Negation in Questions

MSc Thesis (*Afstudeerscriptie*)

written by

Noortje Joost Venhuizen
(born October 19, 1986 in Groningen, the Netherlands)

under the supervision of **Dr Floris Roelofsen** and **Dr Galit Weidman Sassoon**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
August 30, 2011	Dr Floris Roelofsen
	Dr Galit Weidman Sassoon
	Dr Ing Robert van Rooij
	Dr Maria Aloni
	Prof Dr Benedikt Loewe (chair)

Institute for Logic, Language and Computation

Abstract

In this research, we focus on different types of polar questions, in particular the difference between positive polar questions and negative polar questions with high and low negation. We propose a theoretical framework and empirically test some of the predictions of the framework, focusing on the differences in contexts in which different polar question types are licensed. A context is taken to be a combination of speaker belief and contextual evidence. It is shown that positive polar questions and the different types of negative polar questions indeed differ from each other in terms of the contexts in which they occur. Accordingly, the results are formalized in the framework of inquisitive semantics, resulting in an analysis where polar questions coincide with respect to their informative and inquisitive content, but differ from each other in terms of highlighted possibilities and felicity conditions.

Acknowledgements

I would like to thank Floris for being so enthusiastic and involved in this project. And Galit, for being supportive and critical at the times I needed it. Furthermore, Tirza made some wonderful drawings and Merel and Jasper saved me by helping out with the statistics. Thanks to Adrian and Simone for useful and elaborate tips via email and Morgan for his enthusiastic ideas. I would also like to thank Maria for being a great mentor throughout my studies. Finally, of course, my parents who have always supported me and Arón, without whose patience, devotion and great cooking I've never would have made it through. All of you, bedankt!

Contents

Chapter 1

Introduction and Framework

1.1 Polar questions

In the field of linguistics, the study of negation concerns a range of linguistic levels and has been the subject of many semantic and pragmatic analyses. Another area of great interest is the study of questions, which do not conform to the traditional semantic picture of meaning as truth-conditional content. In this research we investigate the combination of these two phenomena and try to formalize how the meaning of a question changes when it contains a negation.

We will focus on polar, *yes/ no* questions. Consider the following examples.

(1) a. Is John coming to the party?
b. Is John not coming to the party?
c. Isn't John coming to the party?
d. Isn't John not coming to the party?

The examples in (1) show a syntactic distinction between different types of polar questions. As opposed to examples (1b)-(1d), example (1a) does not contain any negation, and may thus be called *positive*. Example (1b) contains a negation that syntactically binds to the verb '*coming*', we call this *low negation*. In contrast, the negation in example (1c) syntactically binds to the inverted auxiliary '*is*'; we call this *high negation*. Example (1d) shows that the two types of negation can be combined in one sentence, resulting in a *high/low negation*.

Now consider the following set of examples.

(2) a. John is coming?
b. John is not coming?
c. John is coming, isn't he?
d. John is coming, is he?
e. John is not coming, isn't he?
f. John is not coming, is he?

The examples in (2) differ from the questions in (1) in that they do not have an inverted auxiliary. However, they are inquisitive because they raise an issue between multiple possibilities.[1] We will call (2a) and (2b) *positive* and *negative declarative questions*, respectively, because of their declarative structure, with the auxiliary between the noun and the verb. The questions in (2c)-(2f) are called *tag questions*, where the polarity of the tag may or may not agree with the polarity of the declarative sentence to which the tag is attached (constant polarity tag question and reverse polarity tag question, respectively). When referring to a specific type of tag question, we will explicitly refer to the polarity of the declarative and the tag in the question, e.g. (2c) is a *pos-neg* tag question, (2d) a *pos-pos* tag question, etc.

The syntactic distinction between different types of (negative) polar questions suggests a distinction on the semantic level as well. We will investigate different contexts in which these polar questions can occur and accordingly propose an analysis.

1.2 Inquisitive Semantics

In order to account for the different types of polar questions, we need a framework that can account for the truth-conditional content of a sentence – also called it's *informative* content – as well as for the issue that a question raises – it's *inquisitive* content. For this, we will use inquisitive semantics. In inquisitive semantics the meaning of a sentence is taken to be a combination of the informative and the inquisitive content of the sentence. We will here focus on propositional inquisitive semantics (as proposed in Groenendijk, 2009; Groenendijk & Roelofsen, 2009; Ciardelli & Roelofsen, 2011, among others), but the current analysis can be straightforwardly extended to first-order inquisitive semantics (e.g. Ciardelli, 2010). Some formal notions from inquisitive semantics are introduced to obtain a handle to analyze the data in the next chapter.

1.2.1 Formal properties

What we will call a *sentence* is the linguistic object that expresses a proposition. The syntactic representation of a sentence is called a *clause*, which consists of multiple *constituents*. In inquisitive semantics, the *proposition* expressed by a sentence is a set of possibilities, which can be seen as possible updates of the common ground. A *possibility* is a set of possible worlds (*indices*) and embodies a way to update the common ground.

With respect to sub-sentential semantic and syntactic composition, much notions are adopted from alternative semantics (Hamblin, 1973; Kratzer & Shimoyama, 2002; Alonso-Ovalle, 2006, among others). In particular, we will use the denotation of constituents and the notion of pointwise function application. Each constituent is considered to be of a certain type, which allows them to be combined compositionally. The basic types are e (for entities), s (for indices/

[1] For the influence of intonation, see Gunlogson (2001); Nilsenova (2006).

possible worlds) and t (for truth values), which constitute the domains D_e, D_s and D_t, respectively. For all types σ, τ it holds that $\langle \sigma, \tau \rangle$ is also a type, with $D_{\sigma,\tau}$ as corresponding domain. Most ordinary constituents get as denotation a singleton set containing their traditional denotation. For example, in (3) and (4) below the denotations are shown of the proper name '*John*' of type e and the verb '*play*' of the property type $\langle e, \langle s, t \rangle \rangle$. Other constituents, typically those that express multiple alternatives, like for example disjunctions or questions, denote multi-membered sets of alternatives. Example (5) shows a disjunction, which is again of type e.

(3) $[\![\text{ John }]\!] := \{\text{John}\}$

(4) $[\![\text{ play }]\!] := \{\lambda x \lambda v.\text{play}(x)(v)\}$

(5) $[\![\text{ John or Bill }]\!] := \{\text{John, Bill}\}$

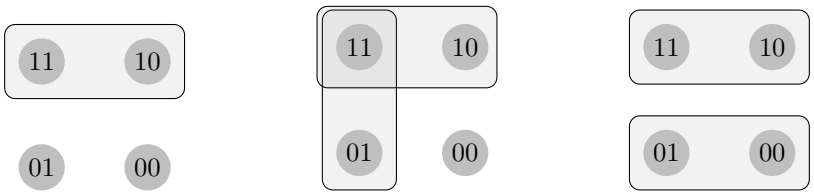
These denotations can accordingly be combined using pointwise function application to form a complete clause. The idea is that by using pointwise function application, a multi-membered set can compose with a singleton set by composing once for each alternative, thus generating a new multi-membered set. Pointwise function application is formalized as follows (Kratzer & Shimoyama, 2002):

Definition 1 (Pointwise function application)**.** If α is a branching node with daughters β and γ, and $[\![\beta]\!] \subseteq D_\sigma$ and $[\![\gamma]\!] \subseteq D_{\langle \sigma,\tau \rangle}$, then
$[\![\alpha]\!] := [\![\beta\gamma]\!] := [\![\gamma\beta]\!] := \{a \in D_\tau | \exists b \exists c[b \in [\![\beta]\!] \wedge c \in [\![\gamma]\!] \wedge a = c(b)]\}$

So, given (4) and (5) above, we can apply pointwise function application, resulting in a proposition of type $\langle s, t \rangle$, as shown in (6).

(6) $[\![\text{ John or Bill plays }]\!]_{w,g} := \{\lambda v.\text{play}(\text{John})(v), \lambda v.\text{play}(\text{Bill})(v)\}$

In inquisitive semantics the *proposition expressed by* a sentence φ is the set of possibilities it denotes. So, a proposition consists of one or more possibilities, where each possibility is a set of indices. If a sentence proposes two or more possibilities, it is called *inquisitive*: it proposes an issue and invites other participants to provide information to resolve the issue. Two possibilities are called *alternative* possibilities when one possibility is not properly included in the other; in other words, each possibility contains at least one index that is not included in the other possibility.
Now, the *inquisitive content* of a sentence consists of the possibilities it proposes, whereas the *informative content* consists of the indices that a sentence *excludes*: the indices that are not part of any of the possibilities that the sentence proposes. Some examples of sets of possibilities are visualized in figure 1.1 (here, 11 is the index where both p and q are true, 10 is the index where only p is true, etc.). As we saw above, the proposition associated with $p \vee q$, shown in figure 1.1(b), is inquisitive; this contrasts traditional semantic theories, where the denotation of a disjunction is taken to be the union of these possibilities. Note that the possibilities proposed by $p \vee q$ are alternative possibilities. As



(a) $[\![p]\!]$ (b) $[\![p \vee q]\!]$ (c) $[\![?p]\!]$

Figure 1.1: Visualization of possibilities for different sentences.

shown in figure 1.1(c), a polar question has no informative content; it excludes no indices, so it only has an inquisitive content.

Note that polar questions consist of two alternative possibilities, corresponding to the classical tautology $p \vee \neg p$. These possibilities are taken to represent the 'complete' answers to a polar question: the question poses an issue and thus requests an answer that picks out one of the possibilities. Since polar questions offer exactly two disjunct possibilities that cover all indices, *yes/ no* answers are taken correspond to these possibilities; in the case above '*yes*' corresponds to the possibility p and '*no*' corresponds to the possibility $\neg p$.

The denotation of questions is realized by the question operator [Q] that operates on a clause φ of type $\langle s, t\rangle$, returning the clause [Q φ], again of type $\langle s, t\rangle$. Semantically, [Q] takes the union of the set of possibilities in the denotation of φ and the set of possibilities that φ excludes. We can define the possibilities that a sentence φ excludes as the set of indices that are included in the possibilities of $\neg\varphi$, represented as the complement of the possibilities of φ: $\overline{[\![\varphi]\!]}$, as shown in definition 2.

Definition 2 (Complement)**.** $\overline{[\![\varphi]\!]} = \{\overline{\bigcup\{\alpha \mid \alpha \in [\![\varphi]\!]\}}\} = [\![\neg\varphi]\!]$

It follows from this definition that the proposition expressed by a negated clause $\neg\varphi$ always contains (at most) one possibility, which consists of all the indices that are not in the union of the possibilities for φ. Thus, in inquisitive semantics φ and $\neg\neg\varphi$ are not fully equivalent, since the proposition expressed by $\neg\neg\varphi$ only contains a single possibility, corresponding to the union of the possibilities for φ.

The proposition expressed by [Q φ] consists of the possibilities of φ itself, plus the possibilities that φ excludes. Definition 3 defines the question operator.

Definition 3 (Question operator)**.** $[\![\text{Q } \varphi]\!] := [\![\varphi]\!] \cup \overline{[\![\varphi]\!]}$

We define the notion of the *sentence radical* of a clause $[Q\ \varphi]$, denoted as $[\![Q\varphi]\!]_R$, in order to capture the common denominator of syntactically different polar questions. For example, the polar questions in (1) and (2) all seem to be 'about' the proposition expressed by 'John is coming (to the party)'. This is recursively defined as follows:

Definition 4 (Sentence Radical)**.** The sentence radical of a question consisting of a question operator and a sentence φ, $[\![Q\varphi]\!]_R$, is defined as follows:

if $\varphi = \neg\psi$:	$[\![Q\varphi]\!]_R = [\![Q\psi]\!]_R$
otherwise:	$[\![Q\varphi]\!]_R = \varphi$

Moreover, we will call the clause to which the question operator attaches the *prejacent* of the question, i.e., φ is the prejacent of [Q φ] and $\neg\varphi$ is the prejacent of [Q $\neg\varphi$].

Definition 5 (Prejacent)**.** $\textsc{prej}([\mathrm{Q}\varphi]) = \varphi$

Note that the prejacent of [Q φ] may differ from its sentence radical: while the sentence radical excludes all external negations, the prejacent may be of the form $\neg\psi$.

On the basis of definition 3, positive polar questions and negative polar questions receive the same denotation: they both express a proposition consisting of the possibilities proposed by $[\![\varphi]\!]$ and the possibilities proposed by $[\![\neg\varphi]\!]$. However, as we will see, positive and negative polar questions are used in different contexts which suggests that they should not be treated equivalently.

In the next chapter we will further investigate this observation and in chapter 3 the predictions are empirically tested. Then, in chapter 4 we will come back to the formal framework and investigate the implications of the results of the experiment for the denotation of positive and negative polar questions. The final chapter presents the conclusions.

Chapter 2

Background

2.1 Earlier Approaches

Negated questions have been extensively studied. This section discusses some of the most prominent accounts on negated questions, which we divide into three types: the semantic approach (Ladd, 1981; Romero & Han, 2004), the pragmatic approach (van Rooij & Šafářová, 2003) and the meta-linguistic approach (Horn, 1989; Reese, 2006, 2007). We will see that all approaches one way or another analyze (negated) polar questions in terms of the context in which they occur; we present the discourse approach (after Büring & Gunlogson, 2000) in section 2.2.

2.1.1 The Semantic Approach

Ladd (1981) formulated an ambiguity between what he calls an *inner* negation and an *outer* negation reading of negative polar questions, which distinguishes them from positive polar questions (see also Borkin, 1971; Ladusaw, 1979). In particular, he focuses on questions with what we above called a *high* negation. Consider some of Ladd's original examples (emphasis added):

(7) (Situation: A and B are former left-wing activists discussing the recent activities of a colleague.)
A: Did you hear John's decided to go to business school?
B: Yeah–I can't believe how much he's changed these days–*didn't he even vote for Reagan?*
A: That's what somebody told me.

(8) (Situation: A and B are staunch republicans)
A: What's Dick been up to these days–I haven't seen him at the Club for ages.
B: Haven't you heard? He says he's disillusioned with two-party politics–he's joined Common Cause, gave a lot of money to the Citizen's Party...
A: *Didn't he even vote for Reagan?*
B: Not as far as I know.

Here, the question with high negation in (7) is considered to be an outer negation question, since the speaker uses the question to confirm the (positive) proposition expressing that John voted for Reagan. In example (8), however, the negative question is used to check the (negative) inference that Dick didn't vote for Reagan, so this Ladd calls an inner negation question. In other words, in the case of outer negation questions the speaker expects a positive answer (in example (7) above, "John voted for Reagan"), whereas in the case of inner negation questions the expected answer contains a negation (in example (8) above: "John did not vote for Reagan").

Ladd (1981) argues that this distinction is based on an actual syntactic/ semantic ambiguity, involving a difference in the scope of the negation, i.e. in the case of the inner negation reading the negation in the question scopes over the prejacent of the question ("John voted for Reagan"), while in the case of the outer negation reading it does not. This idea is motivated by the observation that polarity items can be used to force a specific interpretation of a negative question. For example, *too* and *either* are polarity items, such that *either* is only licensed in negative contexts (NPI) and *too* typically occurs in non-negative contexts (PPI); this is illustrated in (9).

(9) a. Jane is coming too.
b. # Jane is not coming too.
c. # Jane is coming either.
d. Jane is not coming either.

Ladd (1981) observes that these polarity items can also be used to disambiguate between the inner or the outer negation reading of negative polar questions, as is shown in (10).

(10) a. Isn't Jane coming too?
b. Isn't Jane coming either?

Here, (10a) gets an outer negation reading because the use of *too* signals that it is not in the scope of the negation, while the use of *either* in (10b) does signal that it is within the scope of the negation and thus receives an inner negation reading. For an empirical study investigating the distinction proposed by Ladd, see Hartung (2006a,b).

Ladd (1981) does not provide a semantic theory to account for the different types of negative polar questions, but Romero & Han (2004) do provide such an analysis. They do this by introducing a new operator, VERUM. The basic idea is that negated questions contain an implicit VERUM operator which presupposes a backgrounded speaker attitude and interacts with the negation of the question, yielding the inner and outer negation readings proposed by Ladd (1981). The introduction of the VERUM operator is motivated by the observation that negated questions with high negation exhibit an epistemic bias that is similar to questions containing the epistemic adverb *really*. For example, they argue that (11a) and (11b) both require a backgrounded attitude of the speaker towards

the sentence radical that 'Jane is coming'; in the case of (11a) there is a negative epistemic implicature, i.e., the speaker believed or at least expected that Jane is not coming, while in the case of (11b) the implicature is of opposite polarity, i.e., the speaker has the background belief that Jane is coming.

(11) a. Is Jane really coming?
b. Isn't Jane coming?

In order to define the semantic contribution of *really*, Romero & Han (2004) introduce the VERUM operator, which is used to assert that the speaker is certain that the prejacent p should be added to the *common ground* (cg). The definition of VERUM is provided in (12), where $Epi_x(w)$ is the set of worlds that conform to x's knowledge in w, $Conv_x(w')$ is the set of worlds where all the conversational goals of x in w' are fulfilled; a conversational goal may be, for example, to acquire maximal information about the state of the world. $CG_{w''}$ is the common ground in w'', i.e., the set of propositions that the speakers assume in w'' to be true (Stalnaker, 1978).

(12) $[\![\text{VERUM}_i]\!]^{g x/i} = \text{FOR-SURE-CG}_x =$
$\lambda p_{\langle s,t\rangle} \lambda w. \forall w' \in Epi_x(w)[\forall w'' \in Conv_x(w')[p \in CG_{w''}]]$

So, VERUM φ is true in a world w iff some discourse participant x (usually the speaker or the addressee) is certain that in all of the worlds in which the conversation goals of x are met the proposition φ is part of the common ground, i.e., x is sure that φ can be added to the common ground: FOR-SURE-CG$_x$ φ.

Romero & Han (2004) assume that in the case of negated questions, the inverted auxiliary together with the high negation introduce VERUM into the logical form of the question. They follow Groenendijk & Stokhof (1984) in describing the denotation of a question as a partition on the set resulting from intersection of the propositions in the common ground; thus, the Q operator is defined as follows:

(13) $[\![Q]\!] = \lambda p_{\langle s,t\rangle} \lambda w_s \lambda q_{\langle s,t\rangle}[q = p \vee q = \neg p]$

Note that this definition is in effect the same as how we defined the question operator in definition 3 above, only stated in different terms. Now, with the VERUM operator in the logical form of negated questions, Ladd's (1981) ambiguity can be reduced to a difference in scope between the negation and VERUM operator: in the outer negation reading the negation scopes over VERUM and in the inner negation reading VERUM scopes over the negation. Examples (14) and (15) show the logical form and denotation for the outer and inner negation reading of (11b), respectively.

(14) $[\![[_{CP}\ Q$ not [VERUM $[_{IP}$ Jane is coming]]] $]\!] =$
{FOR-SURE-CG$_x$ Jane is coming, $\neg$FOR-SURE-CG$_x$ Jane is coming}

(15) $[\![[_{CP}\ Q$ VERUM [not $[_{IP}$ Jane is coming]]] $]\!] =$
{FOR-SURE-CG$_x$ $\neg$Jane is coming, $\neg$FOR-SURE-CG$_x$ $\neg$Jane is coming}

Thus, the question in the case of (14) addresses x's certainty about whether or not the proposition expressing that 'Jane is coming' is in the common ground; this corresponds to x's positive bias, and thus the question gets an outer negation reading. In (15), on the other hand, the issue posed by the question relates to x's certainty about whether or not the proposition expressing that 'Jane is *not* coming' is in the common ground; so this reflects x's negative bias, and thus the question gets an inner negation reading.

Romero & Han (2004) provide a semantic analysis for the distinction between positive and negative questions, as well as between inner and outer negation questions. Various authors have questioned the assumptions of this approach. The main argument against this theory is based on the felicitous answers to polar questions. As described above, most semantic theories (e.g. Hamblin, 1973; Groenendijk & Stokhof, 1984; Groenendijk & Roelofsen, 2009) assume that the meaning of a question can be formulated in terms of the complete answers it licenses. So, the *yes* and *no* answers to a question are taken to represent the set of *possibilities* a sentence proposes (in terms of the inquisitive semantics defined above). Now, this means that the *yes* and *no* answers to negative polar questions should correspond to the possibilities in the denotation of outer and inner negation questions provided by Romero & Han's (2004) analysis, shown in (14) and (15). Below, (16a-i) and (16b-i) show the interpretation of the *yes* and *no* answers, respectively, to the outer negation reading of the question in (16), as predicted by Romero & Han's (2004) analysis and (16a-ii) and (16b-ii) represent the predicted interpretation of the *yes* and *no* answers to the inner negation reading. It seems, however, that the most intuitive interpretation of *yes* and *no* is provided in examples (16a-iii) and (16b-iii).

(16) Isn't Jane coming?
  a. Yes.
    i. $\Rightarrow$ FOR-SURE-CG$_x$ Jane is coming.
    ii. $\Rightarrow$ FOR-SURE-CG$_x$ Jane is not coming.
    iii. $\Rightarrow$ Jane is coming.
  b. No.
    i. $\Rightarrow$ not FOR-SURE-CG$_x$ Jane is coming.
    ii. $\Rightarrow$ not FOR-SURE-CG$_x$ Jane is not coming.
    iii. $\Rightarrow$ Jane is not coming.

This example provides a strong argument against Romero & Han's (2004) semantic analysis of negated questions. Romero (2005) has offered a response to the criticisms, by suggesting that VERUM contributes to the expressive content rather than the truth-conditional content of a sentence, similar to the use of epistemic modals like '*must*' in the example (17) below.

(17) a. Why isn't Louise coming to our meetings these days?
  b. She must be too busy with her dissertation.

In (17b), '*must*' is used to comment on the truth-conditional content of its prejacent. According to Romero (2005), VERUM shows similar behaviour, such that answers to questions containing VERUM respond to the prejacent of VERUM rather than VERUM itself. The idea is countered by Reese (2007), who states that Romero's (2005) assumption that VERUM is an expressive operator contradicts Romero & Han's (2004) analysis of the inner /outer negation ambiguity as a genuine scope ambiguity, since expressive content is generally not taken to semantically embed with respect to truth-conditional operators (Potts, 2005). Moreover, the analysis seems to assume that the prejacent of inner negation questions does not contain a negation, while this is one of the central assumptions of Ladd's (1981) scope-account of the inner/outer negation ambiguity.

We now leave this approach and turn to a pragmatic account proposed by van Rooij & Šafářová (2003).

2.1.2 The Pragmatic Approach

van Rooij & Šafářová (2003) propose a decision-theoretic approach to account for the use of polar questions. They argue that the semantic approach is not adequate to distinguish between positive and negative polar questions (and alternative questions), because of its two-valued nature: if one is interested in learning if some proposition p is the case, then he should be equally interested in learning if $\neg p$ is the case. From this assumption it follows that positive and negative polar questions are semantically equivalent. van Rooij & Šafářová's (2003) analysis is based on the idea that speaker's beliefs and preferences determine the form of the question, which is formalized in Bayesian decision theory. The account also embraces alternative questions, but we will here focus on the analysis of polar questions.

In short, the idea is that a speaker chooses the use of a certain polar question on the basis of the answer that has the highest estimated *utility* relative to the *beliefs* and *desires* of the speaker. So, each participant is taken to be in a certain belief-desire state, modeled as $\langle P, U\rangle$, where P is a probability function representing beliefs and U is the utility function representing desires. The expected utility (EU) of a belief-desire state is calculated as follows:

(18) $EU(P,U) = \sum_{w \in W} P(w) \times U(w)$

Thus, EU measures the degree to which an agent prefers to be in a particular belief-desire state $\langle P, U\rangle$. Now, in order to determine how good it is to learn the new information that a proposition q is the case, the difference is measured between the expected utility of the information states before and after the agent learned the proposition q. The probability function after conditionalizing P with q is denoted as P_q, i.e., $P_q(w)$ is the probability of the world w given the proposition q. The value (UV) of the new information q is then defined as follows:

(19) $UV(q) = EU(P_q, U) - EU(P, U)$

van Rooij & Šafářová (2003) argue that the choice of a speaker to use a positive or a negative polar question depends on which answer has the highest expected utility for the speaker. While assuming semantic equivalence between positive and negative polar questions, they are taken to differ in terms of their expected utility. In particular, positive polar questions (PPQs) are used when the utility of the positive answer is greater than the utility of the negative answer, and for negative polar questions (NPQs) this is exactly the other way around; this is summarized in (20) below:

(20) a. PPQs: $UV(q) > UV(\neg q)$
b. NPQs: $UV(q) < UV(\neg q)$

The utility function can be seen in terms of *informativity*: for NPQs it holds that $UV(\neg q) > UV(q)$ iff $\text{inf}(\neg q) > \text{inf}(q)$. The informativity (or: *surprisal value*) of q, $\text{inf}(q)$, is defined as follows: $\text{inf}(q) = -log_2 P(w)$.[1] Thus, this means that as the probability of a proposition decreases, its surprisal value increases. So, learning that a proposition q is the case is more informative if q is considered to be unlikely, than if q already has a high probability. It thus follows that for NPQs it holds that $P(q) > P(\neg q)$ according to the speaker's probability function; this explains the positive bias inherent to NPQs. However, in the same way the analysis predicts PPQs to be negatively biased: $P(q) < P(\neg q)$, which is of course not what we want. van Rooij & Šafářová (2003) resolve this by assuming that negative statements are *by default* less informative than positive statements. So, polar questions are by default stated in terms of a positive polar question and thus not because of some negative bias.

With respect to negative polar questions, van Rooij & Šafářová (2003) argue that Ladd's (1981) distinction between inner and outer negation polar questions can be discarded; they are semantically and morpho-syntactically indistinguishable. The difference, they state, lies in the belief-desire state that is used to compute the relative utility of the possible answers, which, in the case of inner negation questions is the *prior* belief-desire state and for outer negation questions is the *current* belief-desire state. Crucially, however, van Rooij & Šafářová (2003) state that this distinction also exists for PPQs, so it is not typical for negative polar questions. Inner negation questions can be seen as grounding questions, in much the same way as PPQs can be used when the speaker is reluctant to accept some new information q. In the case of inner negation questions, however, the new information is a negative proposition, and the new information is not taken into account for the calculation of the utility of each answer. These similar uses of PPQs and inner negation questions are illustrated in example (21) below (from Reese, 2007).

(21) a. A: John is back from Hawaii.
b. B: Was John in Hawaii?
c. B: Didn't John go to Fiji?

[1] See van Rooij & Šafářová (2003) for a formal deduction of this definition.

Both (21b) and (21c) are responses to (21a) that do not accept the information provided in (21a). Assuming that John cannot be in two places at once, the proposition that John went to Fiji implies that he did not go to Hawaii. Thus, the inner negation reading of (21c) indicates that B is reluctant to ground the information that (21a) provides, while the outer negation reading conveys that, considering the counterevidence in (21a), the speaker still considers it more likely that John went to Fiji than that he did not.

However, the assumption that there is no semantic or morpho-syntactic difference between inner and outer negation questions can be challenged. As Ladd (1981) already observed, inner and outer negation questions license different polarity items: *either* and *too* can be used to distinguish between the inner and outer negation reading. Furthermore, as already described in section 1.1, the syntactic distinction between high and low negation, although not conclusive in the decision between inner and outer negation, provides a strong indication in favor of the actuality of a semantic distinction.

2.1.3 The Metalinguistic Approach

Another approach to negated questions is based on the idea that outer negation functions as metalinguistic negation and is proposed by Reese (2006, 2007), following Horn (1989) and Potts (2005).

Horn (1989) proposes that metalinguistic negation is a common function of negation in natural language and describes it as a "device for objecting to an utterance on any grounds whatever, including the conventional and conversational implicata it potentially induces, its morphology, its style or register, or its phonetic realization" (p. 363). Reese notes that he does not assume with Horn that metalinguistic negation is essentially different from truth-functional negation; his use of metalinguistic negation is concerned with the illocutionary act of *denial* or *correction*. Moreover, Horn (1989) argues that metalinguistic negation does not license negative polarity items, nor does it block the use of positive polarity items; this seems to correspond to Ladd's (1981) analysis of the licensing of *either* and *too* in inner and outer negation questions. It was shown that *either* and *too* can be used to disambiguate between the inner and outer negation reading of a question, since *too* forces the outer negation reading while *either* induces the inner negation reading.

Reese argues that the negation in outer negation questions can be seen as a metalinguistic negation, since outer negation questions have similar functions, i.e., denial. Example (22) below illustrates this (adapted from Reese, 2007).

(22) a. A: None of the students turned in their assignment.
b. B: Jane turned in her assignment.
c. B: Didn't Jane turn in her assignment?

Since assertions are taken to commit the speaker to its content, the assertion in (22b) clearly challenges the information provided in (22a). Similarly, the outer negation reading of (22c) provides a denial of the information provided in (22a),

although a less forceful one. Note that the inner negation reading of (22c) is also available in this context, but it does not function as a denial: this reading can merely be used to check the inference that Jane didn't turn in her assignment (cf. the grounding function proposed by van Rooij & Šafářová, 2003).

On the basis of these examples, Reese (2007) considers outer and inner negation questions to have distinct discourse functions, where outer negation questions can have both an assertive and an inquisitive function and inner negation questions only have an inquisitive function (see also Asher & Reese, 2005). Reese (2007) motivates this classification by observing that outer negation questions can co-occur with sentence initial *after all*, which is typical for assertions, as well as with sentence initial *tell me*, which characterizes questions (Sadock, 1974)[2]. Inside negation questions, however, can only occur with inquisitive discourse markers. This is illustrated in examples (23) and (24) (adapted from Reese, 2007).

(23) a. # Tell me, Jane is coming.
b. Tell me, is Jane coming?
c. Tell me, isn't Jane coming {too/ either}?

(24) A: Sue isn't coming, so there'll be no syntacticians here.
B: What do you mean? ...
a. After all, the MIT syntacticians are coming.
b. # After all, are the MIT syntacticians coming?
c. After all, aren't the MIT syntacticians coming too?
d. # After all, aren't the MIT syntacticians coming either?

The felicity of the questions in (23b) an (23c) and the infelicity of the assertion in (23a), according to Reese, directly follow from the fact that both inner and outer negation questions, as well as positive questions, are inquisitive and the assertion is not. With respect to example (24), note that, similar to example (22) above, B does not accept the information provided by A. Reese assumes that *after all* typically occurs in such contexts. We again see that the assertion in (24a) and the outer negation question in (24c) are felicitous after *after all*. Since PPQs (24b) and inner negation questions (24d) do not occur in denial contexts and *after all* is only licensed in these contexts, they can never co-occur with each other.

Reese (2006, 2007) provides a formal analysis of negative polar questions in terms of the Segmented Discourse Representation Theory, SDRT (Asher & Lascarides, 2003). It is beyond the scope of this thesis to go into the details of his analysis. However, the main assumptions are that outer negation questions are assigned a complex speech act type ASSERTION • QUESTION by the grammar and, secondly, that inner negation polar questions and positive polar questions are considered to be semantically equivalent. The provided analysis is based

[2]Crucially, *tell me* should be followed by a comma to obtain this reading. When not followed by a comma, *tell me* can be used to express a desire and can occur with assertions.

on pragmatic mechanisms to model the restricted use of various polar question types; in this respect Reese's (2007) analysis agrees with van Rooij & Šafářová's (2003) approach. However, while van Rooij & Šafářová (2003) abstain from making a distinction between inner and outer negation questions, Reese (2007) considers this to be warranted on the basis of the linguistic evidence he provides. The framework, SDRT, defines the meaning of a sentence in terms of its discourse function as determined by its connection to the preceding context. Given the examples provided by different authors, it seems that it is indeed the case that the context of a negated question crucially determines its interpretation. In the next section we will describe the proposal made by Büring & Gunlogson (2000) and extend it towards a testable analysis of negative polar questions.

2.2 The Discourse Approach

From the examples above it followed that the felicity of polar questions in many cases depends on the context it occurs in. We will work out this idea in the current section, following Büring & Gunlogson (2000), who propose that the *contextual evidence* crucially determines the felicity of (negated) polar questions. We will propose that another contextual factor plays an important role, namely the previous belief of the speaker, which we will refer to as the *speaker belief*.

In order to forestall a self-fulfilling theory, we abstain from making a semantic distinction between inner and outer negation, but rather focus on the syntactic distinction between high and low negation, as proposed in chapter 4.1. We thus investigate the interpretation of positive polar questions (PPQs), high negation polar questions (HNPQs), low negation polar questions (LNPQs) and high/low negation polar questions (HLNPQs). First we will describe Büring & Gunlogson's (2000) proposal for contextual evidence, who base their theory on the inner/ outer negation distinction. This will be extended to account for speaker belief and accordingly we describe the predicted felicity conditions of the syntactic question types on the basis of the proposed theory. The major part of these predictions will be tested in the experiment described in chapter 3.

2.2.1 Contextual Evidence

Büring & Gunlogson (2000) follow Ladd (1981) in assuming that there is a difference in meaning of positive polar questions, inner negation polar questions and outer negation polar questions. As a reinforcement of Ladd's (1981) observation that inner and outer negation polar questions license different polarity items, Büring & Gunlogson (2000) provide an example from German, which shows a clear morpho-syntactic difference between inner and outer negation questions. The example is repeated in (25).

(25) a. Gibt es *nicht ein* vegetarisches Restaurant in dieser Ecke?
gives EXPL not a vergeterian restaurant in this corner
'Is there not some vegetarian restaurant around here?' (*outer*)

b. Gibt es *kein* vegetarisches Restaurant in dieser Ecke?
gives EXPL no vergeterian restaurant in this corner
'Is there no vegetarian restaurant around here?' (*inner*)

Here we see that the PPI '*nicht ein*' is licensed in outer negation questions, while inner negation questions license the NPI '*kein*'. Thus, these items can again be used to disambiguate between the inner and outer negation reading of negative polar questions. Note that the *kein/ nicht ein* distinction in German has an equivalent in English; the translation in (25) shows that the *not a/ no* distinction in English can serve a similar purpose, i.e., disambiguating between inner and outer negation questions.

Büring & Gunlogson (2000) propose that the felicity conditions for polar questions demonstrate that the use of the different question types is restricted by the *contextual evidence*. Consider the following example (adapted from Büring & Gunlogson, 2000).

(26) (Scenario: S and A have a justified previous belief that Carl is right-handed. They encounter Carl cutting bread with his left hand.)
a. S: Is Carl left-handed?
b. # S: Is Carl not left-handed?
c. # S: Isn't Carl left-handed?
d. S: Isn't Carl not left-handed?
e. # S: Is Carl right-handed?
f. S: Is Carl not right-handed?
g. S: Isn't Carl right-handed?
h. # S: Isn't Carl not right-handed?

Because *left-handed* and *right-handed* are antonyms, we can say that the questions with low negation (26b) and (26f) are interchangeable with the respective positive questions with an antonymous complement: (26e) and (26a), respectively. This is the case given the intonational pattern that contracts the low negation with the complement, namely when there is no pause between the negative particle and the adjective. According to the same reasoning, the high/low negation questions in (26d) and (26h) can be considered equivalent to the high negation questions with antonymous complement, (26g) and (26c), respectively. Overall we can thus state that the questions in example (26) have the same sentence radical, namely the proposition expressing that Carl is left-handed (or right-handed, depending on what side of the equivalences we choose). This example shows that the felicity of the different types of polar questions depends on the direct *contextual evidence*, which we define as in definition 6 (adapted from Büring & Gunlogson, 2000). Here, p is taken to be the sentence radical of the question.

Definition 6 (Contextual evidence)**.** Contextual evidence for or against a proposition p w.r.t. a discourse situation σ is evidence for or against p that

has just become mutually available to the participants in the current discourse situation σ.

This definition is vague in at least two respects. First, it is unclear what is meant by the 'current discourse situation'; no theory is provided that formalizes what a discourse situation is. Moreover, there are various ways in which evidence about a proposition can 'become mutually available to the participants', for example extra-linguistically, as in the case of (26) above. However, this definition does provide us with an intuition of how a context may restrict the interpretation of a question and it gives us a flexible notion that can be formalized in different discourse analysis theories (see for example the discourse structure proposed by Farkas & Bruce, 2010). In order to use this definition to define different contexts, Büring & Gunlogson (2000) introduce the notion of *compelling* contextual evidence, which they define as follows.

Definition 7 (Compelling)**.**

i. Evidence *for* p is compelling if, considered in isolation, it would allow the participants to assume p.

ii. Evidence *against* p is compelling if it is compelling evidence for the opposite of p, $W - p$ (with W= the set of all worlds).

Consider again example (26). Given that 'p' expresses the proposition 'Carl is left-handed', the given scenario in (26) provides *compelling contextual evidence for* p. Note that in accordance with definition 7(i), the evidence that Carl is cutting with his left hand would, *in isolation*, justify the inference that p, namely that Carl is left-handed. It follows that the infelicity of examples (26b) and (26e) is based on the context; even though both conversation participants have a previous belief that $\neg p$ is the case, questioning this negated proposition as in (26b) and (26e) is infelicitous.

Overall, we can thus define three types of contextual evidence: contexts with *neutral* contextual evidence with *no* compelling contextual evidence for or against p, contexts with compelling contextual evidence *for* p and contexts with compelling contextual evidence *against* p. In the following, these contexts are described as providing 'neutral', 'p' and '¬p' contextual evidence, respectively. On the basis of these context types, Büring & Gunlogson (2000) formulate evidence conditions for the three question types, which are summarized in table 2.1

Contra Büring & Gunlogson (2000) we propose that it is not the case that the contextual evidence solely defines the felicity conditions for all types of questions. As the infelicity of (26c) and the felicity of (26g) show, felicity conditions may also depend on the previous belief of the speaker (cf. Romero & Han, 2004). This will be discussed in the next section.

2.2.2 Speaker Belief

We propose that questions with high negation may convey a biased attitude of the speaker (cf. Romero & Han, 2004); in the case of (26g) towards the

Semantic type	Evidence condition
PPQ	There is *no* compelling contextual evidence against p (i.e. there is either no evidence or evidence *for* p).
INPQ	There is compelling contexual evidence *against* p.
ONPQ	There is *no* compelling contextual evidence for p (i.e. there is either no evidence or evidence *against* p).

Table 2.1: Evidence conditions on positive polar questions (PPQs), inner negative polar questions (INPQs) and outer negation polar questions (ONPQs).

proposition expressing that Carl is right-handed, which in this case is felicitous. We define the belief state of the speaker as follows (again p is the sentence radical of the question).

Definition 8 (Speaker Belief)**.** The speaker's belief of a proposition p w.r.t. a discourse situation σ is the speaker's belief about p before the current discourse situation σ.

This definition again lacks a formal description, but for the current purposes it will suffice. We will assume that *before the current discourse situation* here means, "before the contextual evidence became available". Following this definition, in (26) there is a speaker belief that Carl is right-handed and not left-handed. Because of this speaker belief, the high negation question in (26g) is acceptable, whereas (26c) is not. This shows that the speaker's belief state may also be relevant for the felicity conditions of a question.

Analogous to contextual evidence, we also have three contexts for speaker belief: contexts where the speaker has no previous belief; contexts where the speaker has a belief that p; and those where the speaker has a belief that $\neg p$: these are expressed as contexts with 'neutral', 'p' (positive) and '¬p' (negative) speaker beliefs (SB), respectively.

2.2.3 Context

We can define a context in terms of a combination of speaker belief and contextual evidence, which can each have three states, *for*, *against* and *neutral* with respect to a proposition. We thus have a total of nine contexts. In other words, there are nine ways in which a context may relate to a proposition, where contextual evidence and speaker belief are two necessary relations. The context is defined as follows:

Definition 9 (Context)**.** A context relates to a proposition p in two ways:

i. Providing compelling contextual evidence *for* p, *against* p, or *neutral* with respect to p.

ii. Providing a speaker belief *for* p, *against* p, or *neutral* with respect to p.

Now we can define the felicity of the the different question types in each of the contexts.

2.3 Felicity conditions

In this section we investigate the felicity of the different question types for the contexts defined above. The predictions made on the basis of these felicity conditions will be tested in the experiment described in chapter 3.

2.3.1 PPQs and LNPQs

We follow Büring & Gunlogson (2000) in assuming that the felicity conditions of positive polar questions and low negation polar questions are restricted by the contextual evidence. As we saw in (26), low negation polar questions like (26b)/ (26e) cannot occur in contexts with compelling contextual evidence *for* the proposition expressing that Carl is left-handed, whereas positive polar questions corresponding to (26a)/(26f) can. More generally, we propose that it is the case that positive polar questions can occur in all contexts except those with compelling contextual evidence *against* the radical of the question, p, and low negation questions can occur only in these contexts. This is shown by the following observations (adapted from Büring & Gunlogson, 2000).

(27) (Scenario: S and A are talking long-distance on the phone.)
   a. S: What's the weather like out there? Is it sunny?
   b. # S: What's the weather like out there? Is it not sunny?

(28) (Scenario: A enters S's windowless office wearing a dripping wet raincoat.)
   a. # S: What's the weather like out there? Is it sunny?
   b. S: What's the weather like out there? Is it not sunny?

In example (27), the contextual evidence is *neutral* because it provides no compelling contextual evidence for or against the proposition p: '*it is sunny*'. Although the positive polar question in (27a) is perfectly acceptable, the low negation polar question in (27b) sounds inappropriate. The context in (28) clearly provides compelling contextual evidence *against* p, since it provides compelling contextual evidence for the complement of p: '*it is raining*'. In this case, it is inappropriate to utter a positive polar question, as in (28a), while the low negation polar question in (28b) is acceptable.

2.3.2 HNPQs and HLNPQs

Similarly, we argue that high negation polar questions can also be restricted by the contextual evidence in the same way as low negation polar questions: they are felicitous in contexts with compelling contextual evidence *against* p.

However, we propose that high negation polar questions can also be restricted by the belief state of the speaker. Consider examples (29)-(31). Examples (29)-(31) represent utterances made by one speaker as a single utterance.

(29) I have no idea who is coming to the party....

a. # ... isn't it the case that John is coming?

b. ? ... isn't John coming?

(30) I heard Sue is coming to the party and if Sue comes, John will also come...

a. ... isn't it the case that John is coming?

b. ... isn't John coming?

(31) I heard Sue is coming to the party and if Sue comes, John will not come...

a. # ... isn't it the case that John is coming?

b. ? ... isn't John coming?

Note the difference between the high negation questions in (a) and (b): we will call questions like (a) *explicit* high negation questions and those in (b) *simple* high negation questions. This example shows that (explicit) high negation questions may occur in a context where the speaker expresses an earlier belief that p is the case, as in (30), but not in contexts where the speaker is either neutral with respect to the proposition or has a prior belief that the negation holds, as (29) and (31) show. Analogously, it can be shown that high/low negation polar questions occur only in contexts where the speaker believes that $\neg p$ is the case.

Note that only simple high negation questions can be restricted by the contextual evidence; while a simple high negation would be felicitous in example (28) above, an explicit high negation would not. Below we will see that the explicit high negation corresponds to the outer negation reading of high negation questions.

2.3.3 Summary

The felicity conditions described above are summarized in table 2.2, which shows the contexts in which each of the different question types can occur.

Syntactic Type	**Restricted by**	**Occurs in**
PPQ	Context. Evidence	'*neutral*', 'p' CE contexts
LNPQ	Context. Evidence	'$\neg p$' CE contexts
HNPQ	Context. Evidence, or Belief State	'$\neg p$' CE contexts, or 'p' SB belief states
HLNPQ	Belief State	'$\neg p$' SB belief states

Table 2.2: Felicity conditions for different question types

Note that there is not a one-to-one correspondence between the restrictive contexts and the syntactic question types; high negation questions are predicted to occur both in $\neg p$ CE contexts and in p belief states. Moreover, comparing the predictions in table 2.2 to Büring & Gunlogson's (2000) evidence conditions shown in table 2.1, we see that the predictions for PPQs coincide, as do the preditions for the questions Büring & Gunlogson (2000) call INPQs and are here called LNPQs. With respect to (simple) HNPQs, we see that the predictions agree

with the predictions made for ONPQs made in table 2.1 in that they are felicitous in contexts with compelling contextual evidence *against* p, but, contrary to Büring & Gunlogson's (2000) analysis, they are not predicted felicitous in contexts with *neutral CE*, but are instead dependent on *positive* speaker belief.

As noted above, the use of *explicit* high negation may function as a disambiguation between the inner and outer negation reading of high negation questions; it provides the outer negation reading. Moreover, because LNPQs are predicted felicitous in the same contexts as inner negation questions, this too seems to be a disambiguation method. Thus, the *inner/outer* negation distinction may be considered to represent the distinction made here between contexts with compelling contextual evidence against p and those with 'p' belief states. Because high negation questions are predicted to be felicitous in both types of context, they can be said to have both an inner and an outer negation reading. As we have seen, there are several other ways to distinguish between these different uses of high negation questions, for example by using an explicit context (e.g. Ladd, 1981), using NPI's (van Rooij, 2003; Asher & Reese, 2005) or with an explicit high negation as we saw above in examples (29)-(31).

This also explains why negative questions intuitively seem to involve some kind of suggestion, sometimes called *bias*, or *expected answer*.[3] This suggestion can be related to the felicity conditions introduced in table 2.2. Because questions are restricted by a particular context, this context functions as a suggestion. LNPQs, for example, signal that $\neg p$ has somehow become salient in the current context. HNPQs, on the other hand, when restricted by the belief state, differ from PPQs because of the negation that occurs in the question. This, intuitively, results in a biased attitude of the speaker towards the proposition p. Similarly, HLNPQs exhibit a biased attitude towards the proposition $\neg p$.

2.3.4 Predictions

The felicity conditions described in section 2.3.3 above are now worked out in an example, which shows the predicted felicitous and infelicitous question types in all contexts.

Table 2.3 gives an illustration of the different contexts and the questions that are predicted to be felicitous on the basis of table 2.2. Here, we take p to be the proposition expressed by "*It is sunny*".

Firstly, note that in the contexts where the speaker belief and the contextual evidence support the same proposition, either p or $\neg p$, the question types that are predicted to be felicitous sound a bit awkward. We suggest that this is because these contexts are non-inquisitive since there is no conflict between the speaker belief and contextual evidence; we call these *rhetorical* contexts. Our analysis will be mainly focused on non-rhetorical contexts, since these are the most natural contexts for the occurrence of polar questions.

[3] All questions and assertions types can convey a suggestion in many ways, for example by a particular word choice, like the use of NPI's (e.g. Asher & Reese, 2005), or by means of special intonation (e.g. Ladd, 1996; Nilsenova, 2006). However, we will not be concerned with these types of suggestion here.

contextual evidence / **speaker belief**	**S and A are talking long-distance on the phone.**	**A enters S's windowless office wearing a sleeveless shirt and sunglasses**	**A enters S's windowless office wearing a dripping wet raincoat.**
S: "I have no beliefs on the matter, so tell me..."	Is it sunny? #Is it not sunny? #Isn't it sunny? #Isn't it not sunny?	Is it sunny? #Is it not sunny? #Isn't it sunny? #Isn't it not sunny?	#Is it sunny? Is it not sunny? Isn't it sunny? #Isn't it not sunny?
S: "I heard the weather was nice, so tell me..."	Is it sunny? #Is it not sunny? Isn't it sunny? #Isn't it not sunny?	?Is it sunny? #Is it not sunny? ?Isn't it sunny? #Isn't it not sunny?	#Is it sunny? Is it not sunny? Isn't it sunny? #Isn't it not sunny?
S: "I heard the weather was bad, so tell me..."	Is it sunny? #Is it not sunny? #Isn't it sunny? Isn't it not sunny?	Is it sunny? #Is it not sunny? #Isn't it sunny? Isn't it not sunny?	#Is it sunny? ?Is it not sunny? ?Isn't it sunny? ?Isn't it not sunny?

Table 2.3: Examples of felicitous and infelicitous inverted auxiliary polar questions for each context in English.

In accordance with the felicity conditions described above, positive polar questions occur in contexts with neutral contextual evidence or evidence for p ("*It is sunny*"), and low negation questions occur in contexts with contextual evidence against p. Also, the uncertainty about the felicity of (29b) and (31b) is now explained; the felicity of high negation questions, as predicted by the *inner/outer* negation distinction, may depend on the contextual evidence, as well as on the speaker belief. This results in the high negation question being predicted felicitous in the context with neutral speaker belief and compelling contextual evidence against the proposition that it is sunny. For (29b) and (31b) it holds that without any further context it is unclear whether the simple high negation question is felicitous, but what is clear is that it is more felicitous than the examples in (29a) and (31a), respectively. We propose that simple high negation questions such as (29b) and (31b) are felicitous in contexts where the speaker has a belief that p and in contexts with compelling contextual evidence against p. In particular, high negation questions are felicitous in contexts where both conditions hold.

Note that, since we focus our analysis on non-rhetorical contexts, the contexts in which high negation questions occur all express some kind of conflict between the speaker's belief and the contextual evidence; in particular, the context with speaker belief p and compelling contextual evidence against p.

speaker belief \ **contextual evidence**	**neutral**	**p**	**¬p**
neutral	PPQ #LNPQ #HNPQ #HLNPQ	PPQ #LNPQ #HNPQ #HLNPQ	#PPQ LNPQ HNPQ #HLNPQ
p	PPQ #LNPQ HNPQ #HLNPQ	?PPQ #LNPQ ?HNPQ #HLNPQ	#PPQ LNPQ HNPQ #HLNPQ
¬p	PPQ #LNPQ #HNPQ HLNPQ	PPQ #LNPQ #HNPQ HLNPQ	#PPQ ?LNPQ ?HNPQ ?HLNPQ

Table 2.4: Felicity and infelicity of all question types for each context

The predictions made in table 2.3 are summarized in tables 2.4 and 2.5. Table 2.4 shows an overview of all the felicitous and infelicitous question types for all contexts, and table 2.5 shows only the felicitous question types for each context.

Note that PPQs and LNPQs (marked red and blue in table 2.5, respectively) together cover all possible contexts, while HNPQs and HLNPQs (marked green and yellow, respectively) only cover a subset of these contexts. This explains why some languages lack HNPQ and HLNPQ sentences (e.g. Dutch), or why these may be uncommon (like the English '*Isn't John not going to the party?*'). On the other hand, the fact that these question types do occur in many other

speaker belief \ contextual evidence	neutral	p	¬p
neutral	PPQ	PPQ	HNPQ, LNPQ
p	PPQ HNPQ	PPQ HNPQ	HNPQ, LNPQ
¬p	PPQ HLNPQ	PPQ HLNPQ	HNPQ, LNPQ HLNPQ

Table 2.5: Felicitous question types for each context.

languages, among which English, shows that in some cases it is be useful for a speaker to express a previous belief about a proposition.

It can be shown that contextual evidence and speaker belief can also be used to account for the felicity of other question types, for example tag-questions and declarative questions. This motivates the application of these factors in determining the felicity of questions. However, it is beyond the scope of this research to go into these examples.

The felicity predictions of the examples above are based on theoretical reasoning and individual judgments. Therefore, before theorizing about the predictions, we will experimentally test them in order to obtain a reliable foundation for a linguistic theory.

Chapter 3

Experiment

3.1 Introduction and General Hypotheses

In the foregoing chapters we have investigated the felicity of polar questions in different types of contexts. In this chapter we will investigate the above described intuitions in an experiment.

We defined two contextual factors that we take to be crucial for the felicity of different types of polar questions: speaker belief (SB) and contextual evidence (CE). Speaker belief is defined as the belief that the person uttering the question has about the sentence radical of the question, which can be *neutral*, *positive* or *negative*. Contextual evidence is the immediate context of the question which may constitute evidence *for* (*positive*), *against* (*negative*) or *neutral* with respect to the sentence radical of the question. Together, these factors constitute $3*3 = 9$ contexts.

The experiment focuses on three types of polar questions: positive polar questions (PPQs), high negation polar questions (HNPQs) and low negation polar questions (LNPQs). For these three types of questions we will investigate whether it holds that (i) their acceptability depends on the contextual factors SB and CE and (ii) the different question types have different acceptability conditions.

As noted above, Hartung (2006a,b) performed a similar experiment on the felicity of negative polar questions in different contexts. Her work also focuses on testing the felicity of negative polar questions in contexts with different types of compelling evidence (in the sense of Büring & Gunlogson, 2000) and speaker belief (in terms of Romero & Han, 2004). However, a crucial difference between her approach and the one pursued here is that while we investigate the difference between various syntactic question types, she focuses on the difference between *inner* and *outer* negation questions by marking them with *too* and *either*, as in Ladd's (1981) example shown in (10). In the current experiment, we purposely abstain from assuming this typology, thereby hoping to acquire unbiased evidence on the felicity of the syntactically differing question types, which may or may not provide evidence for the inner/ outer negation distinction.

SB \ CE	**neutral**	**p**	**¬p**
neutral	1 PPQ	2 PPQ	3 HNPQ LNPQ
p	4 PPQ HNPQ	5 PPQ HNPQ	6 HNPQ LNPQ
¬p	7 PPQ	8 PPQ	9 HNPQ LNPQ

Table 3.1: Theoretical acceptability of question types per context.

Table 3.1 repeats the acceptability table we described in section 2.3.4, which was based on theoretical judgments about the felicity of the different question types. It shows the acceptability of positive polar questions (PPQs), high negation polar questions (HNPQs) and low negation polar questions (LNPQs) The numbering of contexts is added for ease of discussion.

In words, we considered the acceptability of PPQs and LNPQs to be defined solely by the contextual evidence (*neutral* or *positive* for PPQs and *negative* for LNPQs), while the acceptability of HNPQs can be defined by both contextual evidence and speaker belief (*positive SB* or *negative CE*). These acceptability judgments are summarized by the following set of informal hypotheses (Theory 1):

1. PPQs are more acceptable in contexts with *positive* or *neutral* contextual evidence (contexts 1, 2, 4, 5, 7 and 8) than in all other contexts.

2. LNPQs are more acceptable in contexts with *negative* contextual evidence (contexts 3, 6 and 9) than in all other contexts.

3. HNPQs are more acceptable in contexts with *positive* speaker belief or *negative* contextual evidence (contexts 3, 4, 5, 6 and 9) than in all other contexts.

As described above, contexts 5 and 9 are considered to be special contexts: they are *rhetorical*. When the speaker belief and contextual evidence support the same proposition, the context is taken to be *non-inquisitive*, since there is no issue that needs to be resolved. Thus, we formulate a second set of informal hypotheses, which exclude contexts 5 and 9 as acceptable contexts for all question types (Theory 2):

1. PPQs are more acceptable in non-rhetorical contexts with *positive* or *neutral* contextual evidence (contexts 1, 2, 4, 7 and 8) than in all other contexts.

2. LNPQs are more acceptable in non-rhetorical contexts with *negative* contextual evidence (contexts 3 and 6) than in all other contexts.

3. HNPQs are more acceptable in non-rhetorical contexts with *positive* speaker belief or *negative* contextual evidence (contexts 3, 4, and 6) than in all other contexts.

Finally, we test the differences between the nine individual contexts, as opposed to investigating groups of contexts in terms of contextual evidence and speaker belief. Firstly, we assume that there is an interaction between the question types and the different contexts, i.e., in order to predict the acceptability of a question it is necessary to know both the question type and the context in which it occurs. Moreover, the central assumption of the current research, in contrast to other theories, is that not only the contextual evidence, but also the speaker belief is crucial for the acceptability of the question types and that these factors both have a non-additive effect on the licensing of different question types. Thus, for each question type we predict that there is an interaction between speaker belief and contextual evidence. These predictions are summarized by the predictions below (Interactions):

1. There is an interaction between the question type and the context.
2. For PPQs there is an interaction between SB and CE.
3. For LNPQs there is an interaction between SB and CE.
4. For HNPQs there is an interaction between SB and CE.

Provided these interactions, we can formulate a typical context for each of the question types, since this would be the context where the speaker belief and contextual evidence together determine the felicity of the question. Because above we already assumed that the felicity of HNPQs may depend on both speaker belief and contextual evidence, the most typical context for this question type is the context in which both conditions hold: context 6. We assume that most typical contexts of different question types do not coincide because of pragmatic competition (if one most typical question type is available, then it becomes less likely that another question type is used), so therefore LNPQs are most acceptable in context 3. For PPQs, we propose that *neutral* contexts are preferred over non-neutral contexts, so they are most acceptable in context 1. These assumptions are summarized below (Theory 3):

1. PPQs are most acceptable in contexts with *neutral* speaker belief and *neutral* contextual evidence (context 1).
2. LNPQs are most acceptable in contexts with *neutral* speaker belief and *negative* contextual evidence (context 3).
3. HNPQs are most acceptable in contexts with *positive* speaker belief and *negative* contextual evidence (context 6).

We will design an experiment that can test each of these three theories. Section 3.2 describes the method for the experiment, section 3.3 describes the results separately for each of the three theories and section 3.4 describes the overall discussion.

3.2 Method

3.2.1 Participants

We used Amazon's Mechanical Turk[1] (AMT) to recruit participants. AMT is an online labor market place where workers are paid small amount of money to complete small tasks, named HITs (Human Intelligence Tasks). It has been shown that AMT provides a quick and relatively cheap method to acquire high-quality experimental results that do not differ significantly in performance from standard experimental settings (e.g. Snow et al., 2008; Buhrmester et al., 2011).

The participation of participants was restricted by the worker's location, which was to be in the United States and it was made clear in the instructions that the task was designed for native speakers of English. Moreover, the participants were required to have a HIT approval rate of at least 95%, which means that 95% of the worker's total number of submitted HITs were accepted. However, workers without previously submitted HITs also passed this requirement.

Having taken into consideration the ethical issues posed by crowdsourcing methods like AMT (e.g. Fort et al., 2011), we decided on a payment of $0.03 per HIT. The average completing time of one HIT was 19 seconds, which results in an effective hourly rate of $5.70.

We obtained a total of 109 participants.

3.2.2 Material

Each HIT consisted of an explanation, an example, one target or filler item, presented in the form of a cartoon, and the task to rate the naturalness of the question in the cartoon on a 7-point scale.

In order to test the influence of speaker belief and contextual evidence we designed a cartoon that could be used to manipulate the three levels of each of the factors of interest: speaker belief, contextual evidence and question type. An example is shown in figure 3.1 and the dialogue is shown in (32). The question in the third picture of the cartoon, shown in (32iii), is a PPQ, whose acceptability arguably is determined by the foregoing dialogues (32i) and (32ii).

(32) i. I'm going to buy a BMW. I think a red one.
ii. Did you hear, John bought a car. I heard it's awesome!
iii. Did he buy a BMW?

The utterance in the first picture, shown in (32i), is taken to constitute a speaker belief: the utterer of the question in (32iii) receives information about the sentence radical of the question in a situation that takes place before the direct context of the question. The utterance in example (32i) is taken to constitute *positive SB*, because the information that can be inferred from this utterance ('John bought a BMW') coincides with the sentence radical of the question.

[1] https://www.mturk.com/mturk/welcome

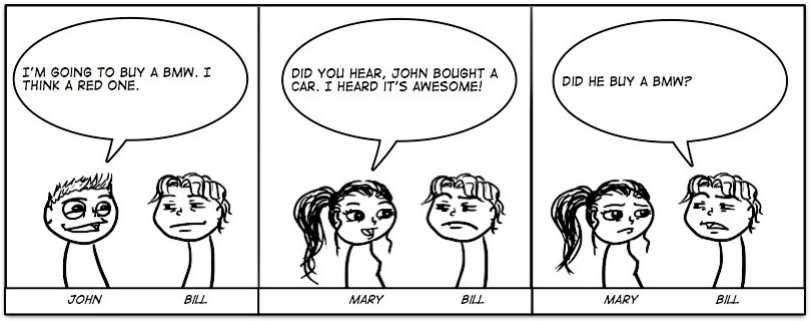


Figure 3.1: Example of target item.

The utterance in the second picture, shown in (32ii), in a similar manner constitutes contextual evidence: the utterer of the question in (32iii) receives information about the sentence radical of the question in the direct context of the utterance of the question, which is signaled by showing the same conversation partners in the second and third picture. The dialogue in example (32ii) above is taken to constitute *neutral CE*, because the information provided here neither supports nor contradicts the sentence radical of the question. Thus, summarizing, figure 3.1 and example (32) represent a PPQ in context 4.

Now consider another instance of the same dialogue:

(33) i. I'm going to buy a car. I think a red one.
ii. Did you hear, John bought a Lexus. I heard it's awesome!
iii. Didn't he buy a BMW?

Here we have a question of type HNPQ against a background of *neutral SB* and *negative CE*. The speaker belief in (33i) is *neutral* with respect to the sentence radical of the question because it neither supports nor contradicts it. The contextual evidence in (33ii) is *negative* because it directly contradicts the sentence radical of the question: the information that 'John bought a Lexus' is incompatible with the proposition that 'John bought a BMW' (in any standard situation), thus we can state that in the current situation 'John bought a Lexus' implies that 'John did not buy a BMW'. So, example (33) is an instance of context 3.

Finally, consider the dialogue in (34).

(34) i. I'm going to buy a Lexus. I think a red one.
ii. Did you hear, John bought a BMW. I heard it's awesome!
iii. Did he not buy a BMW?

Here we have a question of type LNPQ against a background of *negative SB* and *positive CE*. We see that the information provided in (34i) contradicts the sentence radical in the same way as (33ii) above, thus resulting in a *negative*

speaker belief. The contextual evidence provided in (34ii), on the other hand, coincides with the sentence radical of the question in (34iii), resulting in a *positive* contextual evidence.

So, overall we see that all questions in the third picture have the same sentence radical, which contains a specific noun (in the examples above: 'BMW'). With respect to this sentence radical, *positive* speaker belief and contextual evidence is constituted using the same specific noun as in the radical of the question, *neutral* speaker belief and contextual evidence is constituted using a general noun (in the examples above: 'car') and *negative* speaker belief and contextual evidence is constituted using a specific noun that contradicts the specific noun in the sentence radical (in the examples above: 'Lexus').

Moreover, note that the utterances constituting speaker belief and contextual evidence contain an extra sentence after the sentence that provides the context: 'I think a red one.' in dialogue 1 and 'I heard it's awesome!' in dialogue 2. These sentences are added in order to avoid semantic priming of repetitive specific nouns. For example, consider an example of context 5, without these extra sentences:

(35) i. I'm going to buy a BMW.
ii. Did you hear, John bought a BMW.
iii. Did he not buy a BMW?

The felicity of this entire context is taken to be affected by the repetition of the specific noun 'BMW'. Moreover, the earlier occurrences of the specific noun may influence the participants' response to the occurrence of the specific noun in the question, called semantic priming (e.g. Draine & Greenwald, 1998). Adding an extra sentence after each context-marking sentence is taken to increase the naturalness of the subsequent conversations and decrease the effect of semantic priming.

All in all, this resulted in a total of 27 conditions (for each question type 9 contexts, varying by three types of SB and three types of CE). Table 3.2 shows the 9 combinations of SB and CE for one lexical item. Each of these contexts is followed by each of the question types PPQ, LNPQ and HNPQ, with as sentence radical 'John bought a BMW', whose naturalness was to be rated on the basis of the given context.

We had 6 different lexical items differing between each other in terms of the cartoon and the conversation topic. The six lexical items were filled in for each of the 9 contexts for the 3 question types, resulting in a total of $6*9*3 = 162$ target items. The cartoons were designed to enhance the participants' intuition about the naturalness of the question; for this means the individuals in the pictures were given facial expressions. All cartoons can be found in appendix A. Each of the cartoons was paired with a unique lexical item. Table 3.3 shows all lexical items for a PPQ in context 3 ($SB = n$, $CE = \neg p$). So, in order to manipulate the lexical items into the other contexts, we used the general noun highlighted in table 3.3 in line SB to represent a *neutral* SB or CE, the specific noun

Contxt		
1	SB:	I'm going to buy a car. I think a red one.
	CE:	Did you hear, John bought a car. I heard it's awesome!
2	SB:	I'm going to buy a car. I think a red one.
	CE:	Did you hear, John bought a BMW. I heard it's awesome!
3	SB:	I'm going to buy a car. I think a red one.
	CE:	Did you hear, John bought a Lexus. I heard it's awesome!
4	SB:	I'm going to buy a BMW. I think a red one.
	CE:	Did you hear, John bought a car. I heard it's awesome!
5	SB:	I'm going to buy a BMW. I think a red one.
	CE:	Did you hear, John bought a BMW. I heard it's awesome!
6	SB:	I'm going to buy a BMW. I think a red one.
	CE:	Did you hear, John bought a Lexus. I heard it's awesome!
7	SB:	I'm going to buy a Lexus. I think a red one.
	CE:	Did you hear, John bought a car. I heard it's awesome!
8	SB:	I'm going to buy a Lexus. I think a red one.
	CE:	Did you hear, John bought a BMW. I heard it's awesome!
9	SB:	I'm going to buy a Lexus. I think a red one.
	CE:	Did you hear, John bought a Lexus. I heard it's awesome!

Table 3.2: Example dialogues for each of the 9 contexts.

highlighted in line CE to represent *negative* SB or CE and the specific noun highlighted in line Q to represent *positive* SB or CE.

Additionally, we created 164 filler items which are designed in order to obscure the target items, with the same picture design. The filler items consisted of 41 different lexical items with wh-questions in the last picture to be rated. The fillers were balanced on the basis of positive and negative questions, natural and unnatural questions (according to our own judgement) and obvious and less obvious items. For an overview of the fillers, see appendix B. Table 3.4 presents examples of positive questions in an obviously natural context (a), an obviously unnatural context (b) a more or less natural context (c) and a more or less unnatural context (d), taken from different lexical items. The same examples occured with a negative question (the same number of high and low negation questions). Note that the names used in the different examples refer to the different cartoons.

3.2.3 Procedure

Each HIT contained one item, resulting in a total of 326 HITs, which were each completed by 25 different participants. The order of the target and filler was randomized and the presented HITs are alternated between filler and target item, with two subsequent filler items at the beginning. Because the target and filler items were presented alternately every subject (that saw more than one HIT) filled in about 50% fillers and 50% target items. The target items and

Lexical item		
1	SB:	I'm going to buy a car. I think a red one.
	CE:	Did you hear, John bought a Lexus. I heard it's awesome!
	Q:	Did he buy a BMW?
2	SB:	I'm going to watch a movie. I've never seen it before.
	CE:	Did you hear, Mike watched the Lion King. I heard he saw it twice in a row!
	Q:	Did he watch Star Wars?
3	SB:	I'm going to apply for college. I'm going to study medicine.
	CE:	Did you hear, Charles applied for Harvard. I heard he got accepted!
	Q:	Did he apply for Yale?
4	SB:	I'm going to visit my a family member. It's his birthday.
	CE:	Did you hear, Mickey visited his cousin. I heard he had a great time!
	Q:	Did he visit his uncle?
5	SB:	I'm going to go on holiday. I'm going with my boyfriend!
	CE:	Did you hear, Lucy went to Mexico. I heard she loved it!
	Q:	Did she go to Greece?
6	SB:	I'm going to get a pet. I've always wanted one.
	CE:	Did you hear, Kate got a dog. I heard it's so cute!
	Q:	Did she get a cat?

Table 3.3: Example dialogues of context 3 for each of the 6 lexical items

fillers were presented as a group of HITs, each of the HITs consisting of one target or filler item. Workers had the option after every HIT to continue with the next HIT or stop working, resulting in a high deviation between the number of HITs filled in per subject (minimum = 1; maximum= 326; average = 37). So, in effect different participants filled in different sets of HITs, with randomly different order of presentation of the conditions.

Participants were presented the following instructions:

- This task is for native speakers of English only;
- Based on the conversations in the first two pictures, rate the naturalness of the question in the third picture;
- "Completely natural" means that you could have used the question yourself in the given context;
- "Completely unnatural" means that you would protest if someone used the question in the given context.

The instructions were followed by two examples, one of a context with a completely natural question (36) and one of a context with a completely unnatural question (37):

Filler	
a	- I'm going to buy a new laptop. My old one broke down. - Did you hear, Mike bought a new laptop. I heard he bought a MacBook! - How much did he pay for it?
b	- I'm going to a great party tonight! It's going to be awesome! - Did you hear, Mickey partied all night. I heard he left very late. - What language does he speak?
c	- I'm going to move to Japan. I got a job there! - Did you hear, Kate is moving to Japan. I heard she's very excited. - How's the weather there?
d	- I'm going to be a stand-up comedian. Everyone thinks I'm very funny! - Did you hear, Mickey is going to be a stand-up comedian. I heard he's not funny at all! - Who told you he cares?

Table 3.4: Examples of an (a) obviously natural, (b) obviously unnatural, (c) more or less natural and (d) more or less unnatural filler.

(36) The question uttered by B is completely natural as a response to the sentence uttered by A:
A: I went to see a concert.
B: Which concert did you see?

(37) The question uttered by B is completely unnatural as a response to the sentence uttered by A:
A: I went to see a concert.
B: Which car did you wash?

These examples contain a wh-question and are thus more similar to the fillers than to the target items. This was done in order not to influence the participants' rating of the target items. The examples were followed by the cartoon and below a task description and a 7-point scale, where 1 meant 'completely natural' and 7 meant 'completely unnatural', as shown in (38).

(38) Rate the naturalness of the **question in the last picture** on the basis of the foregoing conversations.
Completely natural ◯ ◯ ◯ ◯ ◯ ◯ ◯ Completely unnatural

participants were given a maximum of 5 minutes to read the instructions and complete the task.

3.2.4 Statistical Analysis

The analysis of the results was done using SPSS statistics version 17.0 with standard statistical methods (Field, 2009). We used non-parametric methods to

analyze the data, as well as parametric methods based on rank transformations. Since non-parametric methods do not assume a normal distribution of the data and apply to the data on the basis of the rank order, these tests fitted the data we have. These methods are motivated by the fact that we have an ordinal scale: participants were forced to rate the naturalness on a 7-point scale, without any indication of the distance between the points on the scale. Furthermore, it motivates the use of the *median* (Mdn) rather than the *mean* to report main effects.

We used *Spearman's rank correlation coefficient*, also called Spearman's Rho, to calculate reliability of participants and lexical items. We used the *Mann-Whitney U test* to test whether two independent samples of observations had equally large values and to test multiple independent groups we used the *Kruskal-Wallis one-way analysis of variance* by ranks. We used the *Bonferroni correction* to correct for familywise error caused by multiple comparisons. Finally, a *factorial analysis of variance* - ANOVA (General Linear Model - GLM) was applied to the data transformed to ranks to test interactions.

3.3 Results

The participants completed the 326 HITs, such that each HIT was completed by 25 different participants; since each condition was presented using 6 lexical items, the total number of responses for each condition was $6 * 25 = 150$. Inspection of the data showed an overall non-normal distribution, as the result for condition 1 ($SB = n, CE = n, Q =$ PPQ) in figure 3.2 illustrates.

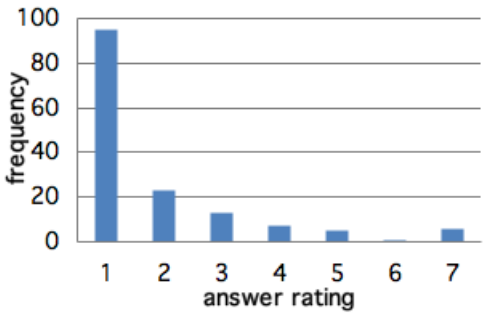


Figure 3.2: Frequency distribution Q1C1.

Taking all contexts together we see that the extreme values 1 and 7 were selected more often than the middle values. However, not all conditions showed a similar extreme value distribution; we will come back to this in the discussion section 3.4.

Table 3.5 reports the median (also known as the 50th percentile or second quartile (Q_2) because 50 percent of the observations fall below this value), as well as the 25th percentile (first quartile: Q_1) and the 75th percentile (third

(SB, CE)	Q_1	Q_2	Q_3
(n, n)	1	1	2
(n, p)	5	6	7
$(n, \neg p)$	3	5	7
(p, n)	1	1	1
(p, p)	5	6	7
$(p, \neg p)$	2	2	5
$(\neg p, n)$	1	3	5
$(\neg p, p)$	3	6	7
$(\neg p, \neg p)$	4	6	7

(a) PPQ

Q_1	Q_2	Q_3
1	2	4
6	7	7
2	3,5	6
1	1	2
5,25	7	7
1	1	2
2	4	6
6	6	7
3	5	7

(b) HNPQ

Q_1	Q_2	Q_3
2	4	6
6	7	7
3	4	6
2	3	4,75
6	7	7
1	1	3
3	5	6
6	7	7
3	6	7

(c) LNPQ

Table 3.5: First, second (median) and third quartiles for (a) PPQs, (b) HNPQs and (c) LNPQs.

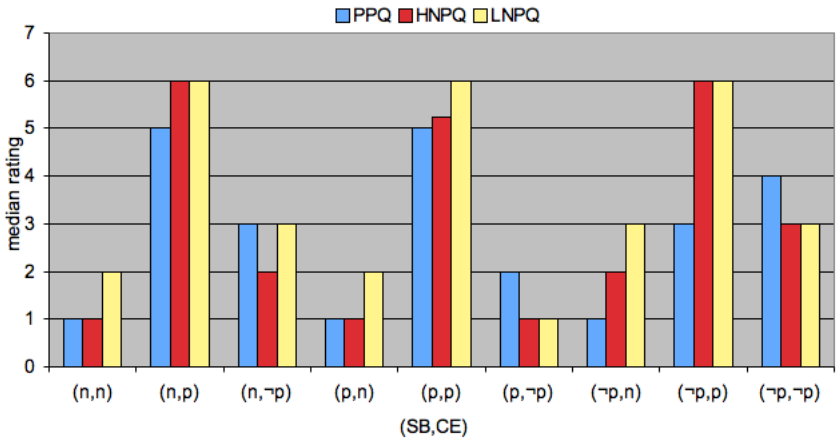


Figure 3.3: Median rating (Q_2) per question type per context.

quartile: Q_3). The median rating (Q_2) for each question type (PPQ, HNPQ and LNPQ) for each of the nine contexts is shown in figure 3.3.

In the following sections, we report the differences found in terms of the different theories proposed in section 3.1. In section 3.3.1 we test the reliability of the results by investigating the correlations between participants and between lexical items. The effects of speaker belief and contextual evidence are tested in section 3.3.2, focusing on theories 1 and 2. The effects of the individual contexts as well as theory 3 are investigated in section 3.3.3.

subject	HITs filled in	targets filled in	Spearman's ρ	significance
26	29	14	.436	$p <$.001
76	47	27	.499	$p <$.001
44	36	15	.564	$p <$.001
16	22	15	.676	$p <$ .002
103	66	33	.696	$p <$ .001

Table 3.6: Correlation results for participants with $\rho < .7$.

3.3.1 Reliability

3.3.1.1 Participants

Hypothesis. In order to test whether participants differed we tested the following null-hypothesis:

> H_0: There is no correlation between the HIT rating of one participant and the ratings of all other participants on the same HIT.

Method. The participants were tested for reliability in order to check whether participants had correctly interpreted the 7-point scale and had not randomly filled in the answers. This was tested by comparing all the answer ratings of one subject to the average ratings for each of the HITs, taken over the 25 different participants that filled in each HIT. Spearman's Rho correlations were computed for all participants that filled in at least 10 target items, such that we only tested participants that affected the overall results. The fillers were included in these correlations. Spearman's ρ is a value between -1 and $+1$, such that $+1$ indicates a perfect positive correlation and -1 a perfect negative correlation. This value is calculated using a function, which compares two samples on whether they are monotonically related (this is the case if for any pair of data values from the first and the second sample, the difference between the values of the former and the difference between the values of the latter coincide). Because this method doesn't require a linear correlation, like Pearson's correlation coefficient does, it is often used as a non-parametric method.

Results. All tested participants showed a significant positive correlation: for 58 of the 63 tested participants the results were such that Spearman's $\rho > .7$ and $p < .001$; the results for the remaining participants are shown in table 3.6, ordered according to increasing ρ. It follows that H_0 can be rejected.

Discussion. The number of filled in target items for both participants 26 and 76 with a relatively low correlation coefficient ($\rho < .5$) was below the average of 37, which means that their ratings do not have a large effect on the overall results. We thus chose not to exclude any participants from our analysis, since all participants showed similar ratings and can thus be considered reliable.

Lexical item	Spearman's ρ	significance
1	.643	$p <$.001
2	.675	$p <$.001
3	.674	$p <$.001
4	.699	$p <$ .001
5	.670	$p <$ .001
6	.679	$p <$ .001

Table 3.7: Correlation results for all lexical items.

3.3.1.2 Lexical Items

Hypothesis. In order to test whether the lexical items differed we tested the following null-hypothesis:

> H_0: There is no correlation between the ratings for one lexical item and the ratings of all other lexical items.

Method. We followed a similar strategy as with testing the reliability of the participants. We calculated Spearman's Rho for the correlation between each lexical item and the average rating over all items.

Results. Table 3.7 shows the results for all lexical items. For all lexical items Spearman's ρ was highly significant, with a value between .65 and .68. This means that all lexical items showed a correlation that is considerably higher than chance level (0) with the average rating overall, thus H_0 can be rejected.

Discussion. The lexical items were considered similar enough to treat the results as (relatively) independent with respect to the lexical items. That is, the similarities across the items were not accidental; the scores reflect similar judgments in all different items, so we can state that the ratings really test how good a given question type is in a given context, for all lexical items. In the discussion section 3.4 we will further investigate the individual differences between the lexical items.

3.3.2 Speaker Belief and Contextual Evidence

We now focus on the results for the target items. In this section we investigate the results of the experiment in terms of speaker belief (SB) and contextual evidence (CE). As formulated in theories 1 and 2 in section 3.1 above, we hypothesized that the acceptability of the different question types can be formulated in terms of speaker belief and contextual evidence; in particular, the question types are taken to depend on one of these, and HNPQs may depend on both speaker belief and contextual evidence.

We first test theory 1 in section 3.3.2.1 and accordingly theory 2 in section 3.3.2.2, which equals theory 1 whilst excluding rhetorical contexts 5 and

9. Then we investigate the different types of contextual evidence and speaker belief separately in section 3.3.2.3, in order to get a better understanding of the differences between the ratings for the different values of SB and CE. Remember that both SB and CE can have 3 values: *neutral* (n), *positive* (p) and *negative* ($\neg p$).

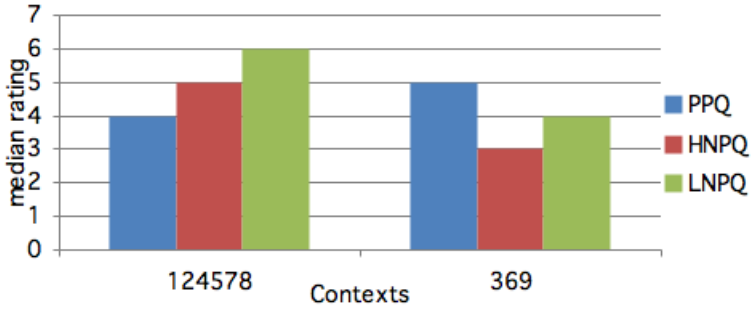
3.3.2.1 Theory 1

Hypotheses. On the basis of the hypotheses of theory 1 described above, we now formulate a set of null-hypotheses which we will try to reject.

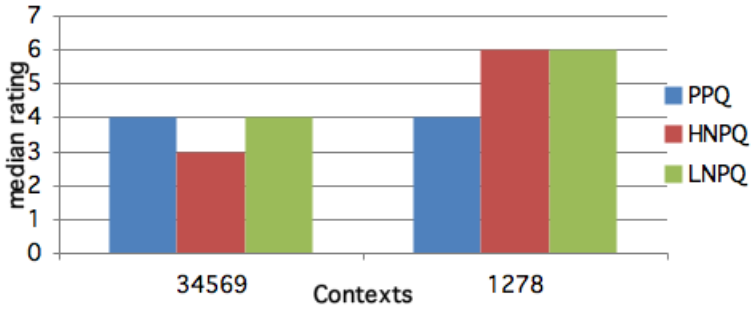
1. ${H_0}^{PPQ}$: There is no significant difference for PPQs between the ratings for the group of contexts 1, 2, 4, 5, 7 and 8 and the ratings for the group of contexts 3, 6 and 9.
2. ${H_0}^{LNPQ}$: There is no significant difference for LNPQs between the ratings for the group of contexts 1, 2, 4, 5, 7 and 8 and the ratings for the group of contexts 3, 6 and 9.
3. ${H_0}^{HNPQ}$: There is no significant difference for HNPQs between the ratings for the group of contexts 3, 4, 5, 6 and 9 and the ratings for the group of contexts 1, 2, 7 and 8.

Method. In order to test the first set of hypotheses, we tested for each question type whether there was a significant difference between contexts 3, 6, 9 and all other contexts; with this we tested hypotheses ${H_0}^{PPQ}$ and ${H_0}^{LNPQ}$, and the same one extended to HNPQs(i.e. "there is no significant difference for HNPQs between the ratings for the group of contexts 1, 2, 4, 5, 7 and 8 and the ratings for the group of contexts 3, 6 and 9"). Secondly, we compared for all question types whether there was a significant difference between contexts 3, 4, 5, 6, 9 and contexts 1, 2, 7, 8; with this we tested hypothesis ${H_0}^{HNPQ}$ and the same one extended to PPQs and LNPQs (i.e. "there is no significant difference for PPQs/LNPQs between the ratings for the group of contexts 3, 4, 5, 6 and 9 and the ratings for the group of contexts 1, 2, 7 and 8."). This results in a total of 6 comparisons. We used the *Bonferroni correction* to correct for familywise error caused by multiple comparisons. This method is based on the idea that the significance level of each individual hypothesis is $1/n$ times what it would be if only one hypothesis were tested, where n is the total number of comparisons. Thus, if we want to claim the truth of a set of comparisons, we need to adjust the significance level (α) for each of the comparisons by dividing the significance level by the number of comparisons. So, in the current case we have a significance level of $\alpha = .05/6 = .008$.

Figure 3.4(a) shows the median ratings for each question type in both groups. We compared for each question type whether there is a significant difference in rating between the two groups. We used a *Mann-Whitney test*, which is a non-parametric method that tests whether two independent samples of observations have equally large values.



(a) non-rhetorical contexts with $CE = n$ or $CE = p$ $(1, 2, 4, 5, 7, 8)$ versus contexts with $CE = \neg p$ $(3, 6, 9)$



(b) contexts with $SB = p$ or $CE = \neg p$ $(3, 4, 5, 6, 9)$ versus all other contexts $(1, 2, 7, 8)$

Figure 3.4: Median rating per question type for different groups of contexts.

So, for each question type we had two independent samples of answer ratings: all the answer ratings for contexts 3, 6 and 9, and those for all other contexts.

Results The Mann-Whitney test revealed that PPQs were rated significantly lower (and are thus more acceptable) in contexts with $CE = n$ or $CE = p$ than in contexts with $CE = \neg p$ $(U = 178044, p < .001, r = .10)$. This means that we can reject ${H_0}^{PPQ}$. LNPQs, on the contrary, were rated significantly lower in contexts with $CE = \neg p$ $(U = 131283, p < .001, r = .29)$, so ${H_0}^{LNPQ}$ was rejected. Also for HNPQs a significant difference was found: they are rated significantly lower in contexts with $CE = \neg p$ than in contexts with $CE = n$ or $CE = p$ $(U = 151217, p < .001, r = .21)$.

Figure 3.4(b) shows the median ratings for each question type in contexts with *positive SB* versus all other contexts.

We tested ${H_0}^{\text{HNPQ}}$ by comparing contexts with positive speaker belief or negative contextual evidence (contexts 3, 4, 5, 6, 9) to all other contexts for all question types. A Mann-Whitney test revealed that HNPQs were rated

significantly lower in contexts 3, 4, 5, 6 and 9 than in other contexts ($U = 150965, p < .001, r = .33$), thus ${H_0}^{HNPQ}$ had to be rejected. However, the same result was found for both PPQs ($U = 205654, p < .007, r = .09$) and LNPQs ($U = 155906, p < .001, r = .30$), which was not predicted by our hypothesis.

Discussion. Regarding the first two hypotheses we saw that ${H_0}^{PPQ}$ and ${H_0}^{LNPQ}$ were rejected as expected: while PPQs were more felicitous in contexts 1, 2, 4, 5, 7, and 8, LNPQs were more felicitous in contexts 3, 6 and 9. Moreover, we found an effect for HNPQs: they were more acceptable in contexts 3, 6 and 9 than in the other contexts. Since contexts 3, 6 and 9 are a subset of the contexts predicted to be felicitous for HNPQs (namely, contexts 3, 4, 5, 6 and 9), this was a partial confirmation for the hypothesis about the felicity of HNPQs.

When comparing the felicitous predicted contexts for HNPQs (3, 4, 5, 6 and 9) to all other contexts, we saw that indeed ${H_0}^{HNPQ}$ could be rejected because HNPQs were more felicitous in the predicted contexts than in the other contexts. Interestingly, however, the same effect was found for LNPQs and PPQs. This was not predicted by our hypotheses. Before making strong conclusions about these results, however, we will investigate whether the same results come out for theory 2.

3.3.2.2 Theory 2

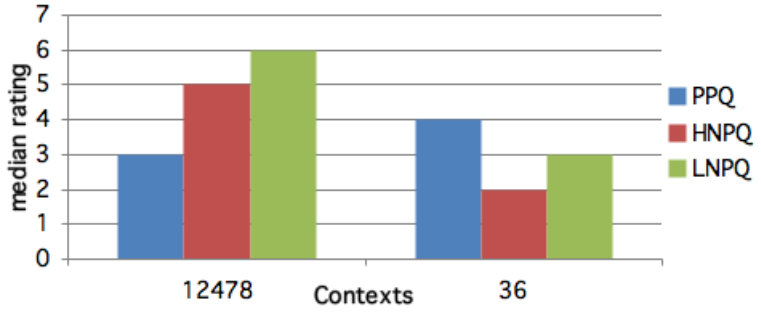
As described above, contexts 5 and 9 are considered to be rhetorical contexts because they are non-inquisitive due to the correspondence between SC and CE in these contexts (both are *positive* in context 5 and both are *negative* in context 9).

Hypotheses. On the basis of the informal hypotheses of theory 2 described above, we again formulate a group of null-hypotheses, which are the same as the hypotheses for theory 1 above, except that contexts 5 and 9 are left out.

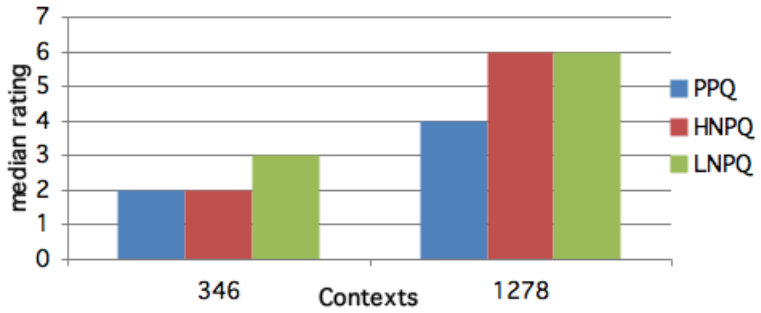
1. ${H_0}^{PPQ}$: There is no significant difference for PPQs between the ratings for the group of contexts 1, 2, 4, 7 and 8 and the ratings for the group of contexts 3 and 6.

2. ${H_0}^{LNPQ}$: There is no significant difference for LNPQs between the ratings for the group of contexts 1, 2, 4, 7 and 8 and the ratings for the group of contexts 3 and 6.

3. ${H_0}^{HNPQ}$: There is no significant difference for HNPQs between the ratings for the group of contexts 3, 4, and 6 and the ratings for the group of contexts 1, 2, 7 and 8.



(a) non-rhetorical contexts with $CE = n$ or $CE = p$ $(1, 2, 4, 7, 8)$ versus contexts with $CE = \neg p$ $(3, 6)$



(b) non-rhetorical contexts with $SB = p$ or $CE = \neg p$ $(3, 4, 6)$ versus all other non-rhetorical contexts $(1, 2, 7, 8)$

Figure 3.5: Median rating per question type for different groups of contexts.

Method. In order to test the first set of hypotheses, we again tested for each question type whether there was a difference between contexts 3 and 6 and all other contexts; with this we tested hypotheses ${H_0}^{PPQ}$ and ${H_0}^{LNPQ}$, and the same one extended to HNPQs. Secondly, we compared for all question types whether there is a difference between contexts 3, 4, 6 and contexts 1, 2, 7, 8; with this we test hypothesis ${H_0}^{HNPQ}$ and the same one extended to PPQs and LNPQs. Again we have a total of 6 comparisons, so applying the Bonferroni correction resulted in a significance level of $\alpha = .05/6 = .008$.

Results. Figure 3.5(a) shows the median ratings for each question type in both groups. The Mann-Whitney test revealed that PPQs were again rated significantly lower (and were thus more acceptable) in contexts with $CE = n$ or $CE = p$ than in contexts with $CE = \neg p$ $(U = 95534, p < .001, r = .12)$. This means that we had to reject ${H_0}^{PPQ}$. LNPQs, on the contrary, were still rated significantly lower in contexts with $CE = \neg p$ $(U = 64580, p < .001, r = .34)$, so we also rejected ${H_0}^{LNPQ}$. Also for HNPQs a significant difference was found:

they were rated significantly lower in contexts with $CE = \neg p$ than in contexts with $CE = n$ or $CE = p$ ($U = 76505, p < .001, r = .26$).

Figure 3.5(b) shows the median ratings for each question type in contexts with *positive SB* versus all other contexts. We tested $H_0{}^{\text{HNPQ}}$ by comparing non-rhetorical contexts with positive speaker belief or negative contextual evidence (3, 4, 6) to all other contexts (1, 2, 7, 8) for all question types. A Mann-Whitney test revealed that HNPQs indeed were rated significantly lower in contexts 3, 4 and 6 than in contexts 1, 2, 7 and 8 ($U = 60003, p < .001, r = .49$), thus $H_0{}^{HNPQ}$ had to be rejected. However, again we found the same result for both PPQs ($U = 108604, p < .001, r = .17$) and LNPQs ($U = 61335, p < .001, r = .47$).

Discussion. The overall results for theory 2 did not differ much from the results for theory 1, where contexts 5 and 9 were included; in both cases the null-hypotheses for all question types were rejected. Again we found that PPQs and LNPQs were significantly more felicitous in opposite contexts: PPQs in contexts 1, 2, 4, 7 and 8 and LNPQs in contexts 3 and 6. And again HNPQs were also more felicitous in contexts 3 and 6. This is not very surprising because taking out contexts 5 and 9 means taking out one context from each of the two compared groups. So, if we assume that there are no large differences between these contexts, then taking out these would not affect the overall differences much.

With respect to the comparisons made for $H_0{}^{HNPQ}$, taking out contexts 5 and 9 may in fact be predicted to have a significant effect. Instead of comparing contexts 3, 4, 5, 6 and 9 to contexts 1, 2, 7 and 8, we compared contexts 3, 4, 6 to contexts 1, 2, 7 and 8. However, we saw that we still found the same results as we had for theory 1 above: for all question types it was the case that they were more acceptable in contexts 3, 4 and 6 than in contexts 1, 2, 7 and 8. As described above, for HNPQs this was predicted. For LNPQs this was also not surprising, since the contexts in which they were predicted (and shown) to be felicitous, namely contexts 3 and 6, are a subset of the contexts tested here (contexts 3, 4, 6). For PPQs, however, we predicted them to be more felicitous in contexts 1, 2, 7 and 8 than in contexts 3, 4 and 6, since the first group of contexts consists of four contexts in which PPQs are predicted to be felicitous and the second group only contains one context in which they are predicted to be felicitous (context 4). It thus seems that there are still some contexts that behave differently from what we expected. In order to see which contexts exactly, in the next section we turn to an investigation of the individual contexts. Firstly, we discuss the results of the comparison of the different types of speaker belief and contextual evidence.

3.3.2.3 Effects of SB and CE

Now we investigate the different types of SB and CE separately in order to get a better understanding of where the effects come from.

Hypotheses. In order to get a better view of the influence of contextual evidence and speaker belief, we compared for each question type the ratings for

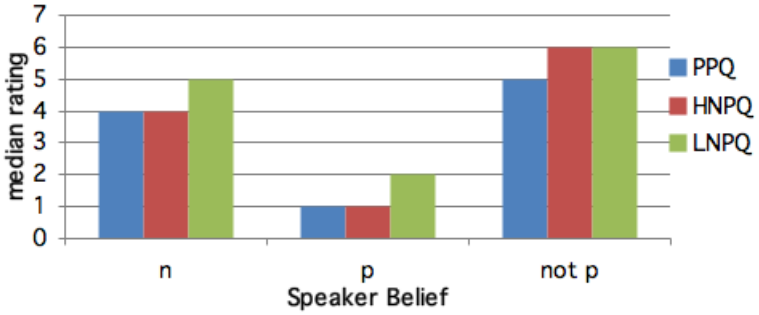


Figure 3.6: Median ratings per question type divided by speaker belief: n (contexts 1, 2, 3), p (contexts 4, 6) and $\neg p$ (contexts 7, 8)

all types of contextual evidence to each other and the ratings for all types of speaker belief to each other.

1. ${H_0}^{SB}$: For each of the question types, there are no significant differences between the ratings for each of the different SB values $n, p, \neg p$ (i.e. there is no significant difference between the ratings of contexts 1, 2, 3, the ratings of contexts 4, 6 and the ratings of contexts 7, 8).

2. ${H_0}^{CE}$: For each of the question types, there are no significant differences between the ratings for each of the different CE values $n, p, \neg p$ (i.e. there is no significant difference between the ratings of contexts 1, 4, 7, the ratings of contexts 2, 8 and the ratings of contexts 3, 6).

Method. We investigated the different types of speaker belief and contextual evidence separately, testing hypotheses ${H_0}^{SB}$ and ${H_0}^{CE}$. Because these hypotheses concern multiple groups (in terms of groups of contexts), we apply the *Kruskal-Wallis one-way analysis of variance* by ranks. This analysis tests the equality of population medians among multiple independent groups and is the non-parametric variant of the one-way analysis of variance. This analysis can be seen as a generalization of the Mann-Whitney test for multiple groups. So, if a significant difference between the groups is found, we still have no information about between which groups the differences were found. In order to obtain this result, we compare each two groups using a Mann-Whitney test.

Again because of multiple comparisons, we apply the Bonferroni correction. In order to test ${H_0}^{SB}$, we perform 3 Kruskall-Wallis analyses (comparing three SB values for each of the three question types), and we accordingly conduct 9 Mann-Whitney tests (comparing three SB values to each other for each question type). Similarly, in order to test ${H_0}^{CE}$ we conduct 3 Kruskall-Wallis analyses and 9 Mann-Whitney tests. So, we have 24 comparisons in total, which means that the significance level is adjusted as follows: $\alpha = .05/24 = .002$.

Question type	comparison	Result
PPQ	$SB = p$ - $SB = n$	$U = 40463, p < .001, r = .35$
	$SB = p$ - $SB = \neg p$	$U = 24334, p < .001, r = .37$
	$SB = n$ - $SB = \neg p$	$U = 63748, p = .189, r = .05$
HNPQ	$SB = p$ - $SB = n$	$U = 24910, p < .001, r = .55$
	$SB = p$ - $SB = \neg p$	$U = 11533, p < .001, r = .60$
	$SB = n$ - $SB = \neg p$	$U = 53034, p < .001, r = .19$
LNPQ	$SB = p$ - $SB = n$	$U = 30807, p < .001, r = .47$
	$SB = p$ - $SB = \neg p$	$U = 13347, p < .001, r = .55$
	$SB = n$ - $SB = \neg p$	$U = 54160, p < .001, r = .17$

Table 3.8: Results of Mann-Whitney tests for contexts per question type.

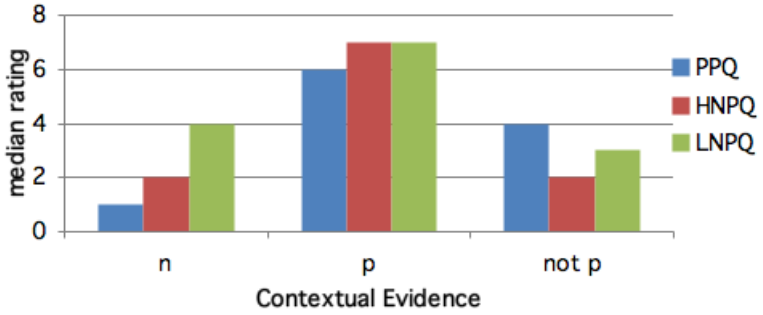


Figure 3.7: Median ratings per question type divided by contextual evidence: n (contexts 1, 4, 7), p (contexts 2, 8) and $\neg p$ (contexts 3, 6).

Results. Figure 3.6 shows the median ratings per question type for all types of speaker belief. The Kruskal-Wallis tests showed that the effect of the speaker belief was significant for all question types (PPQ: $H = 122.62, 2$ d.f., $p < .001$; HNPQ: $H = 324.45, 2$ d.f., $p < .001$; LNPQ: $H = 263.08, 2$ d.f., $p < .001$). This means that ${H_0}^{SB}$ had to be rejected.

In order to test between which of the groups (types of speaker belief) the significant differences are found, we conducted three Mann-Whitney tests per question type to compare the different groups (i.e. for each question type we compared the colons in figure 3.6). The results are shown in table 3.8. Thus, PPQs were rated significantly higher in contexts with *negative SB* than in contexts with *positive SB* and they were rated significantly higher in contexts with *negative SB* than in contexts with *neutral SB*. The difference between *positive* and *neutral SB* was not significant. For both HNPQs and LNPQs all types of speaker belief were significantly different, such that they were most acceptable in contexts with *positive SB* and least acceptable in contexts with *negative SB*.

Similarly, we tested the main effect of CE. Figure 3.7 shows the median ratings per CE type for all question types. The Kruskal-Wallis one-way analyses of variance showed that the effect of the contextual evidence was extremely

Question type	comparison	Result
PPQ	$CE = p$ - $CE = n$	$U = 18824, p < .001, r = .63$
	$CE = p$ - $CE = \neg p$	$U = 29009, p < .001, r = .28$
	$CE = n$ - $CE = \neg p$	$U = 34543, p < .001, r = .43$
HNPQ	$CE = p$ - $CE = n$	$U = 15588, p < .001, r = .66$
	$CE = p$ - $CE = \neg p$	$U = 10190, p < .001, r = .61$
	$CE = n$ - $CE = \neg p$	$U = 66315, p = .673, r = .02$
LNPQ	$CE = p$ - $CE = n$	$U = 23149, p < .001, r = .57$
	$CE = p$ - $CE = \neg p$	$U = 11202, p < .001, r = .60$
	$CE = n$ - $CE = \neg p$	$U = 53378, p < .001, r = .18$

Table 3.9: Results of Mann-Whitney tests for contexts per question type.

significant for all question types (PPQ: $H = 337.98, 2$ d.f., $p < .001$; HNPQ: $H = 396.90, 2$ d.f., $p < .001$; LNPQ: $H = 335.89, 2$ d.f., $p < .001$), so we had to reject ${H_0}^{CE}$.

Again we conducted Mann-Whitney tests to investigate which groups differed significantly from each other. The results are shown in table 3.9. This shows that PPQs were rated significantly lower in contexts with *neutral CE* than in contexts with *negative CE*, in contexts with *negative CE* they were rated significantly lower than in contexts with *positive CE*. For LNPQs we see that in contexts with *neutral CE* they were rated significantly higher than in contexts with *negative CE*, in contexts with *positive CE* they were again rated higher than in both other types of CE contexts. For HNPQs the difference between contexts with *neutral CE* and contexts with *negative CE* was not significant, but HNPQs were rated significantly lower in these contexts than in contexts with *positive CE*.

Discussion. As predicted, ${H_0}^{SB}$ and ${H_0}^{CE}$ were both rejected because for each question type significant differences were found between the different types of speaker belief and the different types of contextual evidence. However, the differences between the individual types of SB and CE were not as expected.

For the different values of SB we found that all question types exhibited the same overall pattern: they were rated lowest in contexts with *positive SB* and highest in contexts with *negative SB* (except that for PPQs the difference between *neutral* and *positive SB* was not significant). Although this pattern was expected for HNPQs, we saw that against our predictions the felicity of PPQs and LNPQs also depended on the speaker belief. With respect to contextual evidence we found that all question types were least acceptable in contexts with *positive CE*. When testing the individual contexts below we will investigate where this result comes from.

Furthermore, we saw that LNPQs were more acceptable in contexts with *negative CE* than in contexts with *neutral CE*, which agreed with our predictions. Conversely, for PPQs we saw that they were more felicitous in contexts with *neutral CE* than in contexts with *negative CE*, which also agreed with our

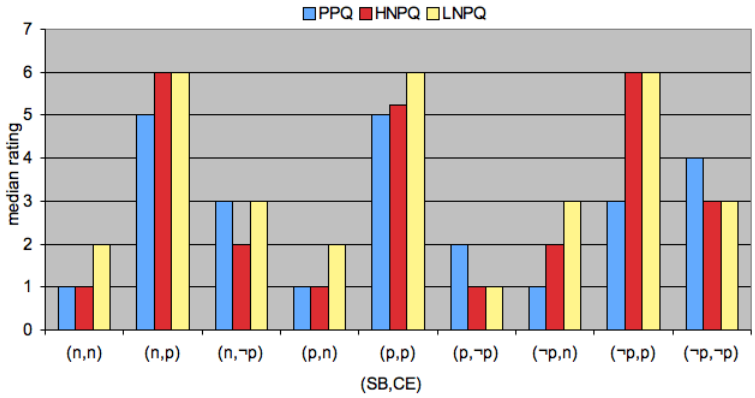


Figure 3.8: Median rating per question type per context overall.

predictions. The non-significant difference found for HNPQs between contexts with *neutral* and *negative CE* did not agree with our predictions: we predicted HNPQs to be more acceptable in contexts with *negative CE*. In order to see which contexts exactly cause this, we will turn to the individual contexts in section 3.3.3.

3.3.3 Contexts

In this section we investigate to what extent the acceptability rating for the different question types depends on the context, viewed as a combination of contextual evidence and speaker belief. So, we tested three question types in nine contexts, so we had a total of 27 conditions.

Firstly, we assume that the rating of a question depended on both the question type and the context in which it occurs. Thus, we assume there to be *main effects* of the context and the question type, as well as an *interaction* between those, i.e., the rating of a question in some context depends on the question type of the question and, the other way around, the rating of a certain question type depends on in which context it occurs. In other words, the effects of the context and the question type are predicted to be non-additive; the *interact*. Similarly, since we now focus on individual contexts (combinations of one type of speaker belief and one type of contextual evidence) instead of on the different values of speaker belief and contextual evidence separately, we predict there to be a main effect of and interaction between these two factors for each question type. These predictions will be tested in the section 3.3.3.1 below.

Secondly, we investigate the individual contexts and the relation to the different question types. Figure 3.8 repeats the graph shown in figure 3.3 and shows the median rating for each question type (PPQ, HNPQ and LNPQ) for each of the nine contexts. We will investigate the differences between the 27 conditions and report the results in the section 3.3.3.2 below.

Thirdly, we investigate the differences between the question types in terms of the contexts in which they occur in section 3.3.3.3. Here, we also investigate the predictions made in theory 3, which focus on typical contexts for each question type. Finally, the other way around, the differences between the contexts between the question types they license are investigated in section 3.3.3.4.

3.3.3.1 Interactions and Main effects

Hypotheses. We predict that there is an interaction between the question types and the contexts; this means that in order to predict the acceptability rating for a question, it is necessary to know both the question type and the context. We tested this on the basis of the following null-hypothesis.

1. ${H_0}^{Q,C}$: There is no interaction between the question type and context.

On the basis of an interaction between question types and contexts, thus, we predicted that for each question type the rating depended on both the speaker belief and the contextual evidence in a way that the effects of the two are not additive, but are needed both in order to predict the rating of a question type. We tested this on the basis of the following null-hypotheses.

1. ${H_{0:PPQ}}^{SB,CE}$: For PPQs there is no interaction between the factors SB and CE.
2. ${H_{0:HNPQ}}^{SB,CE}$: For HNPQs there is no interaction between the factors SB and CE.
3. ${H_{0:LNPQ}}^{SB,CE}$: For LNPQs there is no interaction between the factors SB and CE.

Method. We tested the interaction between context and question type. We used a factorial ANOVA, because we have a factorial design for each of the conditions. Because of the non-normal distribution of the results, we apply a rank transformation method to make the data more appropriate for a parametric analysis (Conover & Iman, 1981).

Results. The 9x3 factorial ANOVA on ranks with context ($1-9$) and question type (PPQ, HNPQ and LNPQ) as between-participants factors revealed a main effect of the context, $F(8, 4023) = 33.13, p < .001, {\eta_p}^2 = .062$, and a main effect of question type, $F(2, 4023) = 6.69, p = .001, {\eta_p}^2 = .003$. There was also found an interaction between context and question type, $F(16, 4023) = 11.16, p < .001, {\eta_p}^2 = .042$, so ${H_0}^{Q,C}$ had to be rejected. This means that in order to know the ranking of the question, it is necessary to both know the context and the question type. Levene's test showed that the variances of the populations from which the different samples were drawn were not equal ($p < .001$). Although this is in contrast with the ordinary assumptions for the ANOVA, because the group sizes are equal we may assume that the analysis

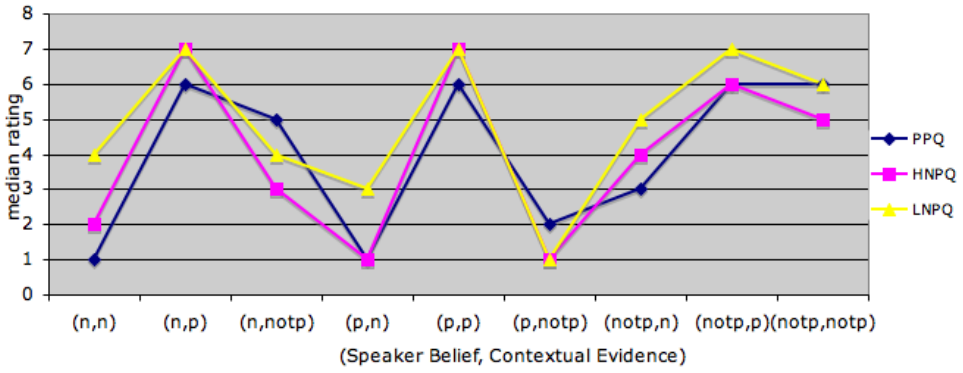


Figure 3.9: Interaction between context and question type.

is robust against heterogeneous variances (Stevens, 2009). The interaction is visualized in figure 3.9. The lines in the graph are not entirely parallel, since the differences between the lines differed on the different data points, e.g. the yellow line of LNPQs is in some contexts clearly above both other lines (for example in the context with $SB = p$ and $CE = n$), while in the context with $SB = n$ and $CE = \neg p$ the yellow line is below the blue line of PPQs, meaning that LNPQs were rated lower and thus more acceptable than PPQs. Thus, this means that the ratings depend on both the question type and the context.

For each question type the interaction between speaker belief and contextual evidence was tested. For PPQs the 3x3 factorial ANOVA on ranks with speaker belief and contextual evidence as between-participants factors revealed a main effect of SB, $F(2, 1043) = 5.879, p < .004, {\eta_p}^2 = .011$, and a main effect of CE, $F(2, 1043) = 15.327, p < .001, {\eta_p}^2 = .029$. There was also found an interaction between SB and CE for PPQs, $F(2, 1043) = 5.302, p < .006, {\eta_p}^2 = .010$, so ${H_{0:PPQ}}^{SB,CE}$ had to be rejected. Again we had equal group sizes and thus robustness against heterogeneous variances. This interaction is visualized in figure 3.10: again the lines are not parallel, e.g. between $CE = n$ and $CE = p$ the ratings for $SB = n$ and $SB = p$ coincide, but for $CE = \neg p$ we see that $SB = n$ is rated clearly higher than $SB = p$.

For HNPQs the 3x3 factorial ANOVA on ranks with speaker belief and contextual evidence as between-participants factors revealed a main effect of SB, $F(2, 1043) = 25.269, p < .001, {\eta_p}^2 = .046$, and a main effect of CE, $F(2, 1043) = 46.578, p < .001, {\eta_p}^2 = .082$. For HNPQs no interactions were found between SB and CE, $F(2, 1043) = 2.015, p = .134, {\eta_p}^2 = .004$, so ${H_{0:HNPQ}}^{SB,CE}$ could not be rejected. Again we had equal group sizes and thus robustness against heterogeneous variances.

For LNPQs the 3x3 factorial ANOVA on ranks with speaker belief and contextual evidence as between-participants factors revealed a main effect of SB, $F(2, 1043) = 3.724, p < .025, {\eta_p}^2 = .007$, and a main effect of CE, $F(2, 1043) =$

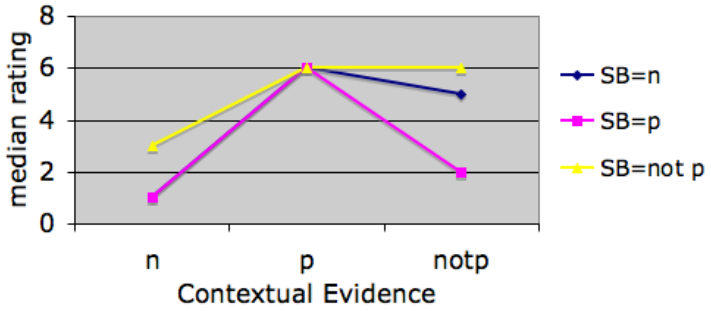


Figure 3.10: Interaction between speaker belief and contextual evidence for PPQs.

$74.558, p < .001, {\eta_p}^2 = .125$. There was also found an interaction between SB and CE for LNPQs, $F(2, 1043) = 8.152, p < .001, {\eta_p}^2 = .015$, so ${H_{0:LNPQ}}^{SB,CE}$ had to be rejected. Again we had equal group sizes and thus robustness against heterogeneous variances.

Discussion. As predicted, we found an interaction between the question type and the context in which the question occurs. This motivated us to test the differences between the question types in terms of the contexts in which they were acceptable and the differences between the contexts in terms of the question types they license, which is done in the sections below.

Note that figure 3.9 shows the median ratings, and not the ranking of the answers on which the analysis was based; this was done to provide a more intuitive display of the results. Moreover, the presented order of the conditions is, although not arbitrary, not crucial; these results can be used to compare any pair of contexts. Similarly, in figure 3.10 the choice to represent contextual evidence on the x-axis and the types of SB as different series is not crucial here: we would also discover an interaction had we presented it the other way around.

For PPQs and LNPQs we found an interaction between SB and CE. For HNPQs no interaction was found. However, we predicted that the felicity of HNPQs may depend on both the speaker belief and the contextual evidence, because it is licensed in contexts with $SB = p$ *or* $CE = \neg p$. Thus, the fact that we found no interaction between SB and CE for HNPQs does not mean that these factors were not both crucial for the acceptability of HNPQs; it merely means that knowing only one of them may suffice to predict the acceptability. In the following sections we will investigate in more detail what determines the acceptability of each of the question types.

SB \ CE	neutral	p	¬p
neutral	1 PPQ	2 PPQ	3 HNPQ LNPQ
p	4 PPQ HNPQ	5 PPQ HNPQ	6 HNPQ LNPQ
¬p	7 PPQ	8 PPQ	9 HNPQ LNPQ

Table 3.10: Theoretical acceptability of question types per context.

3.3.3.2 Overall effects

Hypotheses. With respect to the overall effects, we hypothesize that there are significant differences between the question types: as table 3.1, repeated here as table 3.10, illustrates, we predict question types to be felicitous in a different number of contexts; LNPQs in three (3, 6, 9), HNPQs in five (3, 4, 5, 6, 9) and PPQs in six contexts (1, 2, 4, 5, 7, 8) in total. This would result in an overall higher rating for LNPQs than both PPQs and HNPQs, and an overall lower rating for PPQs than HNPQs. Moreover, as a result of these differences we predict a significant difference between the different contexts. We already explained that contexts 5 and 9 are predicted to get a higher rating than other contexts because of their rhetorical nature. Furthermore, contexts in which multiple question types are predicted felicitous, are predicted to get an overall lower rating. This was tested on the basis of the hypotheses below:

1. ${H_0}^{Qtypes}$: There is no significant difference between the ratings for PPQs, HNPQs and LNPQs over all contexts.

2. ${H_0}^{contexts}$: There is no significant difference between the ratings for contexts 1-9 over all question types.

Results. We tested the differences between the three question types. A Kruskal-Wallis one-way analysis of variance indicated that significant differences were found in the rating of the different question types ($H = 54.61$, 2 d.f., $p < .001$). So we rejected ${H_0}^{Qtypes}$. Mann-Whitney tests with Bonferroni correction ($\alpha = .05/3 = .02$) indicated that the ratings for the PPQ and the HNPQ did not differ significantly ($U = 882867, p = .154, r = .03$), while the ratings for LNPQs were significantly higher than the ratings for both PPQs ($U = 770855, p < .001, r = .14$) and HNPQs ($U = 801796$, $p < .001, r = .11$).

Accordingly, we compared the different contexts, testing ${H_0}^{contexts}$. A Kruskall-Wallis one-way analysis of variance indicated that also the nine contexts differ significantly from each other ($H = 1581,8$ d.f., $p < .001$). So we also rejected ${H_0}^{contexts}$. Again we performed Mann-Whitney tests in order to

see where exactly the significant differences occur. We had 9 contexts that were each compared to each other so we had a total of 36 comparisons, which with the Bonferroni correction resulted in a significance level $\alpha = .05/36 = .001$. Table 3.11 shows the results for the Mann-Whitney tests for all comparisons.

All contexts differed significantly from each other except for contexts 2 and 5; 2 and 8; 5 and 8; 3 and 7; and 4 and 6. The latter, we must note, shows a trend towards significance ($p = .006$) since it is not significant because we chose to apply the Bonferroni correction. As figure 3.3 demonstrates, contexts 2, 5 and 8 are rated high for all question types. A Mann-Whitney test comparing these contexts to the highest rated other context (context 9) showed that contexts 2, 5 and 8 are indeed rated significantly higher than context 9 ($U = 681374, p < .001, r = .52$), and thus higher than all other contexts.

Discussion. We hypothesized that LNPQs would overall be rated higher than HNPQs and PPQs because they are predicted to be acceptable in less contexts; this was indeed found in the experiment. However, HNPQs and PPQs did not differ in their overall rating. Although we predicted that they would differ, this was based on a divergence of one context (six felicitous contexts for PPQs in total versus five felicitous contexts for HNPQs in total), so the fact that they were not significantly different is not very surprising. In the results for group 2 we indeed found that PPQs and HNPQs did differ in terms of the contexts they are licensed by.

With respect to the contexts we hypothesized that contexts 5 and 9 were overall least acceptable (highest rating) because they are considered rhetorical contexts. Furthermore, we predicted that contexts 3, 4 and 6 were overall rated lowest, because they are all predicted to license multiple question types, in contrast to contexts 1, 2, 7 an 8. We found that contexts 4 and 6 were indeed rated lowest, followed by context 1 and then by contexts 3 and 7. The higher rating for context 3 may be explained by the fact that, in contrast with contexts 4, 6 and 1, context 3 is not predicted to be a typical context for any question type; this investigated in the next section. Against our predictions, contexts 2, 5 and 8 did not differ significantly from each other in terms of the overall rating of the question types and were overall rated higher than the other contexts. Notably, contexts 2, 5 and 8 are exactly the contexts with *positive CE*. The high, and thus 'bad' rating of these contexts may be explained by the fact that the information provided by the contextual evidence was explicitly repeated in the question. This is illustrated in (39), which is an instance of context 2, with lexical item 3.

(39) Charles to Ann: I'm going to apply for college. I'm going to study medicine.
Brian to Ann: Did you hear, Charles applied for Yale. I heard he got accepted!
Ann to Brian: Did he apply for Yale?

Although the PPQ in (39) was predicted to be acceptable in the given context, it

Context	1	2	3	4	5	6	7	8
2	U=24200, $p < .001$, $r = .67$							
3	U=63513, $p < .001$, $r = .33$	U=51291, $p < .001$, $r = .44$						
4	U=77506, $p < .001$, $r = .21$	U=13086, $p < .001$, $r = .77$	U=42373, $p < .001$, $r = .52$					
5	U=20924, $p < .001$, $r = .70$	U=96351, $p = .173$, $r = .05$	U=46805, $p < .001$, $r = .48$	U=10988, $p < .001$, $r = .79$				
6	U=88034, $p < .001$, $r = .12$	U=19464, $p < .001$, $r = .71$	U=52326, $p < .001$, $r = .43$	U=91275, $p = .006$, $r = .09$	U=16960, $p < .001$, $r = .74$			
7	U=70801, $p < .001$, $r = .26$	U=50083, $p < .001$, $r = .45$	U=96348, $p = .204$, $r = .04$	U=51128, $p < .001$, $r = .44$	U=45898, $p < .001$, $r = .49$	U=60514, $p < .001$, $r = .36$		
8	U=27803, $p < .001$, $r = .64$	U=99113, $p = .556$, $r = .02$	U=54881, $p < .001$, $r = .41$	U=16141, $p < .001$, $r = .74$	U=94456, $p = .060$, $r = .06$	U=22497, $p < .001$, $r = .69$	U=53268, $p < .001$, $r = .42$	
9	U=49019, $p < .001$, $r = .45$	U=69514, $p < .001$, $r = .28$	U=83030, $p < .001$, $r = .16$	U=31208, $p< .001$, $r = .61$	U=64910, $p < .001$, $r = .32$	U=39891, $p < .001$, $r = .53$	U=79380, $p < .001$, $r = .19$	U=72656, $p < .001$, $r = .25$

Table 3.11: Overall comparisons of all contexts.

sounds odd because of the repetition of information.[2] The conversation between Brian and Ann is likely to proceed by Brian saying: "Duh, that's what I just said!". This can be seen as a refusal to accept the question, and since participants were instructed to rate questions they would refuse to accept as 'unnatural', this may explain the high rating for contexts 2, 5 and 8.

This doesn't mean that contexts with *positive CE* are always problematic, though. When contextual evidence is, for example, constituted extra-linguistically, the repetitive effect is less likely to occur. Future work may focus on testing different question types in contexts with positive extra-linguistic contextual evidence.

3.3.3.3 Contexts per Question type

Hypotheses. By looking at the different contexts for each question type we investigated whether, apart from the overall differences, the acceptability for the different contexts differed between the question types. It is clear from the predicted felicity pattern shown in table 3.10 that we indeed hypothesize that there were significant differences. Moreover, in section 3.1 theory 3 stated that the question types differ in terms of their most typical context. This was also tested below, in terms of the following set of hypotheses:

1. $H_0{}^{\text{PPQ}}$: For PPQs there are no significant differences between the ratings for the 9 different contexts.
    - $H_0{}^{\text{PPQ}C1}$: For PPQs there is no significant difference between the rating for context 1 and the rating for all other contexts.
2. $H_0{}^{\text{HNPQ}}$: For HNPQs there are no significant differences between the ratings for the 9 different contexts.
    - $H_0{}^{\text{HNPQ}C6}$: For HNPQs there is no significant difference between the rating for context 6 and the rating for all other contexts.
3. $H_0{}^{\text{LNPQ}}$: For LNPQs there are no significant differences between the ratings for the 9 different contexts.
    - $H_0{}^{\text{LNPQ}C3}$: For LNPQs there is no significant difference between the rating for context 3 and the rating for all other contexts.

Method. Because a comparison between all contexts for all question types would result in a very large number of comparisons ($3 * 36 = 108$) and thus a very large Bonferroni correction, we select a subset of the contexts to compare to each other in order to maintain reasonable power. Motivated by the fact that contexts 2, 5 and 8 do not differ significantly from each other and the overall

[2]The question can be acceptable in the given context if it has a distinct intonation pattern which, for example, expresses surprise: "Did he apply for *Yale*?". Since no intonation was provided we cannot say whether participants took this possibility into account. We will come back to this point in section 3.4.

Context	1	3	4	6	7
3	U = 3274, p < .001, r = .63				
4	U = 9268, p = .001, r = .19	U = 2118, p < .001, r = .74			
6	U = 6298, p < .001, r = .40	U = 6942, p < .001, r = .34	U = 4376, p < .001, r = .57		
7	U = 7061, p < .001, r = .34	U = 7341, p < .001, r = .30	U = 5589, p < .001, r = .48	U = 11133, p = .874, r = .01	
9	U = 2447, p < .001, r = .70	U = 10085, p = .114, r = .09	U = 1429, p < .001, r = .79	U = 5800, p < .001, r = .42	U = 6340, p < .001, r = .38

Table 3.12: Comparisons of contexts 1, 3, 4, 6, 7 and 9 for PPQs.

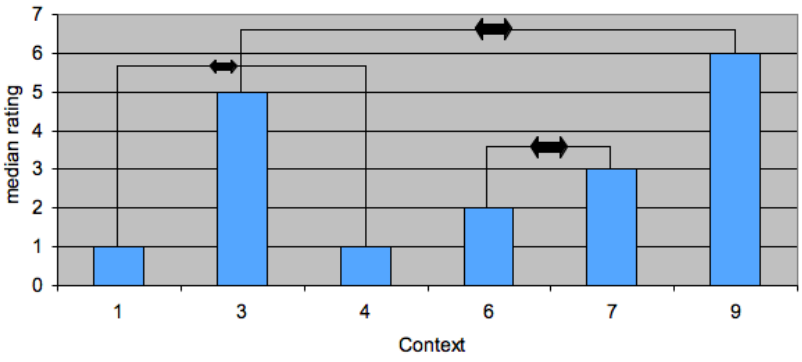


Figure 3.11: Median rating and non-significant differences for PPQs.

higher rating of these contexts, as the Mann-Whitney tests above revealed, these contexts are excluded from the current analysis. So, we compare six contexts to each other for each question type, resulting in $3 * 15 = 45$ comparisons, so the Bonferroni correction results in a significance level $\alpha = .05/45 = .001$.

Results. Table 3.12 shows the results of the comparisons for PPQs and figure 3.11 illustrates the differences between the contexts.

In figure 3.11, the arrow bridges represent a non-significant difference between the colons; smaller arrows represent a smaller significance and thus a lesser similarity between the colons. We chose to represent the non-significant differences rather than the significant differences to increase the clarity of the figure. For PPQs there is a clear distinction between three levels of rating: contexts 1 and 4 were rated low, contexts 6 and 7 were rated averagely high and

Context	1	3	4	6	7
3	U = 8351, p < .001, r = .23				
4	U = 6679, p < .001, r = .37	U = 4546, p < .001, r = .54			
6	U = 7052, p < .001, r = .34	U = 4898, p < .001, r = .51	U = 10973, p = .665, r = .02		
7	U = 7978, p < .001, r = .26	U = 10623, p = .398, r = .05	U = 4707, p < .001, r = .53	U = 5011, p < .001, r = .50	
9	U = 6050, p < .001, r = .40	U = 8671, p = .001, r = .20	U = 3345, p < .001, r = .63	U = 3627, p < .001, r = .61	U = 9464, p = .016, r = .14

Table 3.13: Comparisons of contexts 1, 3, 4, 6, 7 and 9 for HNPQs.

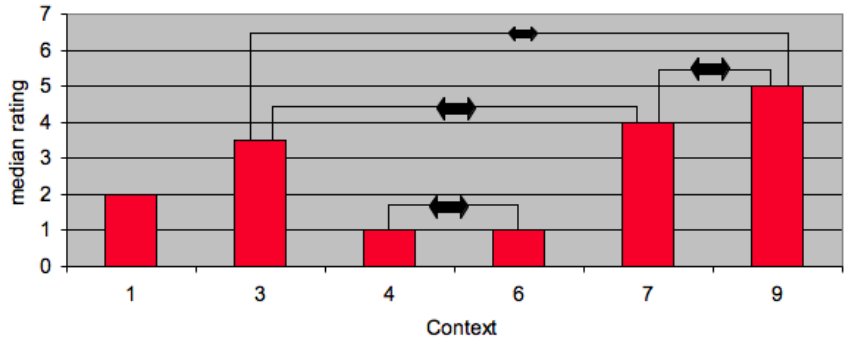


Figure 3.12: Median rating and non-significant differences for HNPQs.

contexts 3 and 9 were rated high. This means that PPQs were most acceptable in contexts 1 and 4 and least acceptable in contexts 3 and 9. So, ${H_0}^{\text{PPQ}}$ has to be rejected and ${H_0}^{\text{PPQ}C1}$ has to be partly rejected because there is no significant difference between contexts 1 and 4 and thus context 1 is not the only typical context for PPQs.

Table 3.13 and figure 3.12 show the results for HNPQs. Contexts 3, 7 and 9 were all rated equally high, while context 1 was rated significantly lower than each of these and contexts 4 and 6 were rated lowest. In other words, HNPQs were most acceptable in contexts 4 and 6 and least acceptable in contexts 3, 7 and 9. Thus, HNPQs differed from PPQs after all; both were least acceptable in (at least) contexts 3 and 9, but they differed in terms of the most acceptable contexts, which for PPQs were contexts 1 and 4, while for HNPQs, these were contexts 4 and 6. So, ${H_0}^{\text{HNPQ}}$ had to be rejected and ${H_0}^{\text{HNPQ}C6}$ had to be

Context	1	3	4	6	7
3	U = 10375, p = .239, r = .07				
4	U = 8799, p = .001, r = .19	U = 7870, p < .001, r = .26			
6	U = 6001, p < .001, r = .41	U = 5189, p < .001, r = .47	U = 7438, p < .001, r = .30		
7	U = 8451, p < .001, r = .22	U = 9287, p = .008, r = .15	U = 6000, p < .001, r = .41	U = 3896, p < .001, r = .58	
9	U = 8205, p < .001, r = .24	U = 9045, p = .003, r = .17	U = 5981, p < .001, r = .41	U = 3733, p < .001, r = .59	U = 10834, p = .573, r = .03

Table 3.14: Comparisons of contexts 1, 3, 4, 6, 7 and 9 for LNPQs.

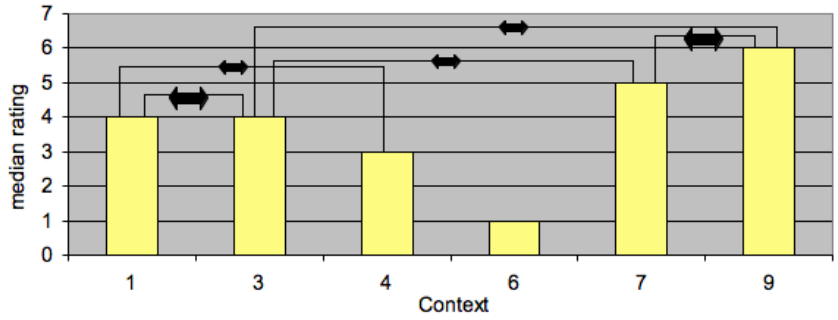


Figure 3.13: Median rating and non-significant differences for LNPQs.

partly rejected because context 6 was not the only typical context for HNPQs.

Table 3.14 and figure 3.13 show the results for LNPQs. For LNPQs it holds that context 6 was rated significantly lower than all other contexts and was thus the context in which LNPQs were the most acceptable. Context 4 showed a trend towards significant difference from context 1 ($p = .001$) and contexts 7 and 9 were the least acceptable contexts for LNPQs. So, ${H_0}^{\text{LNPQ}}$ had to be rejected, but ${H_0}^{\text{LNPQ}C3}$ could not; against our predictions LNPQs were not typical in context 3 but in context 6.

Discussion. With respect to the different question types we have found that PPQs, HNPQs and LNPQs indeed differ in terms of the contexts in which they occur. Table 3.15 summarizes the observed acceptability for each question type. It follows from the most acceptable contexts for PPQs, that ${H_0}^{\text{PPQ}C1}$ was partly rejected: PPQs were most acceptable context 1, but also in context 4 since

Question type	Most acceptable in context(s)	Middle acceptable in context(s)	Least acceptable in contexts
PPQ	1, 4	6, 7	3, 9
HNPQ	4, 6	1	3, 7, 9
LNPQ	6	4, 1, 3	7, 9

Table 3.15: Acceptable contexts per question type

contexts 1 and 4 were not significantly different for PPQs. When looking at the most acceptable and middle acceptable contexts, we found that PPQs are acceptable in all contexts with *neutral CE* (1, 4, 7). Because we have excluded contexts with *positive CE* from the analysis, we cannot make any statements about these contexts. Surprisingly, PPQs were rated significantly lower in context 6 than in context 3. However, the analysis of question types per context showed that in both contexts PPQs were the least acceptable question types, thus it seems that this effect is based on the acceptability of the contexts relative to each other, rather than the acceptability of the PPQ in each of the contexts.

For HNPQs we also saw that ${H_0}^{\text{HNPQ}C6}$ was partly rejected, since context 4 was indeed a most acceptable context, but there was no significant difference between context 4 and context 6. However, this confirmed the prediction that HNPQs were (most) acceptable in contexts with *positive SB*. The fact that HNPQs are middle acceptable in context 1 is difficult to explain. We suggest that this is due to the fact that the contexts, as presented in each HIT, left space for filling in of additional speaker belief; because the cartoons were presented without any further explanation about the speaker beliefs and events in between the pictures (in particular between the first and second picture), we suspect that some participants allowed for implicit speaker belief. For example, consider the instantiation of lexical item 6 of a HNPQ in context 1:

(40) i. I'm going to get a pet.
ii. Did you hear, Kate got a pet.
iii. Didn't she get a cat?

In this context we can imagine that the speaker of the question has a belief that Kate is a 'cat-person' so would be more likely to take a cat as pet than a dog. So even though the context provides no such speaker belief, participants may have based their answer on the reading in which they 'add' speaker belief. Because context 1 has both *neutral SB* and *neutral CE*, this context is taken to be more susceptible for this effect. However, since participants were left no space to comment on what they based their answer, these suspicions cannot be substantiated.

For LNPQs we saw that ${H_0}^{\text{LNPQ}C3}$ could not be rejected: they were most acceptable in context 6. Furthermore, all other contexts were not significantly different from at least one of the others. However, for LNPQs context 3 was a middle acceptable context, whereas for PPQs and HNPQs it was a least acceptable context. This can be seen as a confirmation of the prediction that LNPQs are

most felicitous in contexts with *negative CE*. However, the average acceptability of context 4 again suggests that LNPQs may also be influenced by the speaker belief, resulting in a higher acceptability in contexts with *postive SB*.

In order to put these results in perspective, we now investigate the difference between the contexts in terms of the acceptability of the different question types for each context.

3.3.3.4 Question types per Context

Hypotheses. In order to test the predictions made in table 3.10, we investigated for each context which question types were licensed. On the basis of table 3.10 we predict that for all question types there are significant differences between the question types, and that in contexts where 2 question types occur, these do not differ significantly. Since the null-hypotheses for all contexts coincide, we summarize these using one hypothesis, where x can be replaced by any context number between 1 and 9:

x. ${H_0}^{Cx}$: For context x there is no significant difference between the rating for the question types PPQ, LNPQ and HNPQ.

Method. We compared for each context what was the most acceptable question type. We have already seen that contexts 2, 5 and 8 do not differ significantly in terms of answer rating, nor do contexts 3 and 7 and contexts 4 and 6. However, these results were based on the overall rating; they may still differ in terms of what question types they license. For each context, a Kruskal-Wallis one-way analysis of variance was performed to check whether the question types differed significantly in their rating ($\alpha = .05/9 = .006$).

For contexts 1, 2, 3, 4, 6, 7 and 8 we performed Mann-Whitney tests between all question types, to see which question types were most acceptable. This resulted in $7 * 3 = 21$ comparisons, so the significance level with the Bonferroni correction is $\alpha = .05/21 = .002$.

Results. This resulted in significant differences for contexts 1 ($H = 94.4$, 2 d.f., $p < .001$), 2 ($H = 19.3$, 2 d.f., $p < .001$), 3 ($H = 15.3$, 2 d.f., $p < .001$) and 4 ($H = 143.3$, 2 d.f., $p < .001$). Also, significant differences were found for contexts 6 ($H = 51.5$, 2 d.f., $p < .001$), 7 ($H = 30.8$, 2 d.f., $p < .001$) and 8 ($H = 25.5$, 2 d.f., $p < .001$). However, there is no significant difference between the rating of the different question types for contexts 5 ($H = 6.26$, 2 d.f., $p = .044$) and 9 ($H = 2.62$, 2 d.f., $p = .270$). So, all null-hypotheses of group 3 except for ${H_0}^{C5}$ and ${H_0}^{C9}$ had to be rejected.

Table 3.16 shows the results of the Mann-Whitney tests for contexts 1, 2, 3, 4, 6, 7 and 8 for each comparison of question types.

For context 1 all question types differed significantly, such that PPQ was the most acceptable and LNPQ the least acceptable. For context 2, there was no significant difference between HNPQs and LNPQs, but PPQs were rated lower than both these question types. For context 4 PPQs and HNPQs were not significantly

Context	Compare	Results		
1	PPQ - HNPQ	U = 7080,	p < .001,	r = .34
	PPQ - LNPQ	U = 4473,	p < .001,	r = .54
	HNPQ - LNPQ	U = 7743,	p < .001,	r = .27
2	PPQ - HNPQ	U = 8822,	p < .002,	r = .20
	PPQ - LNPQ	U = 8424,	p < .001,	r = .23
	HNPQ - LNPQ	U = 10832,	p = .539,	r = .04
3	PPQ - HNPQ	U = 8456,	p < .001,	r = .22
	PPQ - LNPQ	U = 9467,	p = .016,	r = .14
	HNPQ - LNPQ	U = 9941,	p = .078,	r = .1
4	PPQ - HNPQ	U = 9563,	p = .004,	r = .16
	PPQ - LNPQ	U = 3469,	p < .001,	r = .64
	HNPQ - LNPQ	U = 5058,	p < .001,	r = .5
6	PPQ - HNPQ	U = 6252,	p < .001,	r = .4
	PPQ - LNPQ	U = 7931,	p < .001,	r = .26
	HNPQ - LNPQ	U = 9551,	p = .012,	r = .15
7	PPQ - HNPQ	U = 9268,	p = .007,	r = .16
	PPQ - LNPQ	U = 7104,	p < .001,	r = .32
	HNPQ - LNPQ	U = 9192,	p = .006,	r = .16
8	PPQ - HNPQ	U = 8344,	p < .001,	r = .24
	PPQ - LNPQ	U = 8072,	p < .001,	r = .26
	HNPQ - LNPQ	U = 10742,	p = .457,	r = .04

Table 3.16: Results

different, and LNPQs were rated higher than both. For context 6, HNPQs and LNPQs were not significantly different but were both rated lower than PPQs. In context 7 the only significant difference was found between PPQs and LNPQs: PPQs were rated lower than LNPQs. The differences between PPQs and HNPQs and between HNPQs and LNPQs were not significant, because we chose to correct the significance levels on the basis of the Bonferroni correction. Finally, in context 8 PPQs were rated lower than both HNPQs and LNPQs, which did not differ significantly.

Discussion. Table 3.17 again repeats the predictions for PPQs, HNPQs and LNPQs from table 3.1. Table 3.18 shows the observed ranking of acceptability of all question types per context; ranking 1 means that the question type was the most acceptable question type in the given contexts, and question types of which the rating was not significantly different from each other get the same ranking (the ranking $1\frac{1}{2}$ means that there was no significant difference with the question type with ranking 1 nor with the question type with ranking 2).

In contexts 5 and 9 there was no difference found between the ratings for the different question types; they were all rated very high. We saw above that for all question types context 9 was part of the least acceptable contexts; context 5 was not considered here but was already shown to be rated significantly higher than all other contexts. This is in agreement with our analysis of contexts 5 and 9 as rhetorical contexts: when SB and CE coincide, the context is non-inquisitive

SB \ CE	neutral	p	¬p
neutral	PPQ	PPQ	HNPQ LNPQ
p	PPQ HNPQ	PPQ HNPQ	HNPQ LNPQ
¬p	PPQ	PPQ	HNPQ LNPQ

Table 3.17: Theoretically predicted acceptability of question types per context.

SB \ CE	neutral	p	¬p
neutral	1. PPQ 2. HNPQ 3. LNPQ	1. PPQ 2. HNPQ 2. LNPQ	1. HNPQ $1\frac{1}{2}$. LNPQ 2. PPQ
p	1. PPQ 1. HNPQ 2. LNPQ	1. PPQ 1. HNPQ 1. LNPQ	1. HNPQ 1. LNPQ 2. PPQ
¬p	1. PPQ $1\frac{1}{2}$. HNPQ 2. LNPQ	1. PPQ 2. HNPQ 2. LNPQ	1. PPQ 1. HNPQ 1. LNPQ

Table 3.18: Observed ranking of question types per context

and thus none of the question types is acceptable.

When comparing the predicted acceptable question types shown in table 3.17 and the observed highest ranked question types in table 3.18 (i.e. ranks 1 and $1\frac{1}{2}$) for all contexts with significant differences between the question types, we found remarkable correspondences. As predicted in table 3.17, PPQs were among the most acceptable question types in contexts 1, 2, 4 and 7. Also as predicted, LNPQs were among the most acceptable question types in contexts 3 and 4. Moreover, HNPQs were indeed among the most acceptable question types in contexts 3, 4 and 6. Contrary to our predictions, HNPQs were also one of the most acceptable question types in context 7. However, as we have seen, the comparison between PPQs and HNPQs in context 7 was not significant due to the Bonferroni correction and thus showed only a trend towards significance ($p = .007$). So, overall we may conclude that the predictions from table 3.17 are substantiated by our findings.

Now, consider again figure 3.3, which shows the medians for all question types for all contexts. The differences described above are indeed represented in the figure. However, in some cases, for example for contexts 2, 5 and 8, there is a difference in height between the colons where there was in fact not found a significant difference. This results from the fact that the Mann-Whitney test is not based on median ratings but also includes the variance of the data points; we should thus keep in mind that the median ratings do not provide the entire picture to determine the differences between data samples.

3.4 Discussion Overall

In the subsections above, we have found various confirmations of our predictions, as well as unexpected effects. In section 3.4.1 we compare the results for the different theories as proposed in section 3.1 and try to formulate some overall conclusions. Subsequently, we discuss some factors that may have affected the results, focusing on the differences in answer distributions between conditions (section 3.4.2), the differences between lexical items in terms of the cartoons (section 3.4.3) and the fact that intonation was excluded from the analysis (section 4.4.3).

3.4.1 Comparing Theories

Theories 1 and 2 focused on the same set of predictions, except that in theory 2 rhetorical contexts 5 and 9 were excluded. However, it seemed that there were no large differences with respect to the significant results between theory 1 and 2. The theories investigated whether the acceptability of PPQs and LNPQs depended solely on the contextual evidence and whether the acceptability of HNPQs depended solely on the speaker belief. We found, however, that for all question types both contextual factors influenced the acceptability rating. This was confirmed by main effects of both contextual evidence and speaker belief for all question types.

When we investigated the individual contexts with respect to the question types, we found some unexpected ratings which may explain some of the results found for theory 1 and 2. Against our predictions contexts 2 and 8 were rated very high overall. Remember that the results for theory 1 and 2 showed that for all question types it was the case that they were more felicitous in contexts with *positive SB* or *negative CE* (contexts 3, 4, 5, 6 and 9) than in other contexts; this was in particular surprising for PPQs. We suggest that these differences were partly induced by the high ratings for contexts 2 and 8 overall. This explains why contexts contexts 3, 4, 5, 6 and 9 were rated significantly lower than contexts 1, 2, 7 and 8 for all question types: the latter group includes contexts 2 and 8 which were rated very high for all question types. Similarly, we can argue that the differences found for the different SB values were caused by the diverting contexts, since *positive SB* (contexts 4, 5, 6) does not include contexts 2 nor 8, while the other SB types both include either one: *neutral SB* (contexts 1, 2, 3) includes context 2 and *positive SB* (contexts 7, 8, 9) includes context 8.

Overall, it seems that it was more useful to look at individual contexts, as done in section 3.3.3 than at groups of contexts, because the contexts differed from each other with respect to the overall rating as well as the licensing of question types. So, this means that instead of considering contextual evidence and speaker belief as separate factors, we should consider these factors in combination to determine the felicity of different question types. Indeed, this was motivated by an observed interaction between contexts and question types and, for PPQs and LNPQs, between contextual evidence and speaker belief. For HNPQs

no interaction was found between contextual evidence and speaker belief, meaning that their felicity does not depend on a combination of these factors but on each factor separately.

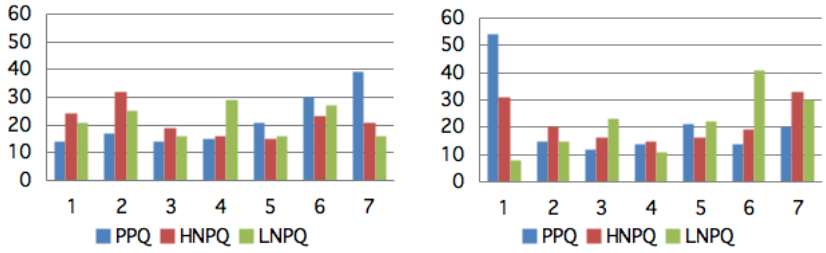
Notably, however, the comparison of theoretically predicted acceptability of question types per context (table 3.17) and the observed ranking of question types per context (table 3.18) showed remarkable correspondences: PPQs were most acceptable (ranked 1 or $1\frac{1}{2}$) in non-rhetorical contexts with $CE = n$ or $CE = p$, HNPQs were most acceptable in non-rhetorical contexts with $SB = p$ or $CE = \neg p$ and LNPQs were most acceptable in non-rhetorical contexts with $CE = \neg p$. So, this again suggests that there is not an interaction between contextual evidence and speaker belief for all question types. How to explain this? We propose that the interaction effects for PPQs and LNPQs are mainly due to rhetorical contexts 5 and 9, as well as contexts 2 and 8 that show abnormal high ratings. With respect to the most interesting contexts 1, 3, 4, 6, and 7, the question types perform as expected, i.e., they depend on speaker belief and contextual evidence separately.

We saw that the results comparing question types per context were indeed much more useful than comparing contexts per question type, for example, context 6 occurs as middle acceptable context for PPQs but the comparison of question types per context shows that PPQs are the least acceptable question type in these contexts. Thus, this result is based more on the overall acceptability of the context than on the acceptability of the question type in the given context. The observation that contexts 3 and 7 occur as low acceptable contexts for all question types will be addressed when we consider the frequency distribution of the answer ratings in section 3.4.2.

So, what do these results mean for the inner/ outer negation distinction for negative polar questions introduced in section 2? We propose that the inner negation reading of negative polar questions is acceptable only in contexts with $CE = \neg p$, while the outer negation reading is only acceptable in contexts with $SB = p$. This confirms the intuition that low negation questions are most often interpreted as inner negation questions (they are most acceptable in contexts 3 and 6) and that high negation question can have both an inner negation and an outer negation reading. In the next chapter we will describe a semantic framework that accounts for these distinctions.

3.4.2 Frequency distribution per condition

We must note that the results were not on all points as uniform as they appeared to be; the ratings of the questions differed much in terms of frequency distribution. We noted above that the extreme values 1 and 7 were selected more often than the middle values (as illustrated in figure 3.2 above). This distribution is not surprising because of the instructions shown above each HIT: participants were instructed to rate questions that they could have used themselves as 'completely natural', and questions that they would not accept in the given context as 'completely unnatural'. The example of a completely natural and a completely unnatural question shown above each HIT also stimulated the



(a) context 3 (b) context 7

Figure 3.14: Frequency distribution for contexts 3 and 7.

usage of extreme values. Of course, since we have not tested this, we cannot make any definite statements about this.

However, the distribution of lexical items was not unvarying over all conditions. In particular, we found dispersed results for contexts 3 ($SB = n, CE = \neg p$) and 7 ($SB = \neg p, CE = n$), as shown in figure 3.14. The effect of this is that for both of these contexts there was only one significant difference between the three question types, which resulted in the middle question type being ranked $1\frac{1}{2}$ in table 3.18.

The dispersed distribution shows that these were the contexts on which the participants were less in agreement. That is, for some reason these contexts create a difference in opinion concerning the felicity of the questions. We propose that this might be because for all lexical items the negative evidence and speaker belief was presented using an antonymous specific noun rather than explicit negation. Thus, because the only non-neutral evidence in contexts 3 and 7 is negative, it might be more natural to formulate a question using this antonymous specific noun. Again, since we have not tested this we cannot make any definite statements about it.

A related issue we discussed earlier was the high overall rating for contexts with *positive CE* (2, 5, 8); we proposed that this high rating was due to the information provided by the contextual evidence being explicitly repeated in the question. These results show that the choices made for the experiment design, e.g. not including extra-linguistic types of evidence and not including questions with antonymous specific nouns, affected the overall results. An interesting issue for future research would be to include these variations in the design in order to get a clear view of the effects.

3.4.3 Cartoons

As described above, each lexical item consisted of a cartoon together with a conversation-theme. Within one lexical item only the contexts were differed between the HITs. The lexical items were constructed by choosing arbitrary conversation-themes and pictures that fitted the general structure (see section 3.2 for experiment design). In order to see the effect of the different car-

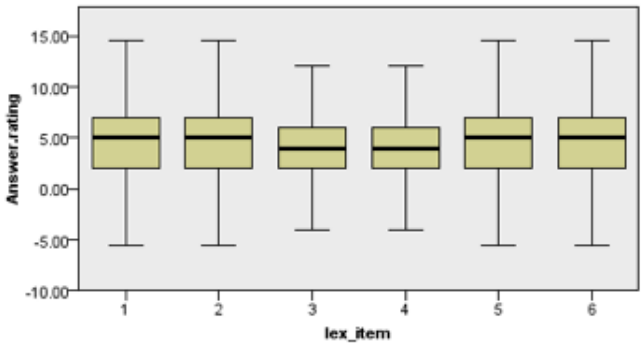


Figure 3.15: Boxplot for Kruskal-Wallis test of lexical items (cartoons).

toons, we investigated the differences between the overall ratings, of which the results are shown below. All cartoons are shown in appendix A

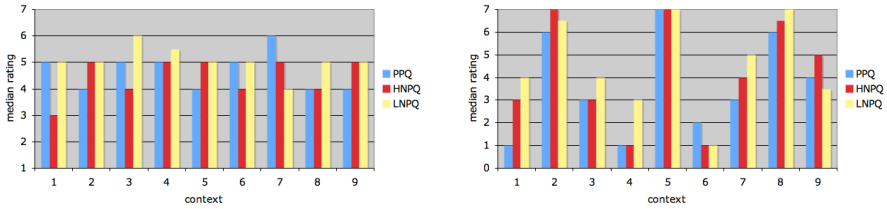
Hypothesis. We tested the following null-hypothesis.

H_0: There is no significant difference between the ratings of the cartoons.

Method. In order to test the differences in results between the lexical items, we performed a Kruskal-Wallis one-way analysis of variance and repeated Mann-Whitney tests on all HITs, thus including the fillers. By including the fillers, we test in particular the influence of the cartoons, because the same cartoons were used for both target and filler items, but not the same conversation themes.

Results The Kruskal-Wallis one-way analysis of variance resulted in a significant difference between the cartoons ($H = 23.21, 5$ d.f., $p < .001$). Closer inspection of the results sugested that this difference is not universal, but only constituted by a subset of the cartoons, as is illustrated in the boxplot in figure 3.15.

Repeated Mann-Whitney tests with Bonferroni correction ($\alpha = .05/15 = .003$) showed that there were no statistical differences between cartoons 1, 2, 5 and 6, and that there were also no statistical differences between cartoons 3 and 4. However, there were significant differences between cartoons 1 and 3 ($U = 880767, p = .001, r = .06$), 1 and 4 ($U = 879895, p = .001, r = .06$), 2 and 3 ($U = 814812, p = .001, r = .06$) and 2 and 4 (U = 814472, p = .001, r = .06). There was a trend towards significant difference between cartoon 5 and cartoon 3 ($U = 889889, p = .006, r = .05$) and cartoon 4 ($U = 889833, p = .006, r = .05$). Also, the differences between cartoon 6 and cartoons 3 ($U = 903553, p = .039, r = .04$) and 4 ($U = 902654, p = .035, r = .04$) were not significant.



(a) lexical items 1, 2, 5, 6 (b) lexical items 3, 4

Figure 3.16: Median rating per question type per context divided by cartoons.

Discussion. These results seem to suggest a division of cartoons into two groups: cartoons 1, 2, 5 and 6 in one group and items 3 and 4 in the other. Cartoon 6 did not significantly differ from 3 and 4, neither did it significantly differ from 1, 2 and 5; because the significance level for the comparisons with 3 and 4 was lower than the significance level for the comparisons with 1, 2 and 5 ($p < .04$ for 3 and 4 versus $0,2 < p < 0,5$ for 1, 2 and 5), it seems more appropriate to consider cartoon 6 as part of the group with cartoons 1, 2 and 5.

Overall, we saw that the first group, containing cartoons 1, 2, 5 and 6 ($Mdn = 5$), was rated higher than the group with cartoons 3 and 4 ($Mdn = 4$). Moreover, looking at the effect of the difference between the two groups of cartoons on the overall results, we saw that the differences between the conditions were much more pronounced for cartoons 3 and 4 than for the other cartoons. Figure 3.16 below represents the same conditions as figure 3.3, divided by cartoons 1, 2, 5, 6 and cartoons 3,4. This shows that the differences observed in figure 3.3 were principally induced by the results for cartoons 3 and 4, where the differences were most prominent. In the foregoing analyses we have included the results for all cartoons and treated the different conditions as independent. This suggests that the ratings do depend on the cartoons, which means that effects may have been even stronger had we only considered cartoons 3 and 4.

Looking at the cartoons, the difference between the two groups of cartoons may be explained by the observation that the facial expressions in cartoons 1, 2, 5 and 6 were more pronounced; in the case of 1, 2 and 6 the expressions seem to suggest a conflict between the individual that asks the question and the addressee. Thus, in any (linguistic) context that did not offer such a conflict, participants may have also been influenced on their acceptability rating for the question by the facial expressions in the pictures. Although it was very clearly stated that participants should base their answers on the foregoing *conversations*, the different results between the cartoons suggest that these also affected the rating. In the case of picture 5 it seems that the expressions of the individuals in the pictures did not accord with the inquisitive context; here there is a lack of conflict between the expressions of the individuals – the addressee in the last picture seems to be happy with any answer.

So, overall we can say that the cartoon (i.e. the facial expressions/ situation in which a question occurs) is another factor that affects the naturalness of the

questions. Future research may aim at investigating the effect of facial expressions on acceptability ratings by including multiple cartoons for each condition in such a way that the effect of the cartoon can be controlled for and separated from the effects of interest.

3.4.4 Intonation

Finally, we should mention the important role intonation has for determining the acceptability of questions, or utterances in general. For example, the difference between a declarative utterance and a question can be determined by intonation (e.g. Gunlogson, 2001). Moreover, intonation is a very important function for expressing bias, in declaratives as well as in questions (Nilsenova, 2006, and references mentioned there).

We already mentioned in the discussion of section 3.3.3.2 that special intonation may be used to express surprise which affects the acceptability of the question. Similarly, consider the two low negation questions in (41), where the underlining signifies emphasis (example of lexical item 4).

(41) a. Did he not visit his uncle?
b. Did he not visit his uncle?

Note how the intonation affects the interpretation of the questions. Arguably, (41a) is interpreted as an outer negation question, whereas (41b) is interpreted as an inner negation question. As we have shown, these different readings affect the acceptability of the question.

It was beyond the scope of the current research to include variations in intonation in the experiment. However, participants may have imagined an intonation pattern themselves, which affected their rating of the question. Future research should incorporate this important feature of interpretation.

3.5 Conclusions and Future work

Despite the large number of comparisons for each analysis which resulted in a large correction on the basis of the Bonferroni method, we have found several significant differences, indicating that the observed differences represent strong effects.

The main result of the experiment was the fact that we indeed found differences between the question types in terms of the contexts in which they are acceptable, as well as between the contexts in terms of the question types they license. We saw that, although contexts 2, 5, 8, and 9 were rated high overall, there were clear differences between the question types with respect to the other contexts. Moreover, the predictions about the felicitous question types per context were effectively satisfied.

Although these results come across as convincing, we must note that this was still a preliminary study. Above we already did some recommendations for future research, and there are many more matters to exploit. These include

but are not limited to: including questions with different syntactic forms (e.g. variation in terms of nouns, auxiliaries, quantifiers, etc.), providing intonation (in particular to the questions), or letting participants formulate the question themselves.

Overall we can state that the theory provided by the theoretical background and linguists' intuitions in many ways correspond to the judgments of arbitrary participants. However, it is interesting to note that participants do not always agree with each other or with the theory, which results in important implications for the further development of linguistic theories.

Chapter 4

Theoretical Implications

4.1 Introduction

In the previous chapter we empirically tested some predictions following from a theory about the felicity conditions of three types of polar questions: positive polar questions (PPQs), high negation polar questions (HNPQs) and low negation polar questions (LNPQs). In the current chapter we will investigate how these observations may be implemented in a theoretical framework. We chose the framework of *inquisitive semantics* to formalize the observed acceptability conditions of (negated) polar questions. This analysis is not intended as providing the unique solution to the issue of negative polar questions, but merely shows a way in which this can be achieved. We believe that other theories, for example van Rooij & Šafářová's (2003) decision-theoretic approach or Reese's (2007) metalinguistic approach, may be extended to account for the same observations.

Inquisitive semantics is not a semantic framework in its traditional sense in that it is not truth-conditional: a sentence is not only associated with the information it provides, but also by the issue it raises. In inquisitive semantics, the meaning of a sentence "directly reflects a primary use of language: the exchange of information in a cooperative process of raising and resolving issues" (Groenendijk & Roelofsen, 2009, p. 1). Thus, the border between traditional semantics and pragmatics becomes shifted, meaning that aspects of language that traditionally could only be accounted for by pragmatic theories, now lay within the scope of inquisitive semantics. This makes inquisitive semantics an ideal framework to address the issue of negative polar questions (see Anderbois, 2011, for a related approach).

In this chapter we will first introduce the notion of highlighted possibilities in section 4.2. Then, in section 4.3, we introduce a new question operator which is applied to account for the different types of polar questions in section 4.4. Finally, section 4.5 offers a comparison of the proposed theory to earlier approaches, as described in section 2.1.

4.2 Highlighting possibilities

In most traditional theories on questions, positive and negative questions are treated as equivalent (Hamblin, 1973; Karttunen, 1977; Groenendijk & Stokhof, 1984, among others)[1]. As we saw, inquisitive semantics based on the definitions described in section 1.2 above does not make this distinction either. However, the results of the experiment showed that the syntactically different question types also differed with respect to the contexts in which they occurred, suggesting a distinction in meaning.

In order to equip inquisitive semantics with a tool to make this distinction, we introduce the highlighting of possibilities (Roelofsen & van Gool, 2010; Pruitt & Roelofsen, 2010). The highlighted possibilities of a sentence φ are denoted as $[\![\varphi]\!]_H$. We will from now on refer to the proposition expressed by φ as $[\![\varphi]\!]_P$, and assume that the highlighted possibilities of φ are a subset of the possibilities φ proposes: $[\![\varphi]\!]_H \subseteq [\![\varphi]\!]_P$.

4.2.1 Motivation

As described above, in inquisitive semantics the meaning of a sentence is a set of possibilities, or proposals to update the common ground in one or more ways. An inquisitive sentence proposes multiple possibilities; this means that it requests a response that provides enough information to establish at least one of the proposed updates. A *response* to a question φ can thus be described as corresponding to one or more of the possibilities proposed by φ. More specifically, in order for a response to φ to be *informative* it must exclude at least one of the possibilities proposed by φ.[2] Polar (*yes/no*) answers are taken to provide such a response, thus 'choosing' one of the possibilities.

Polar answers may involve a polarity particle followed by a constituent or an isolated polarity particle. The former type of responses we will call *joint particle responses*, and the latter we will call *solo particle responses*.[3] Consider the following example of two solo particle responses to a positive polar question.

(42) Is John coming to the party?
  a. Yes. ⇒ he is coming
  b. No. ⇒ he is not coming

The 'yes' answer in (42a) corresponds to the possibility expressing that John is coming and the answer in (42b) corresponds to the possibility that John is not coming. Indeed, we saw above that the possibilities for a polar question

[1] It must be noted that Groenendijk & Stokhof (1984, pp. 321-323) do address the issue, providing an explicit account of the interpretation of *yes* and *no* to capture the difference between positive and negative interrogatives.

[2] Groenendijk & Roelofsen (2009) define the notion of a *compliant* answer and Pruitt & Roelofsen (2010) define *safe* answers. Here, the notion of informativity suffices, but the account can easily be extended to account for different types of answers.

[3] We will not discuss here how these two types of responses are related to each other. See Kramer & Rawlins (to appear) for an ellipsis account.

correspond to the possibilities provided by $[\![p \vee \neg p]\!]$, as shown in figure 1.1(c). Thus, it seems that polar answers correspond to the possibilities proposed by a polar question. However, the licensing and interpretation of polar answers is not straightforward. Consider the following example.

(43) Is the door open?
   a. Yes. ⇒ the door is open
   b. No. ⇒ the door is closed

(44) Is the door closed?
   a. Yes. ⇒ the door is closed
   b. No. ⇒ the door is open

In the inquisitive semantics developed so far, (43) and (44) are entirely equivalent; they both propose two possibilities which correspond to the propositions expressing that the door is open and that the door is closed, respectively. Yet, there is a clear empirical difference between these two questions: in response to (43), *yes* means that the door is open and *no* means that the door is closed, while in response to (44) this is the other way around. We follow Roelofsen & van Gool (2010) and Pruitt & Roelofsen (2010) in assuming that the contrast between (43) and (44) follows from the fact that they highlight different possibilities.

The idea behind highlighted possibilities is that they may serve as antecedents for subsequent anaphoric elements. The answer particles *yes* and *no* are considered to be such anaphoric elements; their felicity can be defined on the basis of highlighted possibilities. This enriches the semantic interpretation, thus providing an extra level of meaning apart from the informative and inquisitive content of a sentence. For example, the distinction between (43) and (44): the question in (43) is taken to highlight the possibility that the door is open and (44) highlights the possibility that the door is closed. Accordingly, we can explain the differences in the interpretation of the polar answers. Pruitt & Roelofsen (2010) propose that *yes*-answers presuppose that there is exactly one highlighted alternative possibility (i.e. there is exactly one highlighted possibility that is not included in any other highlighted possibility) and that a *yes* answer corresponds to this possibility. A *no*-answer is taken to simply reject all the highlighted possibilities of a question (see also Farkas, 2011).

4.2.2 Formalization

Remember that in section 1.2 we defined the prejacent of a question as the clause to which the question operator attaches. Now, we may assume that [Q φ] simply highlights the possibilities that its prejacent highlights: $[\![\text{Q } \varphi]\!]_H = [\![\varphi]\!]_H$. However, the felicitous polar answers to disjunctive questions, as shown in examples (45), suggest that this does not suffice.

(45) Is John or Mary↑ coming to the party?

a. Yes. ⇒ John or Mary is coming.
b. No. ⇒ neither is coming.

Here the underlining signifies that the clause receives emphasis and the arrow represents a rising intonation. In the case of disjunctive questions the intonation is crucial because different intonation patterns may result in different question types; e.g. alternative questions, open questions and, as in example (45) above, polar questions (e.g. Roelofsen & van Gool, 2010). The interpretation of the solo particle response in (45a), expressing that John or Mary is coming, suggests that a polar question should not only highlight the highlighted possibilities of its prejacent, but also the union of these possibilities. Since we assumed that the highlighted possibilities are a subset of the proposed possibilities, we thus also have to adapt definition 3 from section 1.2 which defined the possibilities in the denotation of [Q φ]. Definition 10 thus shows the proposed and highlighted possibilities of the question operator. Note that if the prejacent only has a single possibility, the additional union has no effect on the proposed and highlighted possibilities.

Definition 10 (Q operator)**.**

i. $[\![\text{Q } \varphi]\!]_P = [\![\varphi]\!]_P \cup \overline{[\![\varphi]\!]_P} \cup \{\bigcup [\![\varphi]\!]_P\}$
ii. $[\![\text{Q } \varphi]\!]_H = [\![\varphi]\!]_H \cup \{\bigcup [\![\varphi]\!]_H\}$

Now, example (46) shows that this definition indeed predicts the correct highlighting for the antonymous questions shown in examples (43) and (44). Moreover, if we turn to the distinction between positive and negative polar questions, we see that this definition provides positive polar questions and low negation polar questions with different highlighted possibilities, as shown in (47).

(46) a. $[\![$ Is the door open? $]\!]_H = \{$the door is open$\}$
b. $[\![$ Is the door closed? $]\!]_H = \{$the door is closed$\}$

(47) a. $[\![$ Is John coming? $]\!]_H = \{$John is coming$\}$
b. $[\![$ Is John not coming? $]\!]_H = \{$John is not coming$\}$

Example (47b) provides the highlighting for a low negation polar question (see Farkas & Roelofsen, 2011, for a more elaborate discussion).

However, still we cannot account for the distinction between low negation polar questions and, for example, high negation polar questions. For this, we need a new operator, which is introduced in the next section.

4.3 Negated Question Operator

The *negated question operator* [Q^N] is defined as associating with a negative clause $\neg\varphi$, such that the negation is absorbed by the operator. The negation here functions not as a standard negation but only signals the negated question operator, which means that it is not part of the scope of the operator, i.e.,

what we defined as the *prejacent* of the question. (cf. metalinguistic negation Horn, 1989; Reese, 2007) So, the negated question operator in effect neutralizes the negation of the clause it attaches to, which, as we will see, affects the possibilities it proposes and highlights. This neutralization effect of negation may be compared to the linguistic phenomenon of *negative concord*, where a sequence of seemingly negative clauses can get a simple negative meaning (e.g. de Swart & Sag, 2002).

Since the negation is not part of the prejacent of a Q^N question, we write $[Q^N\ \varphi]$ to refer to a negative question that is bound by the negated question operator. Note that the syntactic representation of a question $[Q^N\ \varphi]$ does contain a negation, but this is not represented in the semantic denotation because of the metalinguistic nature of this negation. Definition 5 is straightforwardly extended to account for the negated question operator, as shown in definition 11.

Definition 11 (Prejacent Q^N)**.**

1. $\textsc{prej}([Q\varphi]) = \varphi$
2. $\textsc{prej}([Q^N\varphi]) = \varphi$

Thus, in the case of the Q and the Q^N operator, the prejacent simply returns the clause to which the question operator syntactically attaches; in the case of the negated question operator this is the sentence φ without the metalinguistic negation. Note, however, that φ may still be a negated clause, which, as we will see below, is the case for high/low negation questions. Intuitively, the question expressed by $[Q\ \varphi]$ may be interpreted as "*Is it the case that* φ*?*" and $[Q^N\ \varphi]$ may be interpreted as "*Isn't it the case that* φ*?*". When discussing high negation questions in section 2.3.2 we called the latter an explicit outer negation reading.

Note that the proposed and highlighted possibilities of $[Q\ \varphi]$ in definition 10 above are defined in terms of φ, which is the prejacent of $[Q\ \varphi]$. Similarly, we can define the possibilities and highlighted possibilities of $[Q^N\ \varphi]$ in terms of its prejacent φ:

Definition 12 (Q^N operator)**.**

i. $[\![Q^N\ \varphi]\!]_P = \overline{[\![\varphi]\!]_P} \cup \{\bigcup [\![\varphi]\!]_P\}$

ii. $[\![Q^N\ \varphi]\!]_H = \{\bigcup [\![\varphi]\!]_H\}$

Thus, the possibilities proposed and highlighted by the negated question operator are a subset of the possibilities proposed a highlighted by the standard question operator; it doesn't include the possibilities of $[\![\varphi]\!]$. The effect of this is that if φ is not inquisitive, i.e. if it contains only one possibility, then the possibilities and highlighted possibilities for $[Q^N\ \varphi]$ and $[Q\ \varphi]$ coincide. Moreover, it follows that $[Q^N]$ highlights always exactly one possibility, because it takes the union of the possibilities highlighted by its prejacent. We will see that this is indeed the desired interpretation.

Figure 4.1 visualizes the possibilities for the different question operators. Here, the gray possibilities represent the highlighted possibilities. Note that a

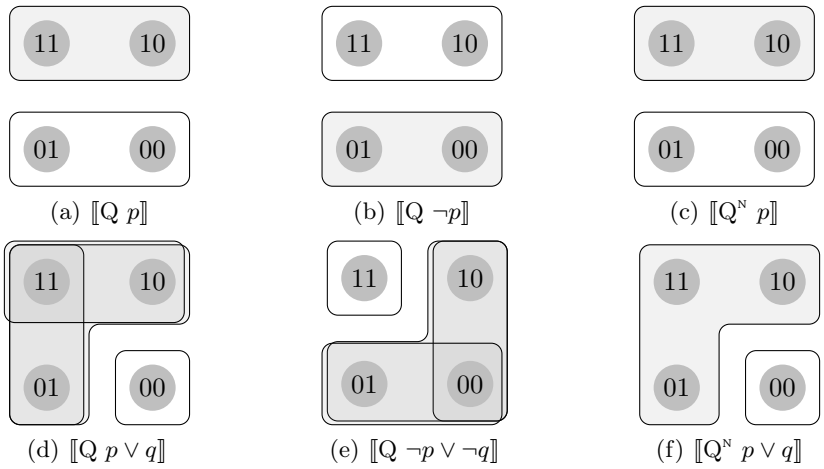


Figure 4.1: Visualization of possibilities and highlights for (a) a positive polar question, (b) a negative polar question with the traditional question operator, (c) a negative polar question with the negated question operator, (d) a positive disjunctive question, (e) a negative disjunctive question with the traditional question operator and (f) a negative disjunctive question with the negated question operator.

negated prejacent may combine both with the standard and with the negated question operator, as shown in figures 4.1(b) and 4.1(c). As described above, the difference between [Q φ] and [Q$^{\text{N}}$ φ] becomes apparent when the prejacent φ is inquisitive itself. This is shown in figures 4.1(d)-4.1(f). We see that 4.1(d) and 4.1(f) indeed contain different highlighted possibilities: while [Q $p \vee q$] highlights the possibilities representing p and q and their union, [Q$^{\text{N}}$ $p \vee q$] only includes the union of the possibilities for p and q. The denotation is motivated by the examples which show the interpretation of solo particle answers to a positive disjunctive question and a high/low negation disjunctive question.

(48) Does Mary like apples↑ or pears↓?
  a. # Yes. ⇒ Yes what?!
  b. No. ⇒ Mary likes neither.

(49) Doesn't Mary not like apples or pears↑?
  a. Yes. ⇒ Mary does not like apples or pears.
  b. No. ⇒ Mary likes both.

Remember, a *yes*-answer presupposes a unique highlighted alternative possibility and a *no*-answer just rejects all highlighted possibilities. For now, let us assume that the PPQ in (48) is interpreted with the [Q] operator, and the HNPQ in (49) is interpreted with the [Q$^{\text{N}}$] operator (see the next section for a motivation of this assumption). Then the *yes*-answer in (48) is indeed infelicitous because

the denotation of the PPQ contains multiple highlighted possibilities, as shown in figure 4.1(d). In the case of (49), however, the denotation only contains one highlighted possibility, namely the union of the highlighted possibilities shown in figure 4.1(e).

So, in this case the expected answer patterns correspond to the predicted pattern on the basis of the provided denotation. In section 4.4.2 we will describe the predicted answer patterns for the different question types in more detail.

4.4 Accounting for different question types

Now, we use these two question operators to account for the difference between positive polar questions, high negation polar questions, low negation polar questions and high/low negation polar questions. Since the negated question operator is signaled by a negation in the clause it attaches to, positive polar questions always combine with the traditional question operator [Q]. High and low negation questions can be combined both with the traditional operator [Q] and the negated question operator [Q^N]. Because of the markedness of interpretations containing double negation and on the basis of example (49) above, we can assume that high/low negation questions always combine with [Q^N]. This results in the following highlight pattern. Here, φ is taken to be the proposition expressed by the sentence 'John is coming'.

(50) Is John coming?
  a. $[\![Q\ \varphi]\!]_H$ = {John is coming}
  b. # $[\![Q^N\ \varphi]\!]$

(51) Is John not coming?
  a. $[\![Q\ \neg\varphi]\!]_H$ = {John is not coming}
  b. $[\![Q^N\ \varphi]\!]_H$ = {John is coming}

(52) Isn't John coming?
  a. $[\![Q\ \neg\varphi]\!]_H$ = {John is not coming}
  b. $[\![Q^N\ \varphi]\!]_H$ = {John is coming}

(53) Isn't John not coming?
  a. # $[\![Q\ \neg\neg\varphi]\!]$
  b. $[\![Q^N\ \neg\varphi]\!]_H$ = {John is not coming}

As we can see in examples (51) and (52), high and low negation polar questions with the same prejacent get the same highlighting with the [Q] operator. With the [Q^N] operator, however, both get the same highlighting as the positive question with the [Q] operator and the same prejacent.

We are now ready to return to the results of the experiment and investigate how the different question operators can be used to explain the felicity conditions

of the different question types. Accordingly, we consider two highlighting signalers to show that the predicted highlighting patterns are correct: the expected answer patterns in section 4.4.2 and intonation in section 4.4.3.

4.4.1 Contextual Conditions

In the experiment we found substantiation for our predictions regarding the felicity of different question types: we found that PPQs were most acceptable in (non-rhetorical) contexts with $CE = n$ or $CE = p$, HNPQs were most acceptable in non-rhetorical contexts with $SB = p$ or $CE = \neg p$ and LNPQs were most acceptable in non-rhetorical contexts with $CE = \neg p$.

Remember that we defined the distinction between *inner* and *outer* negation as based on the context in which a question occurs: inner negation questions occur in contexts with $\neg p$ contextual evidence, and outer negation questions occur in contexts with p speaker belief. In the current analysis, we can thus treat negative questions that combine with a [Q] operator as having an inner negation reading, and negative questions that combine with a [Q^N] operator as having an outer negation reading. Thus, low negation questions principally function as inner negation questions, and high negation questions are good with the inner as well as with the outer negation reading. We thus may assume that low negation questions are most often interpreted using the [Q] operator, while high negation questions commonly occur with the [Q] operator as well as with the [Q^N] operator.

Since the different readings are licensed by different contexts, we may state that the context selects which of the possible question operators is applied. We propose that the question operators are subject to two principles. Basically, these principles state that [Q] questions presuppose certain contextual evidence provided by a context, and [Q^N] questions presuppose a certain belief state of the speaker.

Definition 13 (Acceptability for Q and Q^N)**.** Let γ be a context and s a speaker with belief state σ.

1. [Q $\neg\varphi$] is *acceptable in* γ only if γ provides compelling evidence *against* φ.
2. [Q^N φ] is *acceptable for* s only if $\varphi \in \sigma$.

Note that the acceptability for [Q] questions is defined in terms of questions with a negated prejacent $\neg\varphi$; it follows that if the context γ provides compelling contextual evidence *for* or *neutral* with respect to φ, then [Q $\neg\varphi$] is not acceptable, but [Q φ] is acceptable. For [Q^N] questions it holds that its prejacent, which can be positive or negative, should be in the belief state of the speaker. Thus, since PPQs combine with [Q] and their prejacent is non-negative, they occur in all contexts except those with compelling contextual evidence *against* their prejacent. Inner negation questions also combine with [Q] and presuppose compelling contextual evidence *against* the proposition in their prejacent. On the

other hand, outer negation questions combine with [Q^{N}] and thus presuppose that their prejacent is in the belief state of the speaker.

As we have seen, except for the context there are various other ways to disambiguate between inner and outer negation polar questions, for example using polarity-sensitive lexical items like 'either' and 'too', as shown in (54).

(54) a. Isn't John coming too? ⇒ John is coming
b. Isn't John coming either? ⇒ John is not coming

In the current analysis these features may be seen as signaling a certain context and thus indirectly a certain question operator. So, in (54a) 'too' is a presupposition trigger for the positive proposition expressing that John is coming; this suggests that the speaker believes that John is coming, which licenses the negated question operator [Q^{N}]. In (54b) 'either' presupposes the opposite, namely that John is not coming, which means that the negated proposition has somehow become salient, namely by the contextual evidence; this signals the question operator [Q].

4.4.2 Answer Patterns

As discussed above, the highlighted possibilities are defined as to serve as antecedent for subsequent anaphoric elements, for example polar answers like 'yes' and 'no'. Examples (55)-(58) show the expected answers for the different question types.

(55) Is John coming?
a. Yes, he is coming.
b. No, he is not coming.
c. # Yes, he is not coming.
d. # No, he is coming.

(56) Is John not coming?
a. ? Yes, he is coming.
b. ? No, he is not coming.
c. Yes, he is not coming.
d. No, he is coming.

(57) Isn't John coming?
a. Yes, he is coming.
b. No, he is not coming.
c. ? Yes, he is not coming.
d. ? No, he is coming.

(58) Isn't John not coming?
a. # Yes, he is coming.
b. # No, he is not coming.
c. Yes, he is not coming.
d. No, he is coming.

Remember that 'yes' is taken to confirm the unique highlighted alternative possibility and 'no' is taken to reject all highlighted possibilities. It thus follows from example (55) that a PPQ can indeed only be interpreted as highlighting

its (positive) prejacent, since it only licenses a 'yes'-answer with a positive prejacent, as shown in (55a) (in Farkas's (2011) terms: [same,+]). Similarly, a hlnpq, as shown in (58), only highlights the negated proposition that John is not coming, since 'yes' can only be interpreted with this proposition. For high and low negation questions both highlighting patterns seem possible, although not equally good: a low negation question like (56) seems more likely to be interpreted as highlighting the proposition expressing that John is not coming, while a high negation question like (57) is expected to highlight the proposition expressing that John is coming.

Of course all these observations should be experimentally tested before making strong assertions about them, but it seems that the theoretical predictions made in (50)-(53) agree with the intuitions about licensing of polar answers, as shown in (55)-(58).

4.4.3 Intonation

Another feature that affects highlighting patterns is intonational focus, as shown by Roelofsen & van Gool (2010). They propose that focus affects the computation of highlighted possibilities by making sets of possibilities *collapse*; the highlighting of focused constituents is taken to be the generalized union of the highlighted possibilities of that constituent (for more detail, see Roelofsen & van Gool, 2010). Thus, this means that for a question with intonational focus the highlighting is first affected by the intonation and after that by the question operator. Let us have a look at an example.

(59) a. Isn't John↑ or Mary↓ coming?
b. Isn't John or Mary↑ coming?

Remember, the underlining signifies that the clause receives focus and the arrows represent rising or falling intonation. The intonation pattern in (59a) is often called the *closed* intonation pattern and the pattern of (59b) is called the *block* intonation pattern. Intuitively, it seems more natural to interpret the closed intonation pattern with the [Q] operator and the block pattern with the [Q^N] operator. Indeed, the closed intonation pattern is simply unacceptable in combination with a hlnpq, of which we know it only occurs with the [Q^N] operator, as is shown in example (60)

(60) a. # Isn't John↑ or Mary↓ not coming?
b. Isn't John or Mary↑ coming?

So how can we explain this? Let us analyze example (59) with the [Q^N] interpretation for both questions. Thus, both questions have a non-negated prejacent expressing that John or Mary is coming. According to Roelofsen & van Gool's (2010) analysis this prejacent gets the following highlighting on the basis of the different intonation patterns:

(61) a. $[\![[\text{John}]_F \text{ or } [\text{Mary}]_F \text{ is coming}]\!]_H$
$=$ {John is coming, Mary is coming}

b. $[\![[\text{John or Mary}]_F \text{ is coming}]\!]_H$
$= \{\text{John is coming} \cup \text{Mary is coming}\}$

Now these sets are bound by the negated question operator which, as we discussed above, highlights the union of the highlighted possibilities of its prejacent. This would results in the same highlighting for both questions, namely the union of the possibilities that John is coming and Mary is coming, as shown in (61b). So, if this is unproblematic, why is the [Q^{N}] reading of (59a) excluded? In order to explain this we need to assume a further restriction on the use of intonation. Since intonation is purposely used by speakers to influence the interpretation of a sentence, we may assume that the highlighting forced by the intonation must be reflected in the highlighting resulting from the compositional semantics of the sentence. Thus, because the closed intonation pattern together with the [Q^{N}] operator in (59a) and (60a) results in a highlighting different from the one imposed by the intonation pattern, this interpretation is unacceptable. Note that since the highlighted possibilities of [Q φ] always include the highlighted possibilities of φ, the [Q] operator does not restrict the intonation pattern.

4.5 Implications

We will now compare the proposed analysis to the other approaches to negated questions as described in section 2, this is done in section 4.5.1. In section 4.5.2, we will investigate one of these in more detail, focusing on Reese's (2007) assumption that outer negation questions have an assertive aspect.

4.5.1 Comparison to other approaches

As we saw, the current analysis is an extension and formalization of Büring & Gunlogson (2000), who propose that contextual evidence determines the felicity of all question types. The experiment showed, however, that speaker belief also crucially determines the acceptability of the question types, in particular of HNPQs.

In this respect, the theory conforms to Romero & Han's (2004) idea that the denotation of the different question types should somehow be based on the epistemic bias that in particular high negation question seem to express. However, although for Romero & Han (2004) this resulted in diverging denotations for positive and negative polar questions, in our analysis positive and negative questions propose the same possibilities, they only differ in terms of their highlighting of possibilities and licensing conditions.

The theory proposed by van Rooij & Šafářová (2003) differs from the proposed analysis in many respects. However, the most important difference is that they refrain from making a distinction between an inner and outer negation reading of negative polar questions. It resulted from the experiment, however, that a difference in the interpretation of high and low negation questions can be observed. As we have seen, this distinction can be explained in terms of the inner/ outer negation distinction.

Now let us examine how our analysis deals with Reese's (2007) examples.

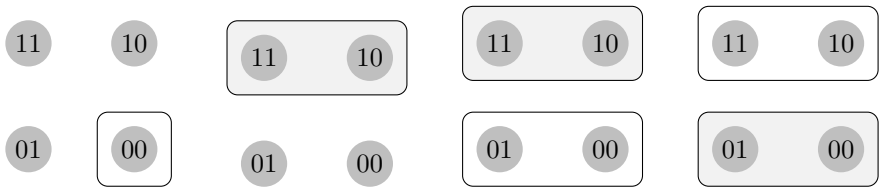
4.5.2 Assertive aspect of outer negation

As we saw above, Reese (2007) treats outer negation questions, in contrast to inner negation questions, as having an assertive aspect; he states that the positive epistemic bias of these question acts as an assertion. He states that, like assertions, outer negation questions can be used as denials. Consider again example (22) from section 2.1.3 above, here repeated slightly adapted as (62). The assertion shown in (62a) constitutes a context with $\neg p$ compelling contextual evidence, where p represents the proposition expressed by 'Jane turned in her assignment'. The sentences in (62b-e) are taken to be responses to the assertion in (62a), which we will call the antecedent assertion.

(62) a. A: None of the students turned in their assignment.
b. B: Jane turned in her assignment.
c. # B: Did Jane turn in her assignment?
d. B: Didn't Jane turn in her assignment?
e. B: Did Jane not turn in her assignment?

Since assertions are taken to commit the speaker to its content, the assertion in (62b) constitutes a conflict context by offering a speaker belief supporting the proposition p, expressing that Jane turned in her assignment; thus, (62b) *denies* (62a). On the basis of these examples, Reese (2007) considers outer and inner negation questions to have distinct discourse functions, where outer negation questions can have both an assertive and an inquisitive function and inner negation questions can only have an inquisitive function. It is unclear, however, how he accounts for the infelicity of the PPQ in (62c). As we have seen above, positive polar questions are unacceptable in contexts that express evidence *against* their prejacent, as is the case in (62c); PPQs thus cannot function as a denial.

In the current analysis we can account for the example above, without considering outer negation questions to be ambiguous between a question and assertive type of expression. Since example (62) illustrates the effect of contextual factors on the felicity of different question types, we turn to the semantic principles introduced in section 4.4.1. We will assume that the high negation question in (62d) is interpreted with the $[Q^N]$ operator and that the low negation question is interpreted with the [Q] operator. Note that in terms of highlighted possibilities, the high negation question in (62d) and the positive question in (62c) are the same: they both highlight the possibility p expressing that Jane turned in her assignment. How can we then explain the difference in felicity between these questions? According to definition 13.2, a $[Q^N]$ question is only acceptable in a context with a speaker belief supporting its prejacent. Since in example (62) we have no evidence about the speaker belief in this context, in particular, we have no evidence that the speaker does *not* believe the proposition p, we can assume that this condition is satisfied and thus the question is felicitous. A [Q]



(a) (62a) (b) (62b) (c) (62d) (d) (62e)

Figure 4.2: Visualization of sentences in (62).

question, however, depends on the contextual evidence. In this example this is defined as providing compelling contextual evidence *against* the proposition expressing that Jane has turned in her assignment. According to definition 13.1, in this context only a [Q] question with the negated proposition as prejacent is acceptable. Since this is not the case in (62c), this question is infelicitous. The low negation question in (62e), on the other hand, does have as a prejacent the negated proposition, which makes the question felicitous.

In order to see why (62d) and (62e) have different functions, namely the former as a denial and the latter as a grounding question, we have to turn to the highlighting patterns of these questions. Figure 4.2 illustrates how these sentences are interpreted in inquisitive semantics with highlighting: figure 4.2(a) represents the assertion in (62a) and 4.2(b)-4.2(d) accordingly show the felicitous responses to this assertion. Again the colored possibilities represent the highlighted possibilities. We see here that the assertion in (62b) denies (62a) by offering a unique and highlighted possibility that does not overlap with the proposed possibility in (62a).

Both negative polar questions (62d) and (62e) propose two possibilities. However, because (62d) highlights the same possibility as the assertion (62b), namely the possibility that does not include the antecedent assertion shown in (62a), it functions as a denial as well, although a weaker one because it also proposes the possibility that overlaps with the antecedent assertion. Because (62e) only highlights this overlapping possibility, it functions merely as a confirming, or grounding, discourse act.

Chapter 5

Conclusions

In this research we have investigated the difference between positive polar questions and different types of negative polar questions. The proposed felicity conditions of the different question types in terms of contextual evidence and speaker belief were empirically investigated. It was found that the proposed contextual factors can indeed for a large part account for the different interpretations of positive polar questions, high negation polar questions and low negation polar questions.

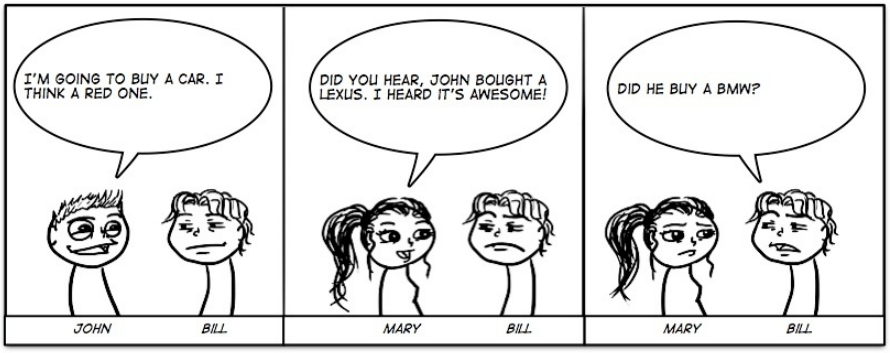
On the basis of the obtained results we formulated a framework in inquisitive semantics that accounted for (i) the semantic equivalence of different types of polar question in terms of the issue they raise (ii) the difference between the *bias* or *expected answer* of the different types of polar questions and (iii) the differences in licensing conditions between the different types of polar questions with respect to the context, viewed as a combination of contextual evidence and speaker belief.

Throughout the thesis several recommendations for future research were mentioned, the most important of which was taking into account the role of intonation in the interpretation of different question types. Moreover, in my opinion the approach advocated here as a combination of theoretical and empirical research should be pursued more often in the field of linguistics, in order to obtain valuable and constructive research.

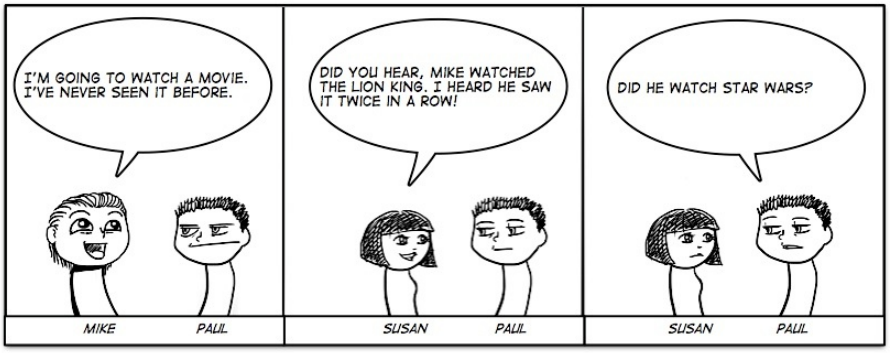
Appendix A

Lexical Items

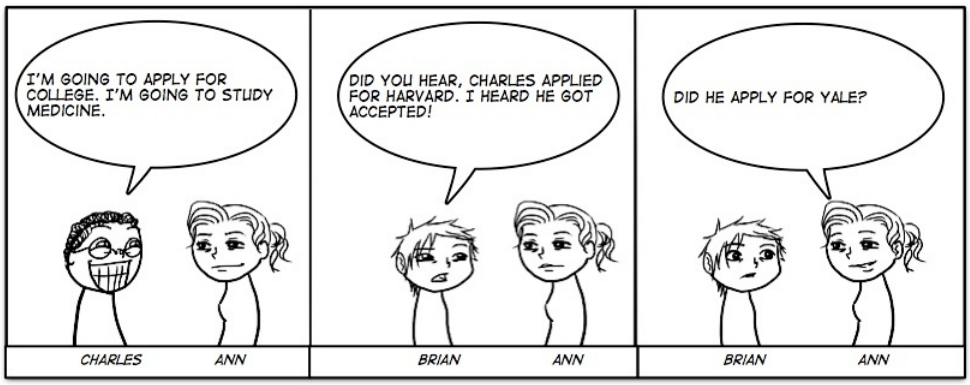
All the cartoons are shown in context 3, such that the first picture (SB) shows the *neutral* noun and the second picture (CE) and third picture (Q) show the two specific nouns, *negative* and *positive*, respectively.



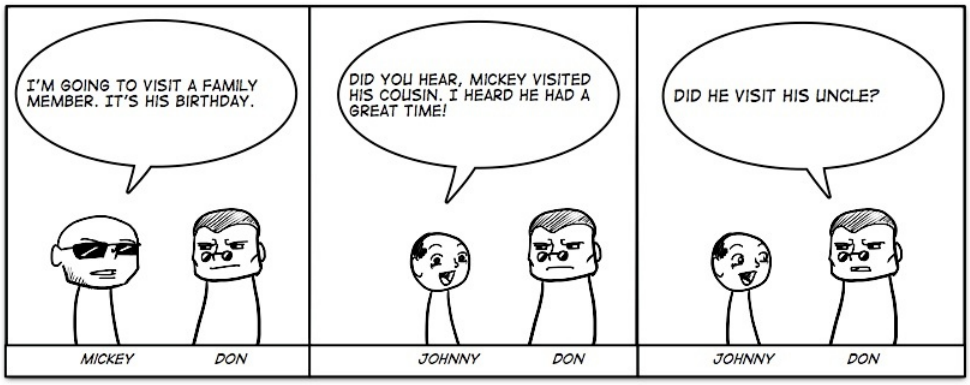
(a) cartoon 1 in context 3



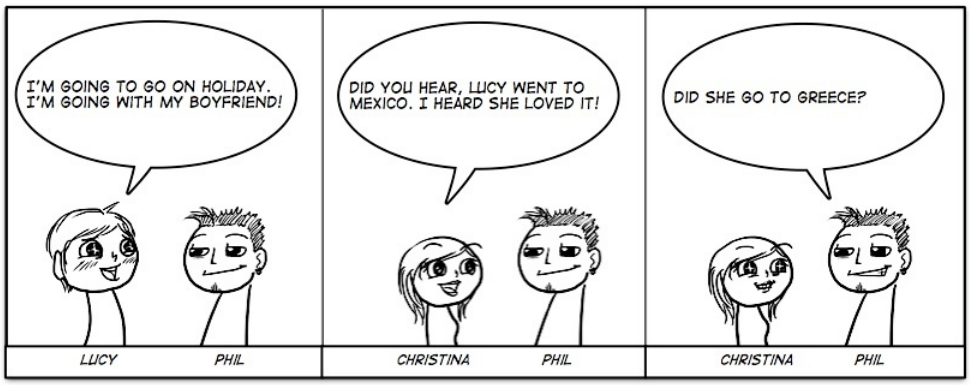
(b) cartoon 2 in context 3



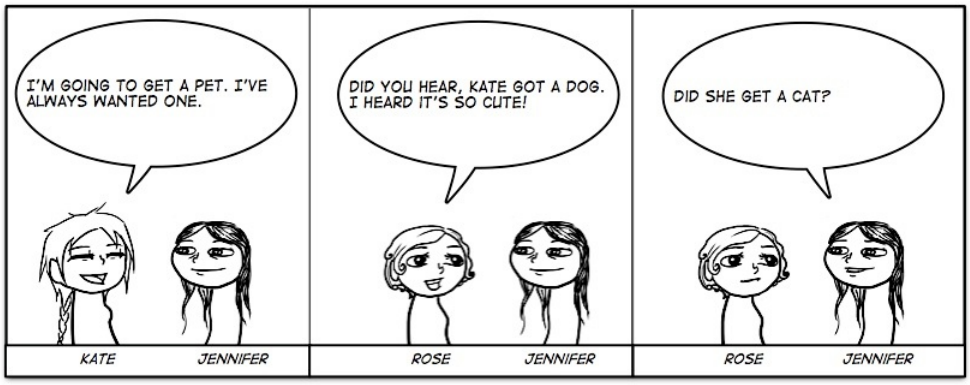
(c) cartoon 3 in context 3



(d) cartoon 4 in context 3



(e) cartoon 5 in context 3



(f) cartoon 6 in context 3

Appendix B

Fillers

The fillers are presented in random order, except that the different versions for each filler item are presented together: a positive question in a (fairly) natural context (GOOD/POS), a positive question in a (fairly) unnatural context (BAD/POS) and the same for negative questions.

Note that the names in the dialogues signal the cartoon that was used to present the filler, e.g. the first set of fillers below was presented with cartoon 1 and the second set with cartoon 4. The cartoons occurred equally often in the fillers.

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1103	1203	1303	1403
I'm going to go to the dentist. I'm a bit scared. Did you hear, John had two of his wisdom teeth removed. I heard it didn't hurt at all. Which ones were removed?	I'm going to go to the dentist. I'm a bit scared. Did you hear, John had two of his wisdom teeth removed. I heard it didn't hurt at all. Which ones did he love?	I'm going to go to the dentist. I'm a bit scared. Did you hear, John had two of his wisdom teeth removed. I heard it didn't hurt at all. Which ones were not removed?	I'm going to go to the dentist. I'm a bit scared. Did you hear, John had two of his wisdom teeth removed. I heard it didn't hurt at all. Which ones did he not love?
1116	1216	1316	1416
I'm going to ask Mary on a date! I'm in love with her. Did you hear, Charles asked Mary on a date! I heard she turned him down. Why did she say no?	I'm going to ask Mary on a date! I'm in love with her. Did you hear, Charles asked Mary on a date! I heard she turned him down. Why did she wear a dress?	I'm going to ask Mary on a date! I'm in love with her. Did you hear, Charles asked Mary on a date! I heard she turned him down. Why did she not say yes?	I'm going to ask Mary on a date! I'm in love with her. Did you hear, Charles asked Mary on a date! I heard she turned him down. Why did she not wear shoes?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1117	1217	1317	1417
I'm going do a presentation. I'm very nervous! Did you hear, John did a presentation. I heard he failed. Why did he fail?	I'm going do a presentation. I'm very nervous! Did you hear, John did a presentation. I heard he failed. Why did he go by car?	I'm going do a presentation. I'm very nervous! Did you hear, John did a presentation. I heard he failed. Why did he not pass?	I'm going do a presentation. I'm very nervous! Did you hear, John did a presentation. I heard he failed. Why did he not go by train?
1125	1225	1325	1425
I'm going to cancel my date tonight. I don't feel like going. Did you hear, John cancelled his date. I heard he chickened out. Why did he cancel?	I'm going to cancel my date tonight. I don't feel like going. Did you hear, John cancelled his date. I heard he chickened out. Why did he try?	I'm going to cancel my date tonight. I don't feel like going. Did you hear, John cancelled his date. I heard he chickened out. Why isn't he going?	I'm going to cancel my date tonight. I don't feel like going. Did you hear, John cancelled his date. I heard he chickened out. Why doesn't he have a beard?
1135	1235	1335	1435
I'm going to change my name. I'm always being bullied. Did you hear, John changed his name. I heard he's now called Smith. What used to be his name?	I'm going to change my name. I'm always being bullied. Did you hear, John changed his name. I heard he's now called Smith. What did he do?	I'm going to change my name. I'm always being bullied. Did you hear, John changed his name. I heard he's now called Smith. Why didn't he choose to be called Skywalker?	I'm going to change my name. I'm always being bullied. Did you hear, John changed his name. I heard he's now called Smith. Why didn't he start a firm?
1140	1240	1340	1440
I'm going to a psychiatrist. I feel depressed. Did you hear, John is depressed. I heard he is taking a break. How long will he be out?	I'm going to a psychiatrist. I feel depressed. Did you hear, John is depressed. I heard he is taking a break. How long is his hair?	I'm going to a psychiatrist. I feel depressed. Did you hear, John is depressed. I heard he is taking a break. How long will he not work?	I'm going to a psychiatrist. I feel depressed. Did you hear, John is depressed. I heard he is taking a break. How long will he not eat?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1101	1201	1301	1401
I'm going to organize a party. It's next friday! Did you hear, Mike organized a party. I heard he invited almost everyone! Who did he invite?	I'm going to organize a party. It's next friday! Did you hear, Mike organized a party. I heard he invited almost everyone! How fast is he?	I'm going to organize a party. It's next friday! Did you hear, Mike organized a party. I heard he invited almost everyone! Who didn't he invite?	I'm going to organize a party. It's next friday! Did you hear, Mike organized a party. I heard he invited almost everyone! How fast isn't he?
1105	1205	1305	1405
I'm going to buy a new laptop. My old one broke down. Did you hear, Mike bought a new laptop. I heard he bought a MacBook! How much did he pay for it?	I'm going to buy a new laptop. My old one broke down. Did you hear, Mike bought a new laptop. I heard he bought a MacBook! How many trees does he own?	I'm going to buy a new laptop. My old one broke down. Did you hear, Mike bought a new laptop. I heard he bought a MacBook! Why did he not buy a PC?	I'm going to buy a new laptop. My old one broke down. Did you hear, Mike bought a new laptop. I heard he bought a MacBook! How many trees does he not own?
1112	1212	1312	1412
I'm going to be a dad! My wife is pregnant! Did you hear, Mike 's wife is pregnant! I heard he is very excited. When did he tell you?	I'm going to be a dad! My wife is pregnant! Did you hear, Mike 's wife is pregnant! I heard he is very excited. When did he sneeze?	I'm going to be a dad! My wife is pregnant! Did you hear, Mike 's wife is pregnant! I heard he is very excited. Who didn't he tell it to?	I'm going to be a dad! My wife is pregnant! Did you hear, Mike 's wife is pregnant! I heard he is very excited. Who didn't he hate?
1115	1215	1315	1415
I'm going to see a movie. I think an action movie. Did you hear, John went to the cinema. I heard he saw a romantic comedy. Who did he go with?	I'm going to see a movie. I think an action movie. Did you hear, John went to the cinema. I heard he saw a romantic comedy. Who did he ignore?	I'm going to see a movie. I think an action movie. Did you hear, John went to the cinema. I heard he saw a romantic comedy. What did he not see?	I'm going to see a movie. I think an action movie. Did you hear, John went to the cinema. I heard he saw a romantic comedy. What did he not feel?

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1110	1210	1310	1410
I'm going home. See you tomorrow! Did you hear, Charles already left. I heard he took a sick leave. How do you know that?	I'm going home. See you tomorrow! Did you hear, Charles already left. I heard he took a sick leave. How do you travel?	I'm going home. See you tomorrow! Did you hear, Charles already left. I heard he took a sick leave. Why don't I know that?	I'm going home. See you tomorrow! Did you hear, Charles already left. I heard he took a sick leave. Why don't you have an apple?
1123	1223	1323	1423
I'm going to be an artist. I'm very creative! Did you hear, Mike is going to be a teacher. I heard he's moving to New York. When did he decide to be a teacher?	I'm going to be an artist. I'm very creative! Did you hear, Mike is going to be a teacher. I heard he's moving to New York. When is he at home?	I'm going to be an artist. I'm very creative! Did you hear, Mike is going to be a teacher. I heard he's moving to New York. When did he decide not to be an artist?	I'm going to be an artist. I'm very creative! Did you hear, Mike is going to be a teacher. I heard he's moving to New York. When is he not sleeping?
1127	1227	1327	1427
I'm going to surprise Cindy! We have our 1st anniversary today. Did you hear, Mike bought Cindy a present! I heard it was very expensive. Why did he buy her an expensive gift?	I'm going to surprise Cindy! We have our 1st anniversary today. Did you hear, Mike bought Cindy a present! I heard it was very expensive. Why did he drink wine?	I'm going to surprise Cindy! We have our 1st anniversary today. Did you hear, Mike bought Cindy a present! I heard it was very expensive. Why didn't he take her to dinner?	I'm going to surprise Cindy! We have our 1st anniversary today. Did you hear, Mike bought Cindy a present! I heard it was very expensive. Why didn't he go by car?
1129	1229	1329	1429
I'm going to be on tv! The musical I played in is going to be broadcasted. Did you hear, Mike played in a musical. I heard it sucked! Who liked it?	I'm going to be on tv! The musical I played in is going to be broadcasted. Did you hear, Mike played in a musical. I heard it sucked. Who likes to dance?	I'm going to be on tv! The musical I played in is going to be broadcasted. Did you hear, Mike played in a musical. I heard it sucked! Who did not like it?	I'm going to be on tv! The musical I played in is going to be broadcasted. Did you hear, Mike played in a musical. I heard it sucked. Who is not a vegetarian?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1104	1204	1304	1404
I'm going to a great party tonight! It's going to be awesome! Did you hear, Mickey partied all night. I heard he left very late. What time did he leave?	I'm going to a great party tonight! It's going to be awesome! Did you hear, Mickey partied all night. I heard he left very late. What language does he speak?	I'm going to a great party tonight! It's going to be awesome! Did you hear, Mickey partied all night. I heard he left very late. When was he not there anymore?	I'm going to a great party tonight! It's going to be awesome! Did you hear, Mickey partied all night. I heard he left very late. What language doesn't he speak?
1108	1208	1308	1408
I'm going to go out tonight. My wife's taking me. Did you hear, Mickey is going to the opera tonight. I heard he's going with his wife. Why are they going to the opera?	I'm going to go out tonight. My wife's taking me. Did you hear, Mickey is going to the opera tonight. I heard he's going with his wife. Why does he eat breakfast?	I'm going to go out tonight. My wife's taking me. Did you hear, Mickey is going to the opera tonight. I heard he's going with his wife. Why are they not going to see a movie?	I'm going to go out tonight. My wife's taking me. Did you hear, Mickey is going to the opera tonight. I heard he's going with his wife. Why does he not buy a car?
1113	1213	1313	1413
I'm going to win the lottery. This time I know it for sure. Did you hear, Charles won $100 in the lottery! I heard he almost spent it all already. What did he spend it on?	I'm going to win the lottery. This time I know it for sure. Did you hear, Charles won $100 in the lottery! I heard he almost spent it all already. What did he study for?	I'm going to win the lottery. This time I know it for sure. Did you hear, Charles won $100 in the lottery! I heard he almost spent it all aleardy. How much did he not spend?	I'm going to win the lottery. This time I know it for sure. Did you hear, Charles won $100 in the lottery! I heard he almost spent it all already. How much did he not investigate?
1111	1211	1311	1411
I'm going to propose to Sue! We've been together for 5 years! Did you hear, John proposed to Sue. I heard Sue rejected him! Why did she say no?	I'm going to propose to Sue! We've been together for 5 years! Did you hear, John proposed to Sue. I heard Sue rejected him! Why did he buy soup?	I'm going to propose to Sue! We've been together for 5 years! Did you hear, John proposed to Sue. I heard Sue rejected him! Why didn't she say yes?	I'm going to propose to Sue! We've been together for 5 years! Did you hear, John proposed to Sue. I heard Sue rejected him! Why didn't he buy bread?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1118	1218	1318	1418
I'm going to my parents' this weekend. I haven't visited them in ages. Did you hear, Charles is going to his parents this weekend. I heard his aunt died. Where did you hear that?	I'm going to my parents' this weekend. I haven't visited them in ages. Did you hear, Charles is going to his parents this weekend. I heard his aunt died. Why did he lose it?	I'm going to my parents' this weekend. I haven't visited them in ages. Did you hear, Charles is going to his parents this weekend. I heard his aunt died. Why didn't he tell me that?	I'm going to my parents' this weekend. I haven't visited them in ages. Did you hear, Charles is going to his parents this weekend. I heard his aunt died. Why isn't grass red?
1121	1221	1321	1421
I'm going to smile all day! I had great news! Did you hear, Charles was very happy today! I heard he didn't stop laughing. When is he ever sad?	I'm going to smile all day! I had great news! Did you hear, Charles was very happy today! I heard he didn't stop laughing. When is he ever happy?	I'm going to smile all day! I had great news! Did you hear, Charles was very happy today! I heard he didn't stop laughing. When isn't he happy?	I'm going to smile all day! I had great news! Did you hear, Charles was very happy today! I heard he didn't stop laughing. When isn't a girl pretty?
1130	1230	1330	1430
I'm going to take Mary's advice. I'm not going to the party. Did you hear, Charles decided not to go to the party. I heard everyone is very sad. Why was everyone sad?	I'm going to take Mary's advice. I'm not going to the party. Did you hear, Charles decided not to go to the party. I heard Mary told him to. Why did he take Sue's advice?	I'm going to take Mary's advice. I'm not going to the party. Did you hear, Charles decided not to go to the party. I heard everyone is very sad. Why is he not going?	I'm going to take Mary's advice. I'm not going to the party. Did you hear, Charles decided not to go to the party. I heard Mary told him to. Why are all trees not the same?
1137	1237	1337	1437
I'm going on holiday. I think somewhere warm. Did you hear, Charles went on holiday. I heard he is glad to be home. Where did he go?	I'm going on holiday. I think somewhere warm. Did you hear, Charles went on holiday. I heard he is glad to be home. Where did he do his laundry?	I'm going on holiday. I think somewhere warm. Did you hear, Charles went on holiday. I heard he is glad to be home. Why didn't he like it?	I'm going on holiday. I think somewhere warm. Did you hear, Charles went on holiday. I heard he is glad to be home. Why didn't he go to Spain?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1122	1222	1322	1422
I'm going to see a doctor today. My throat hurts. Did you hear, Mickey 's doctor told him to quit smoking. I heard he quit immediately. Why do you start smoking?	I'm going to see a doctor today. My throat hurts. Did you hear, Mickey 's doctor told him to quit smoking. I heard he quit immediately. Why do you stop studying?	I'm going to see a doctor today. My throat hurts. Did you hear, Mickey 's doctor told him to quit smoking. I heard he quit immediately. Why don't you quit smoking?	I'm going to see a doctor today. My throat hurts. Did you hear, Mickey 's doctor told him to quit smoking. I heard he quit immediately. Why aren't you coming?
1128	1228	1328	1428
I'm going to be a stand-up comedian. Everyone thinks I'm very funny! Did you hear, Mickey is going to be a stand-up comedian. I heard he's not funny at all! Why does he think he is funny?	I'm going to be a stand-up comedian. Everyone thinks I'm very funny! Did you hear, Mickey is going to be a stand-up comedian. I heard he's not funny at all! Who told you he cares?	I'm going to be a stand-up comedian. Everyone thinks I'm very funny! Did you hear, Mickey is going to be a stand-up comedian. I heard he's not funny at all! Who thinks he's not funny?	I'm going to be a stand-up comedian. Everyone thinks I'm very funny! Did you hear, Mickey is going to be a stand-up comedian. I heard he's not funny at all! Who thinks he's not a good swimmer?
1133	1233	1333	1433
I'm going to flee the country. I hope they don't catch me. Did you hear, Mickey was imprisoned. I heard he was caught for major fraud. Did he confess his crime?	I'm going to flee the country. I hope they don't catch me. Did you hear, Mickey was imprisoned. I heard he was caught for major fraud. Did he get away?	I'm going to flee the country. I hope they don't catch me. Did you hear, Mickey was imprisoned. I heard he was caught for major fraud. Didn't he flee the country?	I'm going to flee the country. I hope they don't catch me. Did you hear, Mickey was imprisoned. I heard he was caught for major fraud. Didn't he murder someone?
1139	1239	1329	1429
I'm going to grow a beard. I think it looks smart. Did you hear, Mickey now has a beard! I heard it looks great on him! How long is it?	I'm going to grow a beard. I think it looks smart. Did you hear, Mickey now has a beard! I heard it looks great on him! How kind is he?	I'm going to grow a beard. I think it looks smart. Did you hear, Mickey now has a beard! I heard it looks great on him! How much time did he not shave?	I'm going to grow a beard. I think it looks smart. Did you hear, Mickey now has a beard! I heard it looks great on him! How much did he not pay?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1102	1202	1302	1402
I'm going fishing with my dad. It's the first time I go with him. Did you hear, Lucy caught a big fish! I heard she broke some record! How big was it?	I'm going fishing with my dad. It's the first time I go with him. Did you hear, Lucy caught a big fish! I heard she broke some record! How married is she?	I'm going fishing with my dad. It's the first time I go with him. Did you hear, Lucy caught a big fish! I heard she broke some record! How big wasn't it?	I'm going fishing with my dad. It's the first time I go with him. Did you hear, Lucy caught a big fish! I heard she broke some record! How married isn't he?
1119	1219	1319	1419
I'm going to go to Paris! I've always wanted to go there! Did you hear, Lucy visited almost every museum in Paris. I heard she loved it. Which museums did she visit?	I'm going to go to Paris! I've always wanted to go there! Did you hear, Lucy visited almost every museum in Paris. I heard she loved it. Which table did she pick?	I'm going to go to Paris! I've always wanted to go there! Did you hear, Lucy visited almost every museum in Paris. I heard she loved it. Which museums did she not visit?	I'm going to go to Paris! I've always wanted to go there! Did you hear, Lucy visited almost every museum in Paris. I heard she loved it. Which table is not round?
1124	1224	1324	1424
I'm going to buy a boat! I love sailing. Did you hear, Lucy never pays her taxes. I heard she has a fake address. How is that illegal?	I'm going to buy a boat! I love sailing. Did you hear, Lucy never pays her taxes. I heard she has a fake address. How are you?	I'm going to buy a boat! I love sailing. Did you hear, Lucy never pays her taxes. I heard she has a fake address. How isn't that illegal?	I'm going to buy a boat! I love sailing. Did you hear, Lucy never pays her taxes. I heard she has a fake address. How aren't people healthy?
1109	1209	1309	1409
I'm going to stop drinking alcohol. It doesn't do much good. Did you hear, Mickey stopped drinking alcohol. I heard he can't always stay away from the booze! When does he drink?	I'm going to stop drinking alcohol. It doesn't do much good. Did you hear, Mickey stopped drinking alcohol. I heard he can't always stay away from the booze! When is the pizza ready?	I'm going to stop drinking alcohol. It doesn't do much good. Did you hear, Mickey stopped drinking alcohol. I heard he can't always stay away from the booze! When does he not drink?	I'm going to stop drinking alcohol. It doesn't do much good. Did you hear, Mickey stopped drinking alcohol. I heard he can't always stay away from the booze! When is the pasta not cooked?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1126	1226	1326	1426
I'm going to class. I'm already late. Did you hear, Lucy didn't do her homework. I heard she got expelled. Why was she expelled?	I'm going to class. I'm already late. Did you hear, Lucy didn't do her homework. I heard she got expelled. Why is she sick?	I'm going to class. I'm already late. Did you hear, Lucy didn't do her homework. I heard she got expelled. Why did she not do her homework?	I'm going to class. I'm already late. Did you hear, Lucy didn't do her homework. I heard she got expelled. Why did she not quit smoking?
1134	1234	1334	1434
I'm going to sue my boss. I haven't been payed in two months! Did you hear, Lucy sued her boss. I heard she won the law suit. How much did she get?	I'm going to sue my boss. I haven't been payed in two months! Did you hear, Lucy sued her boss. I heard she won the law suit. How much time was she in Japan?	I'm going to sue my boss. I haven't been payed in two months! Did you hear, Lucy sued her boss. I heard she won the law suit. Who doesn't she sue?	I'm going to sue my boss. I haven't been payed in two months! Did you hear, Lucy sued her boss. I heard she won the law suit. Who isn't on tv?
1136	1236	1336	1436
I'm going on a diet. I feel fat. Did you hear, Lucy lost a lot of weight. I heard she's now as skinny as last year. How much weigh did she lose?	I'm going on a diet. I feel fat. Did you hear, Lucy lost a lot of weight. I heard she's now as skinny as last year. How many books did she read?	I'm going on a diet. I feel fat. Did you hear, Lucy lost a lot of weight. I heard she's now as skinny as last year. How long did she not eat candy?	I'm going on a diet. I feel fat. Did you hear, Lucy lost a lot of weight. I heard she's now as skinny as last year. How much did she not spend?
1141	1241	1341	1441
I'm going to eat oysters tonight! I've never tasted them. Did you hear, Lucy ate oysters yesterday. I heard she didn't like it at all! How many did she eat?	I'm going to eat oysters tonight! I've never tasted them. Did you hear, Lucy ate oysters yesterday. I heard she didn't like it at all! How did she arrive there?	I'm going to eat oysters tonight! I've never tasted them. Did you hear, Lucy ate oysters yesterday. I heard she didn't like it at all! Why didn't she like them?	I'm going to eat oysters tonight! I've never tasted them. Did you hear, Lucy ate oysters yesterday. I heard she didn't like it at all! Why didn't she see a doctor?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1106	1206	1306	1406
I'm going home. I'm having diner with my boyfriend Did you hear, Kate 's boyfriend cheated on her. I heard she caught him in the act. How is she handling it?	I'm going home. I'm having diner with my boyfriend Did you hear, Kate 's boyfriend cheated on her. I heard she caught him in the act. How annoying is he?	I'm going home. I'm having diner with my boyfriend Did you hear, Kate 's boyfriend cheated on her. I heard she caught him in the act. Who doesn't know that?	I'm going home. I'm having diner with my boyfriend Did you hear, Kate 's boyfriend cheated on her. I heard she caught him in the act. How strange isn't he?
1107	1207	1307	1407
I'm going to move to Japan. I got a job there! Did you hear, Kate is moving to Japan. I heard she's very excited. How's the weather there?	I'm going to move to Japan. I got a job there! Did you hear, Kate is moving to Japan. I heard she's very excited. How's the grass here?	I'm going to move to Japan. I got a job there! Did you hear, Kate is moving to Japan. I heard she's very excited. When is it not raining there?	I'm going to move to Japan. I got a job there! Did you hear, Kate is moving to Japan. I heard she's very excited. When is it not cloudy here?
1114	1214	1314	1414
I'm going to start a new project. It's about global warming. Did you hear, Kate started a new project. I heard she received a grant. How much did she receive?	I'm going to start a new project. It's about global warming. Did you hear, Kate started a new project. I heard she received a grant. How much did she drive?	I'm going to start a new project. It's about global warming. Did you hear, Kate started a new project. I heard she received a grant. How much didn't she receive?	I'm going to start a new project. It's about global warming. Did you hear, Kate started a new project. I heard she received a grant. How much didn't she drive?
1120	1220	1320	1420
I'm going to make dinner tonight. I think I'll make soup. Did you hear, Kate is a terrible cook! I heard she doesn't know it herself. What can she make?	I'm going to make dinner tonight. I think I'll make soup. Did you hear, Kate is a terrible cook! I heard she doesn't know it herself. What does she build?	I'm going to make dinner tonight. I think I'll make soup. Did you hear, Kate is a terrible cook! I heard she doesn't know it herself. What can't she make?	I'm going to make dinner tonight. I think I'll make soup. Did you hear, Kate is a terrible cook! I heard she doesn't know it herself. What doesn't she write?

Appendix B: Fillers

GOOD/POS	BAD/POS	GOOD/NEG	BAD/NEG
1131	1231	1331	1431
I'm going to buy a new watch. Probably a Rolex. Did you hear, Kate lost her new watch. I heard it was very expensive! When did she lose it?	I'm going to buy a new watch. Probably a Rolex. Did you hear, Kate lost her new watch. I heard it was very expensive! When did she find it?	I'm going to buy a new watch. Probably a Rolex. Did you hear, Kate lost her new watch. I heard it was very expensive! When did she not wear it?	I'm going to buy a new watch. Probably a Rolex. Did you hear, Kate lost her new watch. I heard it was very expensive! When did she not know the time?
1132	1232	1332	1432
I'm going on a blind date. My friend set me up. Did you hear, Kate went on a blind date! I heard it was boring! Who did she meet?	I'm going on a blind date. My friend set me up. Did you hear, Kate went on a blind date! I heard it was boring! Who did she kiss?	I'm going on a blind date. My friend set me up. Did you hear, Kate went on a blind date! I heard it was boring! Why did she not like it?	I'm going on a blind date. My friend set me up. Did you hear, Kate went on a blind date! I heard it was boring! Why did she not cry?
1138	1238	1338	1438
I'm going to start working out. I want to increase by stamina. Did you hear, Kate stopped going to the gym. I heard she only went there for a month. Why did she stop going?	I'm going to start working out. I want to increase by stamina. Did you hear, Kate stopped going to the gym. I heard she only went there for a month. Why did she stop drinking?	I'm going to start working out. I want to increase by stamina. Did you hear, Kate stopped going to the gym. I heard she only went there for a month. Why didn't she like it?	I'm going to start working out. I want to increase by stamina. Did you hear, Kate stopped going to the gym. I heard she only went there for a month. Why didn't she walk back?

Bibliography

Alonso-Ovalle, L. (2006). *Disjunction in Alternative Semantics*. Ph.D. thesis, University of Massachusetts at Amherst, Amherst, MA.

Anderbois, S. (2011). *Issues and Alternatives*. Ph.D. thesis, University of California Santa Cruz.

Asher, N. & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Asher, N. & Reese, B. (2005). 'Negative bias in polar questions.' In E. Maier, C. Bary, & J. Huitink (eds.) *Sinn und Bedeutung (SuB), Nijmegen, 01/11/2004-03/11/2004*, 30–43. University of Nijmegen, Nijmegen.

Borkin, A. (1971). 'Polarity items in questions.' *Chicago Linguistic Society* 7, 53–62.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). 'Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?' *Perspectives on Psychological Science* 6, 3–5.

Büring, D. & Gunlogson, C. (2000). 'Aren't positive and negative polar questions the same?' UCSC/UCLA.

Ciardelli, I. (2010). 'A first-order inquisitive semantics.' In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (eds.) *Logic, Language and Meaning*, volume 6042 of *Lecture Notes in Computer Science*, 234–243. Springer Berlin / Heidelberg.

Ciardelli, I. & Roelofsen, F. (2011). 'Inquisitive logic.' *Journal of Philosophical Logic* 40, 55–94.

Conover, W. & Iman, R.L. (1981). 'Rank transformations as a bridge between parametric and nonparametric statistics.' *The American Statistician* 35, 124–129.

Draine, S.C. & Greenwald, A.G. (1998). 'Replicable unconscious semantic priming.' *Journal of Experimental Psychology* 127, 286–303.

Farkas, D. & Roelofsen, F. (2011). 'Polarity particles in an inquisitive discourse model.' To appear.

Farkas, D.F. (2011). 'Polarity particles in English and Romanian.' UCSC.

Farkas, D.F. & Bruce, K.B. (2010). 'On reacting to assertions and polar questions.' *Journal of Semantics* 27, 81–118.

Field, A. (2009). *Discovering statistics using SPSS*. SAGE Publications Ltd, 3rd edition.

Fort, K., Adda, G., & Cohen, K.B. (2011). 'Amazon Mechanical Turk: Gold Mine or Coal Mine?' *Computational Linguistics* 37, 413–420.

Groenendijk, J. (2009). 'Inquisitive semantics: Two possibilities for disjunction.' In P. Bosch, D. Gabelaia, & J.Lang (eds.) *Seventh International Tblisi Symposium on Language, Logic, and Computation*. Springer-Verlag.

Groenendijk, J. & Roelofsen, F. (2009). 'Inquisitive semantics and pragmatics.' In J. Larrazabal & L. Zubeldia (eds.) *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*. University of the Basque Country Press. URL `www.illc.uva.nl/inquisitive-semantics`.

Groenendijk, J. & Stokhof, M. (1984). *Studies in the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.

Gunlogson, C. (2001). *True to Form: Rising and Falling Declaratives and Questions in English*. Ph.D. thesis, University of California Santa Cruz.

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

Hartung, S. (2006a). 'Forms of negation in polar questions.' Universität Tübingen.

Hartung, S. (2006b). *Forms of Negation in Polar Questions*. Master's thesis, Universität Tübingen.

Horn, L.R. (1989). *A natural history of negation*. University of Chicago Press, Chicago.

Karttunen, L. (1977). 'Syntax and Semantics of Questions.' *Linguistics and Philosophy* 1, 3–44.

Kramer, R. & Rawlins, K. (to appear). 'Polarity particles: an ellipsis account.' In *Proceedings of NELS 39*.

Kratzer, A. & Shimoyama, J. (2002). 'Indeterminate phrases: The view from japanese.' In Y. Otsu (ed.) *The Proceedings of the Third Tokyo Conference on Psycholinguistics*, 1–25. Hituzi Syobo, Tokyo.

Ladd, R.D. (1981). 'A first look at the semantics and pragmatics of negative questions and tag questions.' In *Papers from the 17th Regional Meeting of the Chicago Linguistic Society*, 164–171. Chicago Linguistics Society, University of Chicago, Chicago.

Ladd, R.D. (1996). *Intonational phonology.* Cambridge University Press., Cambridge.

Ladusaw, W. (1979). *Polarity Sensitivity as Inherent Scope Relations.* Ph.D. thesis, University of Texas.

Nilsenova, M. (2006). *Rises and Falls. Studies in the semantics and pragmatics of intonation.* Ph.D. thesis, University of Amsterdam.

Potts, C. (2005). 'Lexicalized intonational meaning.' In S. Kawahara (ed.) *University of Massachusetts Occasional Papers*, volume 30, 129–146. MA:GLSA, Amherst.

Pruitt, K. & Roelofsen, F. (2010). 'Prosody and interpretation of disjunctive questions.' In preparation.

Reese, B. (2006). 'The meaning and use of negative polar interrogatives.' In O. Bonami & P.C. Hofherr (eds.) *Empirical Issues in Syntax and Semantics 6: Selected papers from CSSP 2005*, 331–354. Colloque de Syntaxe et Sémantique à Paris.

Reese, B. (2007). *Bias in Questions.* Ph.D. thesis, University of Texas, Austin.

Roelofsen, F. & van Gool, S. (2010). 'Disjunctive questions, intonation, and highlighting.' In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (eds.) *Proceedings of the 17th Amsterdam colloquium conference on Logic, Language and Meaning*, 384–394. Springer-Verlag.

Romero, M. (2005). 'Two approaches to biased yes/no questions.' In J. Alderete, C.H. Han, & A. Kochetov (eds.) *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 352–360. Cascadilla Press, Somerville, MA.

Romero, M. & Han, C.H. (2004). 'On negative *Yes/No* questions.' *Linguistics and Philosophy* 27, 609–658.

van Rooij, R. (2003). 'Negative polarity items in questions: Strength as relevance.' *Journal of Semantics* 20, 239–273.

van Rooij, R. & Šafářová, M. (2003). 'On polar questions.' In R.B. Young & Y. Zhou (eds.) *Proceedings of SALT XIII*, 292–309. CLC Publications, Cornell University, Ithaca, NY.

Sadock, J. (1974). *Towards a Linguistic Theory of Speech Acts.* Academic Press, New York, NY.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A.Y. (2008). 'Cheap and Fast - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks.' In *Proceedings of the 2008 Conference on EMNLP*, 254 – 263. Association for Computational Linguistics, Honolulu.

Stalnaker, R. (1978). 'Assertion.' In P. Cole (ed.) *Syntax and Semantics*, volume 9, 315–332. Academic Press, New York.

Stevens, J.P. (2009). *Applied Multivariate Statistics for the Social Sciences*. Routledge, 5th edition.

de Swart, H. & Sag, I. (2002). 'Negation and negative concord in romance.' *Linguistics and Philosophy* 25, 373–417.