

[26] One additional point to note here is that a general advantage of *wieder nicht* over *nicht wieder*, whatever it may consist of, would help to explain the main effect of **firstword** in Experiment 1b.

[27] See also Romoli and Schwarz (submitted) for comparable results from a slightly different paradigm.

the local responses are based on local accommodation (where the presupposition is interpreted in the very position it was introduced in syntactically). From the perspective of DRT (as well as many other accounts), accommodation is a last resort repair mechanism. More specifically, DRT accounts of presupposition assume that there is a two-phase procedure involved in presupposition evaluation: first, the projection path is followed upwards in the discourse structure in search of an antecedent. If no antecedent is found, global accommodation is attempted, but if this leads to inconsistency, one retreats back down the search path in search for a suitable location for accommodation. On this view, local accommodation is precisely the last stage in this procedure, and it thus is entirely consistent with such an account that local accommodation responses are slower than global ones.[28] In light of this, one might wonder why local accommodation is not more prevalent for our *nicht wieder* sentences in Experiments 1a-c. However, it is quite commonly known that different triggers display varying potential for accommodation in general and local accommodation in particular, so this difference may entirely be due to the fact that *again* is a so-called hard trigger, while factive verbs are soft triggers in the terminology of Abusch (2010). In closing this brief comparison, note that while Chemla and Bott take their results to support Dynamic Semantics, the DRT style analysis we have considered here seems entirely on par with Dynamic Semantics in the relevant respects and thus is equally compatible with their results.

Conclusion

As the experimental exploration of presupposition interpretation in online processing is still in its beginnings, many open questions naturally remain. With respect to the issues discussed

[28] A reviewer raises the worry that such steps in the construction of abstract hierarchical structures posited theoretically should not be mapped onto any temporal effects in processing, because the latter is necessarily linear. While linear processing may be all that is involved in immediate incremental interpretation, it would seem too strong to us to claim that hierarchical linguistic structures have no role in processing at all. In fact, the central conclusion we draw from our results is that a level of hierarchical discourse structure is relevant in comprehension in context. Since many of our projection effects only show up in 'late' processing measures (such as total time), these hierarchical structures may not be present immediately, but they crucially do seem to come into play later on and are central to the integration with the larger discourse context.

here, one particularly interesting point is whether different types of presupposition triggers vary in the way that projection proceeds. In particular, considering the distinction between soft and hard triggers just mentioned, would the former display the same types of effects as we found for *again*? The property of triggers being anaphoric seems equally relevant. Standard DRT accounts see all triggers as anaphoric, but there seem to be good grounds for seeing a difference between triggers in this regard (e.g., comparing *again* to *stop* or *realize*), though the diagnosis of such differences is by no means straightforward and has to be considered carefully for each individual case (Heim, 1990, for example, discusses potential evidence for anaphoric aspects of *stop*). Furthermore, given the parallels that DRT posits between pronouns and presuppositions, the question arises of how to relate the present findings to the extensive literature on pronoun interpretation in processing. Future work will need to assess these issues in further detail. As we've tried to stress throughout, an integration of the relevant formal semantic theories with more comprehensive models of discourse processing is generally called for. What we hope to have shown is that the insights of such formal theories in general, and their specific assumptions about the processes involved in relating presuppositions to preceding linguistic material in particular, have an important role to play in such general models, in particular with respect to the levels of structure that play a role in processing.

References

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy*, *34*(6), 491–535.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics*, *27*(1), 37 –80.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D. M. (2005). Fitting linear mixed models in r. *R News*, *5*, 27-30.

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics*. Stanford, CA: CSLI Publications.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437–457.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 443–467.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Chemla, E., & Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, *38*(3), 241-260.

Chemla, E., & Schlenker, P. (2012). Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics*, *20*(2), 177–226.

Chierchia, G. (1995). *Dynamics of meaning. anaphora, presupposition and the theory of grammar*. Chicago and London: University of Chicago Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path. the use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (p. 320-358). Cambridge: Cambridge University Press.

Geurts, B. (1999). *Presuppositions and pronouns.* Amsterdam; New York: Elsevier.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts.* New York: Academic Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Groenendijk, J., & Stokhof, M. (1990). Dynamic montague grammar. In L. Kálmán & L. Pólos (Eds.), *Papers from the second symposium on logic and language* (p. 3-48). Budapest: Akadémiai Kiadó.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* Unpublished doctoral dissertation, UMass.

Heim, I. (1983). On the projection problem for presuppositions. *Proceedings of the Second West Coast Conference on Formal Linguistics*, 114-125.

Heim, I. (Ed.). (1990). *Presupposition projection.* University of Nijmegen.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language: Proceedings of the third amsterdam colloquium* (Vol. I, p. 227-321). Amsterdam: Mathematical Center.

Kamp, H., & Reyle, U. (1993). *From discourse to logic.* Dordrecht: Kluwer Academic Publishers.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, *4*(2), 169-193.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Lewis, D. (1979). Scorekeepingin a language game. *Semantics from Different Points of View*.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological*, *124*, 372-422.

Romoli, J., Khan, M., Snedeker, J., & Sudo, Y. (submitted). Resolving temporary referential ambiguity using presupposed content. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions.* Springer.

Romoli, J., & Schwarz, F. (submitted). An experimental comparison between presuppositions and indirect scalar implicatures. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions.* Springer.

Romoli, J., Sudo, Y., & Snedeker, J. (2011). An experimental investigation of presupposition projection in conditional sentences. *Proceedings of SALT*, *21*, 592–608.

Schlenker, P. (2008a). Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, *34*(3), 157–212.

Schlenker, P. (2008b). Presupposition projection: The new debate. In T. F. 2008 & S. Ito (Eds.), *Proceedings of SALT 18.* Ithaca, NY: CLC Publications.

Schlenker, P. (2009). Local contexts. *Semantics & Pragmatics*, *3*, 1-78.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics*, *24*(4), 373.

Schwarz, F. (submitted). Presuppositions vs. asserted content in online processing. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions.* Springer.

Simons, M. (2001). Simons, m.: On the conversational basis of some presuppositions. In *Proceedings of semantics and linguistics theory 11.*

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, *2*(4), 447-457.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. K. Milton & P. K. Unger (Eds.), *Semantics and philosophy.* New York: New York University Press.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (Vol. 9, p. 315-322). New York: Academic Press.

Strawson, P. (1950). On referring. *Mind*, *59*, 320–344.

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*(5217), 1632–1634.

Tiemann, S. (2013). *The processing of wieder ('again') and other presupposition triggers.* Unpublished doctoral dissertation, University of Tuebingen.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., ... Beck, S. (2011).

Psycholinguistic evidence for presuppositions: On-line and off-line data. In I. Reich, E. Horch, & D. Pauly (Eds.), *Proceedings of sinn & bedeutung 15* (p. 581-597).

van der Sandt, R. (1992). Presupposition projection as anaphora. *Journal of Semantics*, *9*(333-377).

van der Sandt, R., & Geurts, B. (1991). Presupposition, anaphora, and lexical content. In *Text understanding in lilog* (p. 259-296).

Zwaan, R., & Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.

Author Note

To be filled in