Licensing Question-Sensitive Discourse Particles: Evidence from Grammaticality Judgments, Self-Paced Reading and EEG Studies

Anna Czypionka, Mariya Kharaman,
Josef Bayer, Maribel Romero & Carsten Eulitz
University of Konstanz
anna.czypionka@uni-konstanz.de

1 Introduction

Discourse particles (DiPs) in German are small words like *schon* ('already'), *bloß* ('only'), or *denn* ('then'), which are routinely used in daily conversation. These words modify the illocutionary force of an utterance. Many of them also have other functions in the grammar, e.g. as focus particles or as adverbs. DiPs are sensitive to clause type, so that some DiPs can only occur in declaratives and others only in interrogatives or imperatives. Question-sensitive DiPs (QDiPs in the following) are at the center of the research presented here. While their syntactic and semantic licensing constraints have been studied in depth in theoretical linguistics, so far only little is known about their role in language processing. In the following, we will present an overview of their syntactic and semantic licensing constraints. Based on this theoretical background, we will formulate the research questions guiding our experimental investigations of QDIP licensing in sentence comprehension. The overarching goal of the experiments presented here is to characterize the role of QDiP licensing in sentence comprehension, to assess the relative severity of different kinds of QDiP licensing violations, and to identify electroencephalographic (EEG) correlates for QDiP licensing violations and successful QDiP licensing.

1.1 Background

QDiPs like *denn* reshape the illocutionary force of an utterance, adding a pragmatic dimension to the question. By inserting *denn* into an interrogative clause, the question is linked to the previous discourse context (König 1977; Thurmair 1991, a.o., and Gutzmann 2015; Csipak & Zobel 2016; Theiler 2020 for different formalizations). This makes *denn* pragmatically infelicitous in out-of-context questions (König, 1977; Thurmair, 1991; Wegener, 2002; Grosz, 2005; Bayer, 2012).

As a question-sensitive discourse particle, *denn* can only occur in interrogative clauses like *wh*-questions (1) or polar questions, but not in declaratives (2). Importantly, *denn* in *wh*-

questions in not simply licensed by an occurring *wh*-element but by the standard *wh*-initial form of an interrogative. As shown by (3), *wh*-in-situ / echo questions etc. fail as proper licensers.

(1) *Was hast du denn gesagt?*
what have you QDiP said
'What was it that you said?' / 'What did you say?' (I'm wondering)

(2) *Peter hat das (*denn) gesagt.*
Peter has that QDiP said
'Peter said that.'

(3) *Du hast was (*denn) gesagt?*
You have what QDiP said
'You said WHAT?' (I can't believe you said that!)

The addition of QDiPs like *denn* to an interrogative clause is subject to several licensing constraints. In the following, we will describe a syntactic and a semantic approach to model these constraints.

Syntactic licensing In our account of syntactic licensing conditions for QDiPs, we follow Bayer et al. (2016). In this account, the central question is how QDiPs can communicate with Force, given their distance. Interrogative Force or QForce is seated in the C-projection (either in C or SpecCP). QDiPs, on the other hand, occur in the middle field of the clause, and follow weak pronouns and discourse topics (just like other discourse particles). Different word orders in sentences are derived by scrambling from the *v*P across the DiP into the topic field. Following Bayer & Obenauer (2011) and Bayer (2012), the communication of a QDiP with QForce over this distance is established via probe-goal agreement or feature sharing between QForce (the probe, which has an interpretable interrogative feature *i*Q) and the QDiP (the goal, which has an uninterpretable feature *u*Q). This type of probe-goal agreement only works if the probe and the goal are not separated by barriers, or in minimalist terms phases. Since discourse particles are located in the head position of a 'particle phase' outside *v*P, they are accessible for agreement with Force in CP.

The preceding explanation shows that QDiPs should not be felicitous in declaratives lacking an *i*Q-feature, and neither in non-interrogative embedded clauses, given that this would leave the QDiP in a CP-phase, out of reach of QForce. See Example 4 for an example of an unlicensed QDiP, and Figure 1 for an illustration of the syntactic structure.

(4) *Wer$_i$ hat t$_i$ sich gefreut, dass das Fest (*denn) ausfällt?*
Who has t REFL rejoiced that the party QDIP not.take.place
'Who was happy that the party will not take place?'

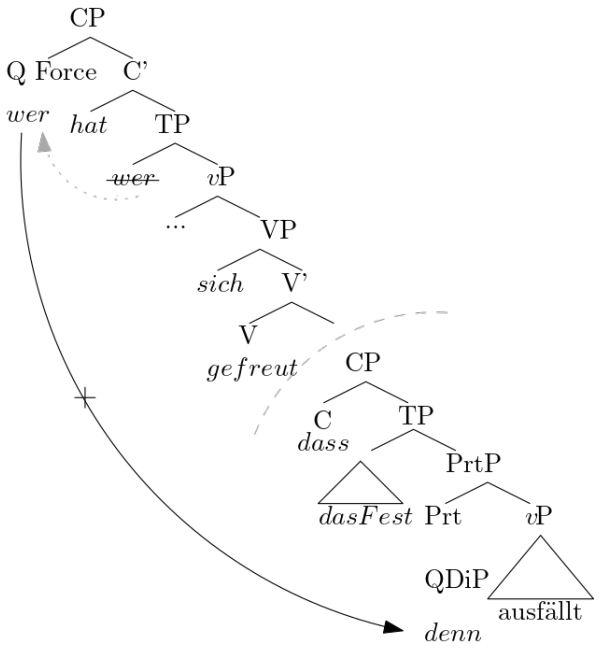

Figure 1. Syntactic tree illustrating the structure of example 4. The CP boundary is marked with a dashed line, the (failed) relationship between the Q-licenser and the QDiP is marked with a crossed-out arrow.

However, this is not an issue if the *wh*-element is extracted from the embedded clause, passing through SpecCP of the embedded clause (and hence through a position where it is accessible for agreement with the QDiP) on its way to the sentence-initial position. This is attested in naturally occurring data (see Example 5) and in experimental findings (Bayer et al., 2016). (In the following, we will also refer to the type of licensing illustrated in 5 as 'QDiP licensing by traces, referring to the fact that a trace of the *wh*-element is left in the SpecCP of the embedded clause.)

(5)

Wie$_i$ denkst du, [$_{CP}$ *t$_i$' dass es denn t$_i$* [$_{vP}$ *weitergehen soll mit euch?*]]
how think you t that it QDiP t on-go should with you
'How do you think that the two of you should carry on?'
http://mein-kummerkasten.de/142829/fremdgehen.html 18-07-2013

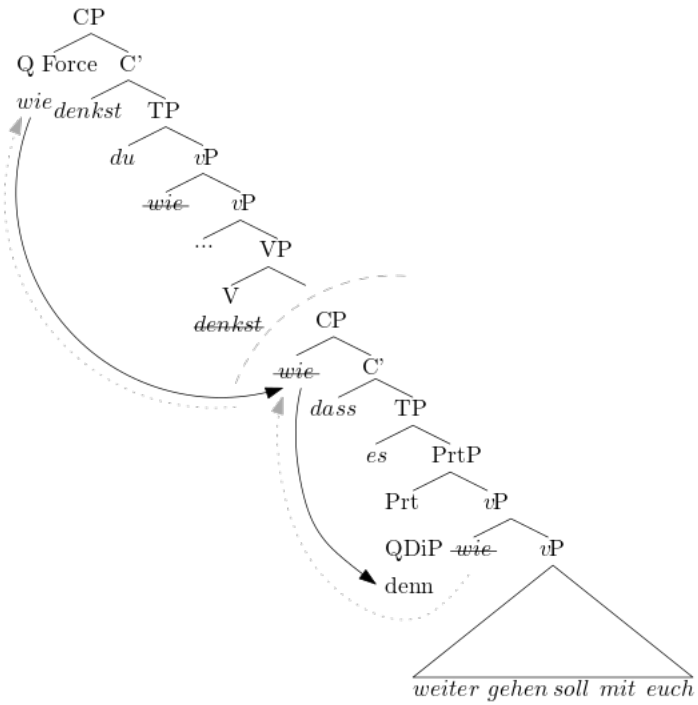

Figure 2. Syntactic tree illustrating the structure of example 5. The CP boundary is marked with a dashed line, the (successful) relationship between the Q-licenser and the QDiP is marked with a series of arrows.

Semantic licensing QDiP licensing can also be described as satisfying semantic licensing conditions, leading roughly to the same predictions as the syntactic licensing conditions outlined above (see Romero 2017 and Czypionka et al. 2020 for a more detailed description). The underlying idea for the theoretical description of semantic licensing is that *wh*-phrases introduce sets of alternatives and are interpreted in their base position (see Hamblin 1973). QDiPs in turn require that the semantic value of their sister be a set of alternative propositions (i.e, of question type). QDiPs then predicate as non-at-issue content some relation between the question meaning constituting the at-issue content and the current common ground. This leads to the following predictions for the felicity of QDiPs in different constructions. If a QDiP occurs in a simple interrogative clause like (1), the syntactic sister of *denn* is of question type introducing a set of alternatives, as in (6), and thus the semantic derivation is successful. In contrast, in declaratives like (2), the syntactic sister of the QDiP will be of a simple propositional type, not of question type, as in (7), and therefore it will not match the QDiPs semantic requirements.

(6) $[\![\text{denn } [_{vP} \text{ du hast was gesagt}]]\!]^w = [\![\text{denn}]\!]^w (\{\lambda w.\text{ you said x in w: x} \in D_e\})$

(7) $[\![\text{denn } [_{vP} \text{ Peter hat das gesagt}]]\!]^w = [\![\text{denn}]\!]^w (\lambda w.\text{ Peter said das in w})$

For interrogatives with long extraction like (5), the syntactic sister of the embedded QDiP is of question type, since the *wh*-phrase introducing a set of alternatives is interpreted in base position, that is, inside the embedded clause. This allows a successful semantic derivation. For interrogatives with short extraction like (4), however, the syntactic sister of embedded *denn*

does not denote a set of alternatives, since the trigger of alternatives – the *wh*-phrase – has its base position in the root clause and not in the embedded clause. Therefore, the semantic derivation crashes.

Taken together, the syntactic and semantic licensing conditions for QDiPs outlined above lead to the same predictions for the behavior of QDiPs in our experiments: QDiPs are predicted to be (i) felicitous in interrogatives if the QDiP is located on the path of the *wh*-chain, (ii) infelicitous in interrogatives if the QDiP is not located on the path of the *wh*-chain, and (iii) infelicitous in declaratives.

1.2 Research questions

While the syntactic and semantic behavior of QDiPs has been the subject of in-depth discussions in the theoretical literature, very little is known about their role in language processing. As outlined above, QDiP licensing is an essentially pragmatic phenomenon that is nevertheless constrained by a number of syntactic and semantic conditions. This in turn makes it a very interesting phenomenon for understanding language processing at the interfaces of syntax, semantics and pragmatics. The four experiments presented in this contribution are aimed to provide a first characterization of QDiP licensing in sentence comprehension. In these experiments, we pursued the following general research questions:

- How do different violations of QDiP licensing constraints affect acceptability?
- What are the EEG correlates for different types of QDiP licensing violations?
- Are there EEG correlates for successful QDiP licensing?

In the following, we will give an overview of the language material used to answer these research questions. We will then present the results of our studies, grouped by experimental methodology (acceptability ratings and EEG measurements).

2 Language Material

Language material consisted of main clauses followed by subordinate clauses. We constructed two different stimulus sets to address different research questions.

2.1 Stimulus set 1: QDiPs in declaratives and interrogatives with short *wh*-extraction

The first stimulus set was built to assess the effects of different types of QDiP licensing violations, namely, QDiPs without a licenser, and QDiPs with a licenser that was syntactically inaccessible. We manipulated the factors DIP TYPE (*denn* and *jetzt*), POSITION (DiP in the root or the embedded clause), and CLAUSE TYPE (interrogative or declarative). The adverb *jetzt* ('now') was chosen because unlike *denn*, it is not constrained by sentence type / illocutionary force. Following this pattern, we constructed 52 item sets.

Example 1 Example item set of stimulus set 1, testing QDiPs in interrogatives and declaratives. This stimulus set was used in experiments 1 and 3.

(a) interrogative, DiP in root clause
Wer hat den Kuchen aus der Bäckerei denn / jetzt aufgegessen?
Who has the cake from the bakery QDIP / nonQDiP up.eat.part
'Who ate the cake from the bakery?'

(b) interrogative, DiP in subordinate clause
*Wer hat gesagt, dass die Oma den Kuchen *denn / jetzt aufessen muss?*
Who has said that the granny the cake QDIP / nonQDiP up.eat must
'Who said that granny has to eat the cake?´

(c) declarative, DiP in root clause
*Robert hat den Kuchen aus der Bäckerei *denn / jetzt aufgegessen.*
Robert has the cake from the bakery QDIP / nonQDiP up.eat.part
'Robert ate the cake from the bakery.'

(d) declarative, DiP in subordinate clause
*Robert hat gesagt, dass die Oma den Kuchen *denn / jetzt aufessen muss.*
Robert has said that the granny the cake QDIP / nonQDiP up.eat must
'Robert said that granny has to eat the cake.´

2.2 Stimulus set 2: QDiPs in interrogatives with short and long wh-extraction.

The second stimulus set was built to monitor QDiP-licensing by *wh*-traces. We manipulated the factors DIP TYPE (*denn* and *jetzt*); POSITION (DiP in the root or the embedded clause), and EXTRACTION (short *wh*-extraction from the root clause, or long *wh*-extraction from the embedded clause). Following this pattern, we constructed 56 item sets.

3 Acceptability ratings

To assess the acceptability of the different constructions presented above, we ran two Magnitude Estimation studies, following Bader (2012). We chose Magnitude Estimation over other acceptability rating methods to allow for a resolution of subtle distinctions, and also to allow for a direct comparison to earlier results reported in Bayer et al. (2016).
Note: Experiment 2 is also reported in a paper accepted for publication in Glossa, see Czypionka et al. (2020). To allow for a better discussion of the corresponding EEG experiment (Experiment 4), we report the results of Experiment 2 here.
Our expectations are as follows: We expect all baseline conditions containing *jetzt* to be acceptable, and to receive similar ratings. In contrast, we expect that *denn* in declarative sentences (Examples 1 (c) and 1 (d)) will be rated as unacceptable, given that it lacks a licenser. For interrogatives with short *wh*-extraction, we expect *denn* in embedded clauses to be rated worse than the corresponding *jetzt* baseline (Examples 1 (b) and 2 (b)), reflecting the fact that there is a clause boundary between *denn* and a potential *wh*-licenser, making the licenser inaccessible. However, for interrogatives with long *wh*-extraction, we expect that *denn* should be acceptable both in root and embedded positions, and should receive ratings similar to the corresponding *jetzt* baseline conditions. In Experiment 1, we expect this to surface as a three-way interaction

Example 2 Example item set of stimulus set 1, testing QDiPs in interrogatives with short and long *wh*-extraction. This stimulus was used in experiments 2 and 4.

(a) short extraction, DiP in root clause
Wer hat denn / jetzt gemeint, dass der Türsteher den Musiker abweisen
Who has QDiP / nonQDiP meant that the bouncer the musician away.turn
soll?
should
'Who QDiP / nonQDiP said that the bouncer should turn away the musician?'

(b) short extraction, DiP in subordinate clause
*Wer hat gemeint, dass der Türsteher den Musiker *denn / jetzt abweisen*
Who has meant that the bouncer the musician QDiP / nonQDiP away.turn
soll?
should
'Who said that the bouncer should QDiP / nonQDiP turn away the musician?´

(c) long extraction, DiP in root clause
Wen hast du denn / jetzt gemeint, dass der Türsteher abweisen soll?
Who.acc have you QDiP / nonQDiP meant that the bouncer away.turn should
'Who did you QDiP / nonQDiP say that the bouncer should turn away?'

(d) long extraction, DiP in subordinate clause
Wen hast du gemeint, dass der Türsteher denn / jetzt abweisen soll?
Who.acc have you meant that the bouncer QDiP / nonQDiP away.turn should
'Who did you say that the bouncer should QDiP / nonQDiP turn away?´

between clause type (interrogative and declarative), DiP type (*denn* or *jetzt*) and DiP position (root or embedded), with an interaction for clause type and DiP position for *denn*, but not for *jetzt*. In Experiment 2, we expect a three-way interaction between extraction (short or long *wh*-extraction), DiP type and DiP position, with an interaction of DiP type and position for short, but not for long extraction.

3.1 Material and methods

Language Material The language material for Experiment 1 was based on stimulus set 1. We constructed eight different lists with 52 critical items distributed evenly across conditions; each participant saw 6 or 7 items per condition. Critical items were interspersed with 40 fillers. The language material for Experiment 2 was based on stimulus set 2. We constructed four different lists with with 112 critical items distributed evenly across conditions; each participant saw 14 items per condition. Critical items were interspersed with 84 fillers. In both experiments, the sentences were rated relative to the reference sentence *Die Mitarbeiter haben, dass der Chef Probleme hat, wohl nicht sofort bemerkt.* ('The employees were probably not aware that the boss was having problems.'); acceptability for this reference sentences was set to 50.

Participants Data from 88 participants were recorded for Experiment 1 (one participant was excluded because he/she did not rate the sentences, for the remaining 87 participants: age span 20 - 40 years, mean age 25 years, s.d. = 3.49 years, 66 female). Data from 56 participants

were recorded from Experiment 2 (one participant was excluded because he/she occasionally assigned ratings over 1000, for the remaining 55 participants: age span 19 - 34 years, mean age 23 years, s.d. = 3, 39 female).

Procedure Experiment 1 was conducted as a web-based experiment on the platform SoSciSurvey (Leiner, 2019), Experiment 2 was conducted as lab-based experiment using the software *Linger* (Rohde, 2003).

Analysis Values deviating more than two standard deviations from a participant's condition mean were removed as outliers (3% for Experiment 1, 4.5 % for Experiment 2). For both experiments, logarithmized *z*-scaled ratings were analyzed using a series of linear mixed effects models in R (R Development Core Team, 2019), using the package *lme4* (Bates et al. 2015, *lme4* function), and *LMERConvenienceFunctions* (Tremblay & Ransijn 2015, *summary* function). Effects were resolved hierarchically, beginning with a full interaction of all three factors (DIP:POSITION:CLAUSE TYPE for Experiment 1, and DIP:POSITION:EXTRACTION for Experiment 2). Only statistically significant effects will be reported, unless explicitly stated otherwise.

3.2 Results

An overview of the mean ratings per condition over participants is given in Table 1. A graph illustrating the results is given in Figure 3.

Table 1. Normalized *z*-scaled mean ratings over participants for all conditions. Standard deviations are given in parentheses.

Experiment 1		Experiment 2	
Condition	norm. ratings	Condition	norm. ratings
decl.-denn-root	-.78 (.73)	short-denn-root	.55 (.35)
decl.-denn-embedded	-.67 (.64)	short-denn-embedded	.09 (.55)
decl.-jetzt-root	.77 (.35)	short-jetzt-root	.43 (.36)
decl.-jetzt-embedded	.71 (.32)	short-jetzt-embedded	.44 (.37)
interr.-denn-root	.70 (.42)	long-denn-root	-.15 (.47)
interr.-denn-embedded	.15 (.46)	long-denn-embedded	-.24 (.54)
interr.-jetzt-root	.70 (.35)	long-jetzt-root	-.17 (.45)
interr.-jetzt-embedded	.67 (.28)	long-jetzt-embedded	-.08 (.47)

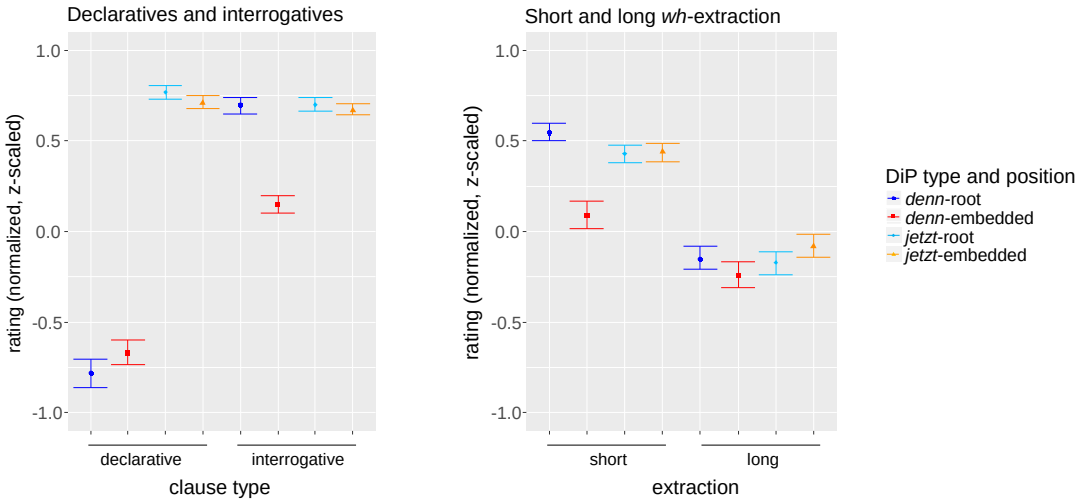

Figure 3. Results for both rating studies side by side. The graph on the left illustrates the normalized mean ratings for Experiment 1, the graph on the left illustrates normalized mean ratings for Experiment 2. Error bars indicate the standard error of the mean.

Results Experiment 1 Mean normalized ratings for all conditions are given in Table 1 (left side), and illustrated in Figure 3 (left panel). Descriptively speaking, *denn* in declaratives received very low ratings (-.78 in root positions, and -.67 in embedded positions); declaratives with *jetzt* were rated as acceptable (.77 for root positions, and .71 for embedded positions). Interrogatives with *denn* in the root clause received high ratings (.70), similar to the *jetzt* baseline (.70), confirming their acceptability. In contrast, interrogatives with *denn* in embedded clauses received lower ratings than the corresponding *jetzt* baseline (.15 for *denn*-embedded, .67 for *jetzt*-embedded). In addition to other main effects and interactions, there was a three-way interaction of CLAUSE TYPE, DiP and POSITION (t = 10.07, p <.001). For conditions with *jetzt*, there was a main effect of CLAUSE TYPE (t = -2.00, p <.05). For conditions with *denn*, there were main effects of CLAUSE TYPE (t = 18.84, p <.001), POSITION (t = 2.92, p <.01), and an interaction of CLAUSE TYPE and POSITION (t = -11.99, t <. 001). The simple main effect of POSITION was significant in interrogatives (t = -8.72, p <.001) and in declaratives (t = 2.69, p <.05).

Results Experiment 2 Mean normalized ratings for all conditions are given in Table 1 right side), and illustrated in Figure 3 (right panel). Descriptively speaking, the pattern for conditions with short *wh*-extraction was similar to the one found for interrogatives in Experiment 1: Ratings were higher for *denn* in root clauses (.55) than in embedded clauses (.09), while ratings for *jetzt* were not affected by position (.43 in root clauses, .44 in embedded clauses). With long *wh*-extraction, ratings for *denn* were similar both in root clauses (-.15) and embedded clauses

(-.24); this was similar to the *jetzt* baselines (-.17 in root clauses, -.08 in embedded clauses). In addition to other main effects and interactions, there was a three-way interaction of DIP TYPE, POSITION and EXTRACTION (t = -4.16, p <.001). For short extraction conditions, there were main effects of POSITION (t = -5.24, p <.001) and DIP TYPE (t = -3.82, p <.001), and an interaction of DIP TYPE and POSITION (t = 5.00, p <.001). For long extraction conditions, there was an interaction of DIP TYPE and POSITION (t = 2.37, p <.05); the resolution of this interaction did not reveal any significant simple main effects.

3.3 Summary acceptability ratings

Ratings for the first stimulus set re-affirm the widely known fact that QDiPs in declaratives (i.e., with absent licensers) are unacceptable. In interrogatives with short extraction, QDiPs in root clauses receive similar ratings to the baseline sentences. Acceptability ratings drop if the QDiP is positioned in the embedded clause instead of the root clause; this becomes visible in both Experiments 1 and 2. This drop in ratings matches the idea that QDiP licensing does not work smoothly across clause boundaries. The absolute ratings, however, are surprisingly high for a construction that should be syntactically/semantically ill-formed, as short-extraction interrogatives with QDiPs in embedded clauses are rated as more acceptable than QDiPs in declaratives, than QDiPs in long extraction conditions, and even higher than the reference sentence in both Experiments. We will return to this surprising finding in the general discussion. In contrast, there is no difference in ratings between QDiPs in embedded and root clause positions if the *wh*-element is extracted from the embedded clause. This finding supports the idea that QDiPs can be successfully licensed by intermediate *wh*-traces.

4 EEG studies

We ran two EEG studies using the language material outlined above. The aim of these studies was to find correlates for different types of QDiP licensing violations and possibly also for successful QDiP licensing. Another aim was to assess qualitative and quantitative differences between QDiP licensing violations due to inaccessible and absent licensers.

We expected effects of QDiP licensing and licensing violations to surface during two time windows corresponding to different components of the event-related brain activity, namely, the N400 and P600 time windows. We based these predictions on the literature on the processing of negative polarity items (NPIs). This is motivated by the fact that there are some important parallels between the licensing of QDiPs and NPIs: Both types of licensing are subject to syntactic, semantic and pragmatic licensing constraints. Both involve a dependency between a licenser ([+Q] for QDiPs, [+NEG] for NPIs) and licensee that need to be in a certain syntactic configuration. In both cases, the licenser needs to c-command the licensee. In the case of QDiPs, the licenser needs to be locally c-commanded by FORCE, in the case of NPIs, there are examples that require strictly local (clause-internal) licensing, and others which allow trans-clausal licensing in the complement CP. While NPI licensing is thought to rely on a c-commanding licenser, there is evidence for so-called "intrusive" licensing, leading to erroneous acceptance of NPIs in the context of a syntactically inaccessible licensers, inaccessible due to a failure of

c-command (see Drenhaus et al. 2005; Saddy et al. 2004; Vasishth et al. 2008; Xiang et al. 2009, 2013; Parker & Phillips 2016). This is reminiscent of the unexpectedly good ratings found for embedded *denn* in interrogatives with short extraction in both rating studies reported above.[1] While studies with English stimuli report an enhanced P600 for unlicensed NPIs (Xiang et al., 2009), studies with German stimuli have reported an enhanced N400 (Saddy et al., 2004), sometimes followed by a P600 (Drenhaus et al., 2005).

Language Material and stimulus presentation The language material for Experiment 3 was 42 item sets from stimulus set 1, the language Material for Experiment 4 was 40 item sets from stimulus set 2.

Stimuli were presented in the center of the screen word-by-word, apart from arguments in the embedded clause, which were presented as full DPs (i.e., article and noun on one screen). Sentences were preceded by a 500 ms asterisk, followed by a 200 ms blank screen. For Experiment 3, each chunk was presented for 750 ms with an ISI of 150 ms; for Experiment 4, each chunk was presented for 800 ms with an ISI of 200 ms. Grammaticality judgments were asked after 20% of the sentences in Experiment 3, and after 10% of the sentences in Experiment 4. For both experiments, all participants saw all stimuli, interspersed with unrelated fillers. Recordings for both experiments were split into two separate sessions conducted on different days.

Participants For Experiment 3, 35 participants were tested (4 removed, 14 male, mean age = 24.5 years, s.d. = 3.3). For Experiment 4, 30 participants were tested (3 removed, 13 male, mean age = 23.96, s.d. = 3.04).

Procedure The EEG was recorded with 61 Ag/AgCl sintered ring electrodes attached to an elastic cap (EasyCap, Herrsching) and connected to an Easy-Cap Electrode Input Box (EiB32). Electrodes were positioned in the equidistant 61-channel arrangement provided by EasyCap (see http://easycap.brainproducts.com/e/electrodes/13M10.htm for electrode layout). Eye movements were monitored by recording the electrooculogram (IO1, IO2, Nz). The ground electrode was located on the right cheek.

Data preparation Data were processed using the Brain Vision Analyzer 2 software (Brain Products, Gilching). Data were prepared by a manual raw data inspection, followed by an ICA blink correction and filtering (low cutoff 0.5 Hz, high cutoff 40 Hz, 50 Hz notch filter), topographic interpolation via triangulation, and a second semiautomatic raw data inspection before segmentation. Segments began 200 ms before stimulus onset and ended 850 ms after stimulus onset for Experiment 3, and 1000 ms after stimulus onset for Experiment 4. Mean

[1] Undoubtedly there are also some major differences between both licensing phenomena, and the theoretical literature on NPI licensing is considerably more detailed than described here. We are reserving an in-depth discussion of the parallels and differences between NPI and QDiP licensing for a future publication.

amplitudes were exported for three time windows: the N400 time window (350-450 ms), the early P600 time window (600-700 ms) and the late P600 time window (700-800 ms). We chose two different P600 time windows to be able to resolve small differences between words in different sentence positions.

Analysis Data were analysed with spatial downsampling to a 25 electrode subset. For the N400 time window, the electrode subset was LO1, FT9, M1, PO9, O9, AF3, FC3, C5, P3, O1, FPz, Fz, Cz, Pz, Oz, AF4, FC4, C6, P4, O2, LO2, FT10, M2, PO10, O10. For the P600 time window, the electrode subset was AF7, FT7, T7, TP7, P7, AF8, FT8, T8, TP8, P8, FP2, C4, P2, O2, F2, FP1, C3, P1, O1, F1, FPz, Cz, Pz, Oz, Fz. These electrode subsets were chosen to allow a good resolution of the skull areas relevant for the respective EEG components. Electrodes in both subsets were parametrized in five medial-lateral and five anterior-posterior regions, added as two topographical factors to the subsequent analyses. Analyses of all reported time windows pursued the influence of the linguistic factors DIP, POSITION and CLAUSE TYPE for Experiment 3, and DIP, POSITION and EXTRACTION for Experiment 4 (just as for the behavioral experiments). In addition, the two topographical factors were added (MEDIAL-LATERAL position and ANTERIOR-POSTERIOR position, with five levels each). For both experiments, analysis set out with an ANOVA with full five-way interaction between linguistic and topographical factors. Interactions were pursued in a hierarchical fashion, beginning with a resolution for different levels of position, and aiming for contrasts between *denn* and *jetzt*. For the sake of brevity and readability, we will limit the report of our results to the effects that are relevant for our research questions. All reported effects are statistically significant unless stated otherwise.

4.1 Results Experiment 3

We were interested in main effects of DIP, and interactions involving DIP and CLAUSE TYPE.

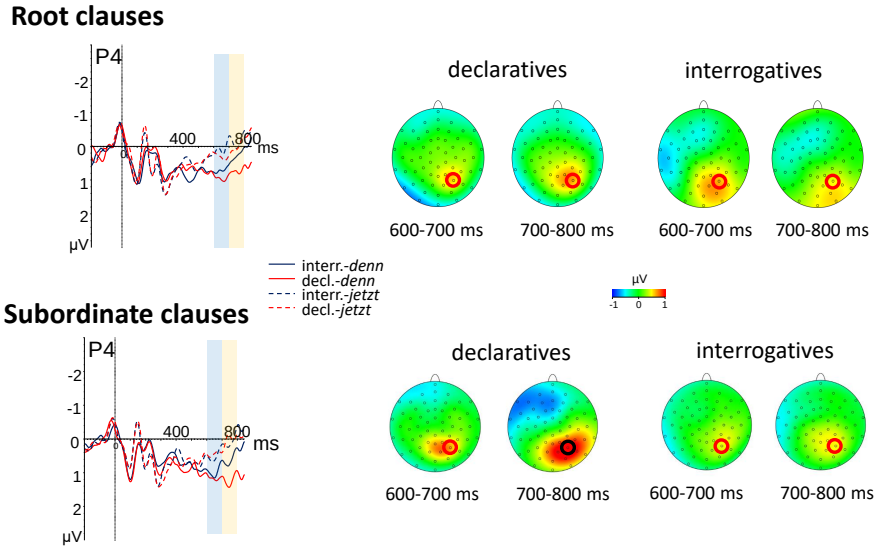

Figure 4. Results for Experiment 3. The upper half illustrates the P600 found in root clauses, the lower half illustrates the P600 found in subordinate clauses. Grand average ERP waveforms for one representative EEG channel and all 4 experimental conditions are shown at the left side. The right side of the figure shows the mean topographies in the early and late P600 time windows for the difference between the corresponding *denn* and *jetzt* conditions. Red or black circles indicate the position of the representative EEG channel.

N400 time window: 350-450 ms There were no interactions of DIP and CLAUSE TYPE, or DIP, CLAUSE TYPE and POSITION.

Early P600 time window: 600-700 ms There was an interaction of DIP, ANTERIOR-POSTERIOR and MEDIAL-LATERAL (F16,480) = 4.65, Greenhouse-Geisser's ε = .48, p <.001), and a marginally significant interaction of DIP, POSITION and CLAUSE TYPE (F(1,30) = 3.81, p <.07). There were no statistically significant main effects or interactions of DIP and POSITION in either root or embedded clauses.

Late P600 time window: 700-800 ms There was an interaction of DIP, POSITION and CLAUSE TYPE (F(1,30) = 8.16, p <.01), and a marginally significant interaction of DIP, POSITION, CLAUSE TYPE, LATERAL-MEDIAL and ANTERIOR-POSTERIOR (F(16,480) = 1.80 ε = .56, p <.07). The interaction of DIP and CLAUSE TYPE was significant in root clauses (F(1,30) = 7.53, p <.05), with marginally significant main effects of CLAUSE TYPE for both *jetzt* (F(1,30) = 3.84, P <.06) and *denn* (F(1,30) = 3.57, p <.07). There were no significant main effects or interactions of DIP and CLAUSE TYPE in embedded clauses.

The interaction of DIP, POSITION, LATERAL-MEDIAL and ANTERIOR-POSTERIOR was pursued separately for root and subordinate clauses. For root clauses, there was an interaction of DIP and CLAUSE TYPE (F(1,30) = 7.53, p <.05), with marginally significant main effects of CLAUSE TYPE for both *jetzt* (F((1,30) = 3.84, p <.06) and *denn* (F(1,30) = 3.57, p <.07). In sub-

ordinate clauses, there was an interaction of DiP, CLAUSE TYPE and ANTERIOR-POSTERIOR ($F(4,120) = 3.68$, $\varepsilon = .41$, $p < .05$). For interrogatives there was a marginally significant main effect of DiP ($F(1,30) = 3.69$, $p < .07$). For declaratives, there was an interaction of DiP and ANTERIOR-POSTERIOR ($F(4,120) = 12.60$, $\varepsilon < .41$, $p < .01$). The simple main effect of DiP was significant in anterior-posterior regions anterior ($F(1,30) = 11.77$, $p < .01$), anterior-medial ($F(1,30) = 15.90$, $p < .001$), posterior-medial ($F(1,30) = 15.77$, $p < .001$), and posterior ($F(1,30) = 11.07$, $p < .01$).

Summary results Experiment 3 In the early P600 time window, the most prominent effect is an enhanced P600 for *denn* relative to *jetzt*, independently of clause type and position. While there was an interaction of DiP, position and clause type, there were no significant results in its resolution. In the late P600 time window, effects were different for root and embedded clauses. In root clauses, declaratives were more positive-going than interrogatives; this held for both *denn* and *jetzt*, but was slightly more pronounced for *denn*. In subordinate clauses, there was an enhanced P600 for *denn* relative to *jetzt* in declaratives; this was descriptively weaker and only marginally significant in interrogatives.

Taken together, this suggests an increase in processing cost for *denn* relative to *jetzt* in all conditions. For embedded clauses, this effect wears off more quickly for *denn* in interrogatives (with an out-of-reach licenser) than in declaratives (without a licenser). The correlate of QDiP licensing violations in this experiment is therefore an enhanced P600. Violations of QDiP licensing constraints did not elicit an enhanced N400 in this experiment.

4.2 Results Experiment 4

We were interested in effects including interactions of DiP with POSITION, or DiP with POSITION and EXTRACTION. Only results relevant to these effects will be reported below.

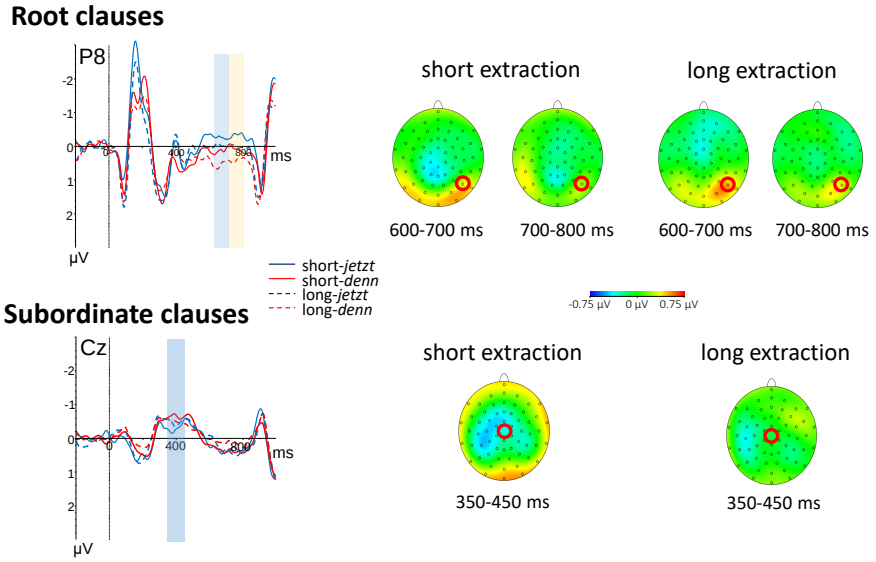

Figure 5. Results for Experiment 4. The upper half illustrates the P600 found in root clauses, the lower half illustrates the N400 found in subordinate clauses. Grand average ERP waveforms for two representative EEG channels and all 4 experimental conditions are shown at the left side. The right side of the figure shows the mean topographies in the early and late P600 time windows, or the N400 time window, respectively, for the difference between the corresponding *denn* and *jetzt* conditions. Red or black circles indicate the position of the representative EEG channel.

N400 time window: 350-450 ms There was a statistically significant five-way interaction of DIP, POSITION, EXTRACTION, LEFTRIGHT and FRONTBACK ($F(16,432) = 1.96$, $\varepsilon = .53$, $p < .05$). The interaction of DIP, EXTRACTION, LEFTRIGHT and FRONTBACK was pursued separately for root and embedded clauses. For root clauses, there were no main effects or interactions involving DIP. For embedded clauses, there was an interaction of DIP and EXTRACTION ($F(1,27) = 6.39$, $p < .05$). The main effect of extraction was statistically significant for *denn* ($F(1,27) = 4.62$, $p < .05$), but not for *jetzt* ($p > .4$).

Early P600 time window: 600-700 ms There was an interaction of DIP, POSITION, EXTRACTION, LATERAL-MEDIAL and ANTERIOR-POSTERIOR ($F(16,432) = 2.45$, $\varepsilon = .46$, $p < .05$). For root clauses, there was an interaction of DIP, MEDIAL-LATERAL and ANTERIOR-POSTERIOR ($F(16,432) = 2.47$, $\varepsilon = .41$, $p < .05$). The interaction of DIP and ANTERIOR-POSTERIOR was significant in medial-lateral regions lateral-right ($F(4,108) = 9.10$, $\varepsilon = .55$, $p < .001$;), lateral-medial-right ($F(4,108) = 8.45$, $\varepsilon = .66$, $p < .001$), medial ($F(4,108) = 4.51$, $\varepsilon = .70$, $p < .01$), and lateral-medial-left ($F(4,108) = 4.65$, $\varepsilon = .65$, $p < .01$), with DIP becoming significant at numerous electrode positions, mainly at posterior-lateral-right regions. For embedded clauses, there were no main effects or interactions involving DIP.

Late P600 time window: 700-800 ms There was a main effect of DIP ($F(1,27) = 4.62$, $p < .05$), and an interaction of DIP with ANTERIOR-POSTERIOR and LATERAL-MEDIAL ($F(16,432)$

= 2.47, ε = .53, p <.05). There were no significant interactions involving DiP and POSITION, or DiP and EXTRACTION.

Summary results Experiment 4 In root clauses (where all conditions are grammatical), there was an enhanced P600 for *denn* relative to *jetzt*. This P600 was restricted to root clauses in the early time window. In the later time window, there were no statistically significant differences between root and embedded clauses; however, it became descriptively weaker in root clauses and was not visible in embedded clauses. In embedded clauses, there was an additional enhanced N400 for *denn* with short extraction (i.e., with an out-of-reach licenser) relative to the other conditions (that were all grammatical).

4.3 Summary EEG studies

In most instances, there was an enhanced P600 for *denn* relative to *jetzt*. This becomes visible in grammatical conditions (e.g., root positions in Experiment 4), and in conditions with unlicensed *denn* (e.g., root and embedded positions in Experiment 3, especially early P600 time window). The P600 deflection is shorter for licensing violations that are due to out-of-reach licensers than for those due to absent licensers (embedded conditions in Experiment 3, late P600 time window).

For Experiment 4 only, a violation of syntactic locality in QDiP licensing (subordinate QDiPs with out-of-reach licensers) elicited an enhanced N400. The same condition did not elicit an N400, but rather an early P600 in Experiment 3. This suggests that the correlate elicited by these conditions is sensitive to the stimulus context in the experiments. A possible explanation could be that Experiment 3 contained violations judged as more severe (*denn* in declaratives) than the condition eliciting an N400 (*denn* with an out-of-reach licenser), and that in the latter experiment, the relatively mild violation became more prominent. This is also interesting in light of the fact that NPI licensing violations have been associated with an enhanced P600 with English stimuli (Xiang et al., 2009), but with an enhanced N400 (Saddy et al., 2004) or N400 - P600 sequence (Drenhaus et al., 2005) with German stimuli. Our findings suggest that licensing processes at the interfaces are sensitive to experimental context.

For Experiment 3, results for root clause positions are more difficult to interpret than the other conditions. This is because the curves for interrogative-*jetzt* are more negative-going than for declarative-*jetzt* (a difference that is not an issue in the other comparisons). While it still allows us to conclude that there is a difference in the P600 amplitude for *denn* and *jetzt*, it does not allow us to claim a bigger P600 amplitude difference for declaratives (i.e. outright violations) in contrast to interrogatives for these positions. However, as our other results allow us to link P600 amplitude to QDiP processing (Experiment 3, embedded clauses, for violations of licensing constraints; Experiment 4, root clauses, for successful QDiP licensing), the general results of our EEG studies are not affected. Future studies should aim to replicate this contrast, using stimuli tailored to monitor the correlates of QDiP licensing violations in root clauses with more parallel baselines to complete the picture.

Taken together, both the successful checking of QDiP licensing constraints and the detection of QDiP licensing violations are associated with an increase in the P600 amplitude. When compared directly, the enhanced P600 is longer-lasting for absent than for inaccessible Q-licensers. So far, an enhanced N400 did only occur for inaccessible Q-licensers, and only in the absence of stronger violations. For now, we cannot assess the nature of the contradictory findings for inaccessible Q-licensers. They could reflect the violation of the syntactic / semantic licensing constraints outlined in the introduction; additional workload connected to processing dispreferred, but ultimately acceptable sentences (as per the rating study); or the processing of sentences erroneously considered acceptable due to shallow online processing (parallel to the reports of intrusive licensing for NPIs with inaccessible licensers, see below).

5 General Discussion and Conclusion

Our findings show, perhaps trivially, that QDiPs need to be licensed by a [+Q] operator or semantically by a question meaning. Unlicensed QDiPs are rated as unacceptable, and are associated with an enhanced P600 in all investigated positions.[2] Successful QDiP processing in well-formed structures is also associated with an enhanced P600 for *denn* relative to *jetzt* (root clauses in Experiments 3 and 4). For now, it remains open if this enhanced P600 reflects a simple increase in processing cost due to the workload for checking different licensing constraints or for processing additional not-at-issue content (see Dörre et al. 2015, 2018, for examples of increased processing cost for discourse particles relative to their at-issue counterparts), or another difference in processing between QDiPs and non-QDiPs.

As predicted by theory, and in replication of Bayer et al. (2016), QDiPs can be licensed by an intermediate *wh*-trace. QDiPs in both root and subordinate clauses receive high acceptability ratings when the *wh*-element is extracted from the subordinate clause, fitting in with the idea of cyclic *wh*-movement. The difference between long and short *wh*-extraction did not affect the amplitude of the *denn* vs. *jetzt* contrast in root clauses, and neither in embedded clauses, suggesting no significant workload differences related to *denn* licensing. For reasons of space, we omit an in-depth discussion of these findings, and refer to Bayer et al. (2016).

Our findings also suggest that QDiP licensing does not work entirely smoothly across clause boundaries. In interrogatives with short *wh*-extraction, there is a drop in ratings for embedded relative to root clause QDiPs. Importantly, the drop in ratings is much less pronounced than we predicted based on the licensing conditions formulated in the introduction, with these syntactically ill-formed conditions being rated above reference. This (surprisingly mild) drop in ratings is however still reflected in the EEG results, as a short P600 in Experiment 3, and an N400 in Experiment 4. This suggests that this violation is not completely ignored. Findings from Experiment 3 (early P600) alone might be interpreted as a reflection of a processing error, with participants erroneously 'licensing' a QDiP even in syntactically ill-formed structures; the P600 would then reflect the increased workload associated with processing QDiPs that are per-

[2] In a series of parallel self-paced reading experiments, we found that these conditions are associated with longer reading times. See Czypionka et al. (2020) for details.

ceived as well-licensed. [3] Experiment 4 however shows a clear divide between reflections of well-formed and illicit structures, with a P600 associated with *denn* in well-formed structures and an N400 for *denn* with an out-of-reach licenser. This suggests a qualitative difference in processing for *denn* with inaccessible licensers relative to accessible licensers. Taken together, our data fit the idea that these structures (interrogatives with short *wh*-extraction and embedded *denn*) are not syntactically/semantically well-formed (hence the drop in ratings relative to root clause *denn*; hence the enhanced N400 or P600). At the same time, they suggest that these sentences are nevertheless interpretable (hence the surprisingly good ratings relative to well-formed conditions; hence the shorter P600 relative to the conditions with absent licensers). We speculate that these seemingly contradictory findings can be explained as the result of an alternative licensing strategy that is employed when the syntactic/semantic licensing constraints are not strictly met.

What would this alternative licensing look like? Romero (2017) proposes to "relax" the semantic analysis into a pragmatic analysis, so that, when semantic licensing is not satisfied, the meaning derivation can still be saved pragmatically and hence the QDiP licensed. The idea is roughly as follows. The QDiP *denn* needs to combine, as before, with a question meaning $Q_{<<s,t>,t>}$. This question meaning may be provided directly by the semantics of its sister constituent, as before, or it may be a Question-Under-Discussion (QUD) retrieved from the context and reflected in the focus structure of the sister constituent. Independent support for this second, pragmatic line of licensing comes from naturally occurring examples like (8), which is easily understood as having focus on *ernsthaft* 'seriously' and hence giving rise to the set of alternatives in (9). This set of alternatives signals that the question 'To what extent does this man want a relationship?' is a salient QUD in the context:

(8) Glaubst du, dass dieser Mann denn ernsthaft eine Beziehung führen möchte?
'Do you think that this man DENN seriously wants to be in a relationship?'
(Bayer et al., 2016)

(9) { 'that this man seriously wants a relationship',
'that this man tentatively wants a relationship',
'that this man jokingly wants a relationship', ...}

Note that (8) and our partially degraded experimental items share the same syntactic/semantic configuration (short *wh*-dependency and embedded QDiPs). Hence, in neither case does syntactic/semantic licensing succeed. However, the fate of naturally occurring sentences like (8) and of our experimental items splits when we come to the sketched alternative pragmatic licensing. In naturally occurring examples like (8), the rich context of utterance and the ease to focus *ernsthaft* 'seriously' in this context pragmatically provide a QUD meaning that can serve as the

[3] A similar phenomenon in the licensing of NPIs is called 'intrusive licensing' (see Drenhaus et al. 2005; Parker & Phillips 2016; Saddy et al. 2004; Vasishth et al. 2008; Xiang et al. 2009, 2013; Yurchenko et al. 2013, a.o.). Explanations for intrusive NPI licensing include errors during cue-based retrieval (Vasishth et al., 2008), and an overapplication of pragmatic licensing processes (Xiang et al., 2009; Parker & Phillips, 2016). While this discussion provides important pointers for future work, our current results do not allow us to claim to what extent illusory licensing plays a role in QDiP processing.

$Q_{<<s,t>,t>}$ argument of *denn*. This suffices to license the QDiP. In contrast, in our experimental studies, no previous context or focus structure was provided that could help the reader identify the intended QUD meaning. This means that the pragmatic search for a salient QUD fails and, thus, the sentences are judged as degraded for pragmatic reasons. For further details about how this alternative pragmatic line of licensing applies to other cases, see Czypionka et al. (2020).

References

Bader, M. (2012). The German *bekommen* passive: A case study on frequency and grammaticality. *Linguistische Berichte*, *231*, 249–298.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bayer, J. (2012). From modal particle to interrogative marker: A study of German *denn*. In L. Brugè (Ed.), *Functional Heads (The Cartography of Syntactic Structures)*, volume 7 (pp. 13–28). New York: Oxford University Press.

Bayer, J., Häussler, J., & Bader, M. (2016). A new diagnostic for cyclic *wh*-movement: Discourse particles in German questions. *Linguistic Inquiry*, *47*(4), 591–629.

Bayer, J., & Obenauer, H.-G. (2011). Discourse particles, clause structure, and question types. *The Linguistic Review*, *28*(4), 449–491.

Csipak, E., & Zobel, S. (2016). Discourse particle *denn* in the antecedent of conditionals. In C. Piñón (Ed.), *Empirical Issues in Syntax and Semantics 11* (pp. 31–60). Paris: CSSP.

Czypionka, A., Romero, M., & Bayer, J. (2020). Question-sensitive discourse particles at the interfaces of syntax, semantics and pragmatics - an experimental approach. *Glossa*, (accepted).

Drenhaus, H., Saddy, D., & Frisch, S. (2005). Processing negative polarity items: When negation comes through the backdoor. *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 145–165.

Dörre, L., Czypionka, A., Trotzke, A., & Bayer, J. (2015). The at-issue and non-at-issue meaning of modal particles and their counterparts. In *Proceedings of the 28th Annual CUNY Conference on Human Sentence Processing, Los Angeles, USA*.

Dörre, L., Czypionka, A., Trotzke, A., & Bayer, J. (2018). The processing of German modal particles and their counterparts. *Linguistische Berichte*, *255*, 58–91.

Grosz, P. (2005). *"Dn" in Viennese German. The syntax of a clitic version of the discourse particle "denn"*. Diplomarbeit, University of Vienna.

Gutzmann, D. (2015). Continuation-based semantics for modal particles. Deriving syntax from semantics. *MIT Working Papers in Linguistics*, *75*, 133–150.

Hamblin, C. L. (1973). Questions in Montague grammar. *Foundations of Language*, *10*(1), 41–53.

König, E. (1977). Modalpartikeln in Fragesätzen. In H. Weydt (Ed.), *Aspekte der Modalpartikeln* (pp. 115–130). Tübingen: Niemeyer.

Leiner, D. (2019). Sosci survey (version 3.1.06) [computer software]. https://www.soscisurvey.de.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339.

R Development Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rohde, D. (2003). *Linger: a flexible platform for language processing experiments, version 2.94*.

Romero, M. (2017). German discourse particles in questions. In *Invited talk at the workshop "Inquisitiveness Below and Beyond the Sentence Boundary" (InqBnB2). University of Amsterdam, December 18th-19th*.

Saddy, D., Drenhaus, H., & Frisch, S. (2004). Processing polarity items: Contrastive licensing costs. *Brain and Language*, *90*, 495–502.

Theiler, N. (2020). *Denn* as a highlighting-sensitive particle. *Linguistics and Philosophy*, 1–40.

Thurmair, M. (1991). Zum Gebrauch der Modalpartikel 'denn' in Fragesätzen. Eine korpusbasierte Untersuchung. In Klein, E., Pouradier Duteil, F., & Wagner, K. H. (Eds.), *Betriebslinguistik und Linguistikbetrieb. Akten des 24. Ling. Koll. Bremen, September 1989*, (pp. 377 – 387). de Gruyter.

Tremblay, A., & Ransijn, J. (2015). *LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models*. R package version 2.10.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Wegener, H. (2002). The evolution of the German modal particle 'denn'. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 379 — 393). Amsterdam: Benjamins.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in psychology*, *4*, 708.

Yurchenko, A., Den Ouden, D.-B., Hoeksema, J., Dragoy, O., Hoeks, J. C., & Stowe, L. A. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia*, *51*(1), 132–141.