Aren’t prosody and syntax marking bias in questions?

Journal:	*Language and Speech*
Manuscript ID	LAS-18-0085.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Arnhold, Anja; University of Alberta Faculty of Arts, Department of Linguistics; University of Konstanz, Linguistics Braun, Bettina; University of Konstanz, Linguistics Romero, Maribel; University of Konstanz, Linguistics
Keywords:	yes-no question, speaker certainty, bias, Ladd's ambiguity
Abstract:	As first observed by Ladd (1981. A first look at the semantics and pragmatics of negative questions and tag questions. Proceedings of Chicago linguistic society 17, pp. 164–171), English polar questions with high negation (HiNQ, e.g. Aren’t they adding a menu item?) can be used both to check the speaker’s belief that the proposition p is true (e.g. p = they are adding a menu item) and to check the addressee’s belief that p is not true (¬p). We hypothesized that this ambiguity can be disambiguated prosodically. We further hypothesized that the prosodic disambiguation is absent in German, because the checked proposition can be marked morpho-syntactically, with HiNQs checking p and low negation questions (LowNQ, e.g. Are they not adding a menu item?) checking ¬p. A production study tested these hypotheses with 24 speakers of Western Canadian English and German each (764 and 767 utterances total, respectively). The results showed that, when the speaker originally believed p and the addressee implied ¬p, English speakers preferred HiNQs over LowNQs, confirming Ladd’s observation, and used intonation to mark whose proposition they were checking, as hypothesized. By contrast, German speakers marked this distinction morpho-syntactically, realizing mostly HiNQs to check their own proposition and LowNQs to check the addressee’s proposition. Their prosody, in turn, was largely determined by the morpho-syntactic question form. The study further manipulated the speaker’s certainty of the checked proposition, but, in contrast to studies on Romance languages, found that certainty itself

	was not marked.

Full report

Corresponding Author:

Anja Arnhold, University of Alberta, Department of Linguistics, 4-24 Assiniboia Hall, Edmonton, AB, T6G 2E7, Canada

Email: anja.arnhold@gmail.com

Aren't prosody and syntax marking bias in questions?

Anja Arnhold[1,2] Bettina Braun[2] and Maribel Romero[2]

[1]University of Alberta, Edmonton, AB, Canada

[2]University of Konstanz, Konstanz, Germany

Abstract

As first observed by Ladd (1981. A first look at the semantics and pragmatics of negative questions and tag questions. *Proceedings of Chicago linguistic society* 17, pp. 164–171), English polar questions with high negation (HiNQ, e.g. *Aren't they adding a menu item?*) can be used both to check the speaker's belief that the proposition *p* is true (e.g. *p = they are adding a menu item*) and to check the addressee's belief that *p* is not true (¬*p*). We hypothesized that this ambiguity can be disambiguated prosodically. We further hypothesized that the prosodic disambiguation is absent in German, because the checked

proposition can be marked morpho-syntactically, with HiNQs checking *p* and low negation questions (LowNQ, e.g. *Are they not adding a menu item?*) checking ¬*p*.

A production study tested these hypotheses with 24 speakers of Western Canadian English and German each (764 and 767 utterances total, respectively). The results showed that, when the speaker originally believed *p* and the addressee implied ¬*p*, English speakers preferred HiNQs over LowNQs, confirming Ladd's observation, and used intonation to mark whose proposition they were checking, as hypothesized. By contrast, German speakers marked this distinction morpho-syntactically, realizing mostly HiNQs to check their own proposition and LowNQs to check the addressee's proposition. Their prosody, in turn, was largely determined by the morpho-syntactic question form. The study further manipulated the speaker's certainty of the checked proposition, but, in contrast to studies on Romance languages, found that certainty itself was not marked.

Keywords

yes-no question, bias, Ladd's ambiguity, speaker certainty

1. Introduction

Polar questions (or yes-no questions) are questions that expect an answer of either *yes* or *no*. The form of the question can be positive or negative (e.g. *Are you hungry?* vs. *Aren't you hungry?* or *Are you not hungry?*), and even though positive and negative forms have traditionally received the same semantic representation (e.g. a set containing the checked proposition and its negation), more recently several authors have argued that they are not in fact equivalent in terms of use-conditions (Asher & Reese, 2007; Büring & Gunlogson,

2000; Romero & Han, 2004, 2004; van Rooij & Šafářová, 2003; Sudo, 2013). Instead, the felicity of their use depends both on the speaker's prior beliefs regarding the checked proposition *p* and the new evidence provided by the conversational context. For example, Domaneschi, Romero and Braun (2017) manipulated the original speaker bias (e.g., (1a); speaker bias for the proposition *p*) and contextual evidence (e.g. (1b), evidence for ¬*p*) and found that in such situations with conflicting biases, English participants most often chose negative polar questions with negation in a high position, as in (1c); by contrast, positive polar questions as in (1d) were almost never uttered by participants as a response in the same context and thus seem inappropriate (see Domaneschi et al., 2017, for empirical results on these and other combinations of biases).

(1) Polar questions checking proposition *p* (*p = there is a train in the early morning*)

a. Tomorrow, you need to go from Nottingham to Sheffield very early. Your brother goes there quite frequently and you remember he told you that he usually takes a train in the early morning, before 7:00.

b. You go to the ticket office and you ask for a train for the next morning. The operator answers to you: "The only train available is at 11:00."

c. Isn't there a train in the early morning?

d. #Is there a train in the early morning?

(from Domaneschi et al., 2017, p. 12)

In this article, we further investigate negative questions with high negations such as (1c) in German and English. While these questions have been argued to be ambiguous between two readings in English, their German counterparts are not. We tested whether

the two readings are disambiguated by prosody in English and how the same difference is signalled in German.

As first pointed out by Ladd (1981), English negative polar questions with high negation (HiNQ, e.g. *Isn't there a vegetarian restaurant around here?*) show an interesting ambiguity, later termed 'Ladd's ambiguity' in the semantic literature: Speakers can either use HiNQs to check the proposition *p* that they are biased for (*p = there is a vegetarian restaurant*) or, having just inferred that the addressee seems to believe that the proposition is not true, check this inference that ¬*p*. Ladd's original examples are given in (2) and (3) below. In (2b) Kathleen uses a HiNQ to double-check her original belief that *p* is true. In (3c), Bob, who also had originally assumed that *p* was true, uses the same HiNQ to check ¬*p*, which he has just inferred from Kathleen's previous statement.

(2) HiNQ checking proposition *p* (*p = there is a vegetarian restaurant*)

(Situation: Kathleen and Jeff have just come from Chicago on the Greyhound bus to visit Bob in Ithaca.)

a. Bob: You guys must be starving. You want to go get something to eat?

b. Kathleen: Yeah, isn't there a vegetarian restaurant around here--- Moosewood, or something like that?

c. Bob: Gee, you've heard of Moosewood all the way out in Chicago, huh? OK, let's go there. (from Ladd, 1981, p. 164)

(3) HiNQ checking proposition ¬*p*

(Situation: Bob is visiting Kathleen and Jeff in Chicago while attending CIS.)

a. Bob: I'd like to take you guys out to dinner while I'm here-we'd have time to go somewhere around here before the evening session tonight, don't you think?

b. Kathleen: I guess, but there's not really any place to go in Hyde Park.

c. Bob: Oh, really, isn't there a vegetarian restaurant around here?

d. Kathleen: No, about all we can get is hamburgers and souvlaki.

(from Ladd, 1981, p. 164)

This ambiguity is not present in German, as first pointed out by Büring and Gunlogson (2000). They state that, when the negative question checks proposition *p* as in (2) – 'outer negation' in Ladd's terminology –, the question translates into German as (4), where the negation *nicht* precedes the indefinite determiner *ein.* When the negative question checks the proposition ¬*p* as in (3) – 'inner negation'–, German is said to use the equivalent negative polar question with the negation in a low position (LowNQ) in (5), where the negation and the determiner are merged into *kein* 'none'.

(4) Gibt es hier nicht ein vegetarisches Restaurant? (HiNQ)
Gives EXPL here not a vegetarian restaurant
'Isn't there a vegetarian restaurant around here?'
(from Büring and Gunlogson, 2000, p. 10)

(5) Gibt es hier kein vegetarisches Restaurant? (LowNQ)
Gives EXPL here no vegetarian restaurant
'Is there not a vegetarian restaurant around here?'
(from Büring and Gunlogson, 2000, p. 9)

Hence, while German may disambiguate syntactically between checking *p* vs. ¬*p*, English may not.[1] We hypothesized that, in English, Ladd's ambiguity is disambiguated prosodically, i.e. that the intonation of English HiNQs is different when they are employed to check *p* than when they are used to check ¬*p*. We therefore constructed a production experiment comparing the prosody of English and German negative polar questions – namely HiNQs and LowNs – checking *p* vs. checking ¬*p*.

Although to our knowledge it has not previously been tested whether prosody marks the checked proposition of English HiNQs, an observation by Domaneschi et al. (2017) hints that this may be the case. They showed that in a context in which a speaker with an existing bias for *p* encountered evidence that their addressee believed ¬*p* (as in (1) above), when German participants selected HiNQs in the experiment they uniformly realized them with a nuclear pitch rise, whereas British English participants sometimes used rises and sometimes falls. Domaneschi et al. (2017) hypothesized that German speakers uttering HiNQ always used them to check their original bias for *p*, whereas the two different intonational realizations for English may have been used to mark the difference between speakers checking their original bias for *p* vs. the newly presented bias for ¬*p*. However, since their experimental design did not control which proposition participants checked, the authors were unable to indicate which of the two contours would be used to check which bias.

Note that to test how the checked proposition is marked, it is crucial to use a so-called contradiction scenario where the speaker's original belief is *p* and the addressee implies ¬*p*, as in (1) and (3), and not a so-called suggestion scenario as in Ladd's context given

in (2). It has been shown in the literature that, while HiNQs checking *p* can be used both in contradiction and in suggestion scenarios, HiNQs checking ¬*p* are only felicitous in contradiction scenarios (AnderBois, 2011, p. 187-189; Romero & Han, 2004, p. 618ff.; Sudo, 2013, table 19): In contrast to scenario (2), Bob could utter both *Oh, really, isn't there some vegetarian restaurant around here?* and *Oh, really, isn't there any vegetarian restaurant around here?* in scenario (3). Since we are interested in whether prosody marks the difference between checking *p* and ¬*p* in the case of ambiguity for English HiNQs, the relevant contexts are contradiction scenarios.

1.1 Prosody of polar questions

For both English and German, scholars have suggested various different, but related explanations for differences in the intonation of questions, with most of the discussion centered around the difference between final pitch rises and falls. Rises have been described as signaling interrogativity and falls as making questions more statement-like, with authors generally acknowledging that there is no one-to-one correspondence between intonation and interrogativity and that other factors, such as politeness, play a role (e.g. for English: Halliday & Greaves, 2008; O'Connor & Arnold, 1973; Palmer, 1922; Schubiger, 1958; for German: Altmann, 1984; von Essen, 1964; Féry, 1993; Isačenko & Schädlich, 1966; Pheby, 1975; Zacharias & Stock, 1973; for other prosodic markers of interrogativity also see Michalsky, 2017; Petrone & Niebuhr, 2014). In the following, we are going to summarize more concrete analyses of the meaning of question intonation, first for English and then for German.

For English, Pierrehumbert and Hirschberg (1990) argue that the primary meaning of phrase-final high pitch in English is to signal that the phrase should be interpreted as related to a following phrase, whereas low pitch signals that it should not (see also O'Connor & Arnold, 1973, p. 47; Pike, 1945, p. 51; Peters, 2006, adopts this interpretation for German). A partially overlapping proposal by Bartels (1997b, 2013) claims that final high pitch is associated with continuation dependence, whereas final low pitch is associated with assertiveness. Similarly, Nilsenová (2006) posits an association between final pitch rises and epistemic uncertainty on the part of the speaker. Steedman (2004), expanding a suggestion by Gussenhoven (1983, p. 201-202), interprets final high pitch as the speaker attributing responsibility for the propositional content to the addressee, whereas falls mark a commitment of the speaker (also see Gunlogson, 2001, on falls and rises in declarative questions vs. statements, and Escandell-Vidal, 1998, 1999, 2017, for a similar account of Peninsular European Spanish question contours).

Conversely, another line of research has taken syntactic and pragmatic factors as its starting point and investigated their influence on the prosodic realization of questions. The role of the checked proposition is generally not explicitly addressed in studies describing the prosody of English or German HiNQs. For English, Romero and Han (2002) suggested that the preposed negation in HiNQs is focused *per se* (it contributes the VERUM operator, see also Romero & Han, 2004) and that the complex of auxiliary and *n't* carries a different pitch contour and higher pitch than the auxiliary in a positive polar question. They left open whether to analyse this contour as a rising (L+H*) or falling (H*+L) accent, but stated that it is followed by a low (L*) accent on the focus exponent,

followed by a high (H-H%) boundary tone, i.e. a sentence-final pitch rise. Hedberg and Sosa (2002) compared different question types in a corpus of spontaneous speech, and found that the auxiliary in negative polar questions overwhelmingly carried a rising (L+H*) accent (eight out of nine cases), but in terms of utterance-final pitch movements, they were almost evenly split between final falling and final rising pitch. Nilsenová (2006) also found an equal number of final rises and falls on the nine negative polar questions in her corpus (four HiNQs, five positive polar questions with negative tags), all of which she classified s as biased towards a positive answer. Without distinguishing positive and negative polar questions (or HiNQs and LowNQs), Palmer (1922) stated that polar questions implying an answer of opposite polarity are realized with falling pitch, whereas without such a speaker bias, a final rise is most common. Pierrehumbert and Hirschberg (1990) contrasted the standard polar question contour rising from a low level (L* H-H%) with a rise from a high level (H* H-H%), which they claimed is used more often for confirmation questions. They only provide examples of positive confirmation questions with a positive speaker bias. Based on experimental data from Western Canadian English, Dehé and Braun (2019) showed that information-seeking questions are predominantly realized with a high rise (H-H%), while rhetorical questions are realized with a high plateau (H-L%).

In summary, research on the prosody of English polar questions has described differences between positive and negative (HiNQ) questions, as well as marking of the expected answer. Thus, while both morpho-syntactic question type and bias have been argued to influence the prosodic realization of English questions, the picture is fragmented and incomplete, in particular regarding the role of checked proposition.

For German, no differences between the prosody of positive and negative polar questions or between HiNQs and LowNQs have been described in the literature yet (but for differences between polar questions, *wh*-questions, alternative questions and questions with declarative syntax in their association with final pitch rises vs. falls, see e.g. Altmann, 1984; Bierwisch, 1966; von Essen, 1964; Isačenko & Schädlich, 1966; Kohler, 2004; Pheby, 1975; Stock, 1999). Some authors have interpreted falls as signaling that the speaker is focused on facts, but rises as indicating a focus on the addressee or the relationship between the speaker and addressee (Kohler, 2004; Stock, 1999; Stock & Zacharias, 1973) or suggested that rises signal openness and an allocation of choice to the addressee, whereas falls restrict the choice (Bartels, 1997a; von Essen, 1964; Kohler, 2004), relating the two to extralinguistic factors such as perceived friendliness and interest in the addressee (von Essen, 1964, p. 59-60; Kohler, 2004; Peters, 2006). Crucially for our present topic, Pheby (1975, p. 154-155) states that the intonational realization of polar questions depends on the presence of bias, distinguishing three cases. First, final pitch rises generally mark neutral questions requesting information. Second, a final fall appears when speakers use a question to assure themselves of something or to be assured that the addressees are certain. Third, questions with final falling-rising pitch convey an expectation on the part of the speaker with respect to the truth of the checked proposition. The third case is most relevant in the present context, since it seems to correspond to the cases of interest in our study, i.e. biased polar questions where the speaker checks the truth of a proposition p (or its negation $\neg p$). To illustrate this case, Pheby gives examples of both positive and negative polar questions carrying what he calls a ‘negative’ expectation, i.e. a speaker bias towards answers of the opposite polarity,

namely *Darfst du das?* 'Are you allowed to do that?' expecting that the addressee is not allowed to do it and *Darfst du das nicht?* 'Are you not allowed to do that?'[2] expecting that the addressee is allowed to do it. But he states that the polarity of the expected answer may also match that of the question, supporting his claim with an example of a positive polar question with the same fall-rise contour carrying a positive expectation. Interestingly, Pheby (1975, p. 171) further subdivides this third case, questions with falling-rising intonation, into questions with a smaller vs. a larger pitch range for the fall-rise, with the latter signalling a stronger expectation of an opposite-polarity answer.[3]

In line with that, Batliner (1989) suggests that the size of the final pitch rise in a question can make a question more or less prototypically question-like ("fragehaltig", Batliner, 1989, p. 155), whereby a less question-like question can for example express a bias for a certain answer or signal a lack of interest in the answer, whereas a more question-like one is neutral with respect to the expected answer and shows strong interest in the response. Note, however, that Batliner is discussing gradient differences in meaning here, which according to him can only be employed when the size of the final pitch rise is not needed to mark a categorical distinction. Thus, prosody can modify how prototypical a morpho-syntactic polar question is, but cannot be used in the same way when it is needed to mark the contrast between question and declarative in so-called declarative questions, which are morpho-syntactic statements and marked as questions only by prosody (Batliner, 1989, p. 155-157).

Similarly, Kügler (2003) observed that positive polar questions in semi-spontaneous and spontaneous Upper Saxon German were always realized with rising nuclear accents (L*H), but that these accents were followed by a final pitch rise for information-seeking

questions, i.e. questions without a bias, and by a pitch fall for confirmation-seeking questions, i.e. questions with a positive speaker bias. He specifically noted that the absence of falling accents in polar questions in his Upper Saxon data constitutes a difference from Standard German.[4]

German hence uses the final contour to mark bias in polar questions (Pheby, 1975; Kügler, 2003) and seems to use gradual differences in pitch range to mark the strength of an expectation of an opposite-polarity answer (Pheby, 1975). There are no experimental data, however, on whether the checked proposition affects the morpho-syntactic question form or the prosodic realization.

Similar pragmatic differences have been found to affect the prosody of polar questions in various other languages. In particular, a prosodic distinction between information-seeking and confirmation-seeking questions or between biased and non-biased questions has been found in experimental studies for several varieties of Romance languages, with authors frequently pointing to the relevance of other factors like the speaker's confidence in inferred information or the strength of their bias, also termed speaker certainty or commitment (e.g. Armstrong, 2017; Armstrong & Prieto, 2015, for Puerto Rican Spanish; Prieto & Borràs-Comes, 2018, for Central Catalan; Grice & Savino, 1995, 2003, for Bari Italian; Michelas, Portes, & Champagne-Lavau, 2016, for French; Vanrell, Mascaro, Torres-Tamarit, & Prieto, 2013, for Majorcan Catalan; Vanrell, Ballone, Schirru, & Prieto, 2014, for Sardinian).

In sum, various factors have been argued to influence the intonation of polar questions in the literature. We believe that the most important factors are: (i) syntactic form or question type (e.g. HiNQ vs. LowNQ), (ii) pragmatic function (neutral information-

seeking use vs. biased questions seeking reassurance/confirmation or expecting a given answer), (iii) the degree of the speaker's certainty or commitment to the proposition in the question, which may be related to more interpersonal aspects like the speaker's perceived friendliness, surprise or disbelief and (iv) the propositional content – p vs. $\neg p$ – checked by the question. Importantly, while most authors acknowledge the relevance of several factors, experimental studies rarely control more than one factor at a time and often implicitly or explicitly conflate factors that can at least conceptually be differentiated. Furthermore, factor (iv) is at this point a suggestion in Domaneschi et al. (2017) and has not been tested experimentally.

1.2 Experiment

Here, we report an experiment on negative polar questions that independently manipulated and crossed the checked proposition (speaker's/p vs. addressee's/$\neg p$) and the speaker's certainty regarding the checked proposition (high vs. low level of certainty) to investigate morpho-syntactic marking (use of HiNQ vs. LowNQ) and prosodic realization (distribution of boundary tones and accent types). Thus, we controlled factor (ii) from section 1.1. to be constant for all target items (biased questions used to resolve a conflict between speaker's and addressee's beliefs) while systematically varying factors (iii) and (iv) and investigating the effects on morpho-syntactic question type (i) and prosody (as well effects of question type on prosody). We concentrated on English and German negative polar questions, which present an ideal test case for these factors due to the differences between the languages described above with respect to Ladd's ambiguity. Based on previous studies, we tested the following hypotheses:

1. Speakers mark which proposition is being checked by their question. German uses question type, English uses prosody.
    a. English speakers use HiNQs both to check the speaker's proposition *p* and to check the addressee's proposition ¬*p* (e.g. Ladd, 1981).
    b. English speakers use prosody, in particular final intonation, to mark the checked proposition (suggestion in Domaneschi et al., 2017).
    c. German speakers use HiNQs to check their own proposition *p* and LowNQs to check the addressee's proposition ¬*p* (Büring & Gunlogson, 2000)
2. The speaker's level of certainty regarding the checked proposition influences the prosodic realization of the question (e.g. Pheby, 1975, for German; for Romance languages Armstrong, 2017; Armstrong & Prieto, 2015; Prieto & Borràs-Comes, 2018; Michelas et al., 2016).

Note that since effects of speaker certainty have not been directly investigated for either German or English, we left the precise nature of this factor's influence open in hypothesis 2. Similarly, it was not possible to specify the nature of the prosodic marking of the checked proposition in English in hypothesis 1b, since it has not previously been investigated. In this paper we focus on categorical differences in boundary tones and accent (placement and type).

2. Method

2.1 Participants

2.1.1 English

For the English version of the experiment, 34 participants were recruited at the University of Alberta in Edmonton, Canada. Of these participants, we excluded seven participants for the following reasons: four were native speakers of other varieties of English than Western Canadian, two were bilingual or heritage speakers of another language and one participant was both. Of the remaining 27 participants, we chose 24 such that each of the two basic experimental lists (see below) contained an equal number of participants. The data that were analyzed came from 24 monolingual speakers of Western Canadian English (15 female, 8 male, 1 other; age 18-33, mean 21.7).

Participants received partial course credit or CAD15 as reimbursement for their time. The study was reviewed and approved by the Research Ethics Board 2 of the University of Alberta.

2.1.2 German

For the German version of the experiment, 32 participants were tested at the University of Konstanz, Germany. Except for one, all of them were native speakers of German who had not acquired any other language before age six. Data from 24 monolingual speakers of German were evaluated to match the number of speakers in the English data (20 female, 4 male; age 18-30, mean 22.5). Again, evaluated participants were equally distributed across experimental lists. All participants received EUR8 as reimbursement for their time. The study was reviewed by the ethics board at the University of Konstanz.

2.2 Procedure and Materials

The experiment manipulated two factors, the proposition checked by asking a negative polar question – HiNQ or LowNQ – and the speaker's level of certainty regarding the checked proposition, which were crossed to create four experimental conditions (A: checking p, 90% sure, B: checking p, 60% sure, C: checking ¬p, 60% sure, D: checking ¬p, 90% sure). Both factors were manipulated within-subjects. Procedure and equipment were the same for both languages.

Participants read descriptions of dialogue scenarios between two fictional characters and were told to imagine themselves as one of the characters (see Table 1 for an English example item).

The text appeared paragraph by paragraph (rows 1.-6. in Table 1), with participants pressing a button on a keypad to proceed when they had read the paragraph and were ready for the next one. After a short introductory statement (1.), the second paragraph stated the participant's positive bias towards a proposition *p*, which was made stronger in conditions A and B (2., here: *p = they are planting a water lily*) than in conditions C and D. Next, the interlocutor in the scenario, their addressee, was introduced as having a certain level of expertise with respect to *p,* which was lowest in condition A and highest in condition D (3.). The addressee's belief that ¬*p* was then introduced in all conditions (4.) before the participant was prompted to resolve the contradiction with a question (5.). The participant's belief was always *p*, while the addressee's belief was ¬*p,* but conditions varied in the following way: In conditions A and B, participants were instructed to utter a question to check their own belief *p*, whereas they were told to check the addressee's belief ¬*p* in conditions C and D (below: 'checked proposition'). Additionally, they were told about their level of certainty of the *checked* proposition, which was 90% for

conditions A and D and 60% for conditions B and C (below: 'certainty'). Finally, participants saw a HiNQ and a LowNQ checking *p*/¬*p* and had to choose one of them by uttering it out loud (6.; participants did not see the labels HiNQ and LowNQ).

Table 1. English example item in the four conditions: A (checking p, 90% sure), B (checking p, 60% sure), C (checking ¬p, 60% sure) and D (checking ¬p, 90% sure).

	Condition A	Condition B	Condition C	Condition D
1.	Your next-door neighbours, the Kapoors, have a very nice garden. Looking out of the kitchen window, you have just spotted them in a corner of the garden by the pond.			
2.	You are sure that they are planting a water lily.		You think that they are planting a water lily.	
3.	Your husband, who really needs new glasses, has noticed the Kapoors, too.	Your husband, who probably needs new glasses some time soon, has noticed the Kapoors, too.	Your husband, who has excellent eyesight, has noticed the Kapoors, too.	Your husband, who has better eyesight than you, has noticed the Kapoors, too.

4.	He says: “Look! Our neighbours are just lounging in the sun for once. Should we invite them over for a tea?”			
5.	You are surprised to hear this, but you are still 90% convinced you are right. However, to make sure, you want to check your assumption. You ask:	You are surprised to hear this, but you are still 60% convinced you are right. However, to make sure, you want to check your assumption. You ask:	You are surprised to hear this, but you are 60% convinced that he is right. However, to make sure, you want to check his assumption. You ask:	You are surprised to hear this, but you are 90% convinced that he is right. However, to make sure, you want to check his assumption. You ask:
6.	Aren’t they planting a water lily? [HiNQ] Are they not planting a water lily? [LowNQ]			

There were 16 different experimental scenarios of this structure for English and German each. The target questions all followed the same pattern, given in Table 2, where V denotes a verbal stem and N a compound noun. Compound nouns were all quadrisyllabic, with primary stress on the first syllable and secondary stress on the third syllable; they contained only voiced segments to facilitate prosodic analysis. Verbs were less

prosodically controlled, but consisted of at least two syllables (see Appendix A for the complete list of target questions). We also controlled information structure to ensure that participants used the question to check the truth of the proposition *p*/¬*p* as a whole and all questions were uttered with broad/VP focus (instead of a possible interpretation with narrow focus on the final compound noun). To this end, we took particular care in constructing the text introducing the addressee's belief that ¬*p*. For example, for the item illustrated in Table 1, we made sure that the addressee's turn in 4. established a conflict with the speaker's belief not only with respect to what the neighbours are planting (leading to narrow focus on the compound noun *water lily* in 6.), but with respect to whether the neighbours are doing anything at all (leading to broad/VP focus in 6.).

Table 2. Structure of target questions. V denotes a verbal stem, N a compound noun.

	HiNQ	LowNQ
English	Aren't they V-ing a(n) N?	Are they not V-ing a(n) N?
German	V-en sie nicht eine(n) N? V-INF they not a N?	V-en sie keine(n) N? V-INF they not.a N?

The 16 items x 4 conditions were split up to create two basic lists, which in turn consisted of two experimental blocks each. Conditions and items were counterbalanced so that each list contained every item twice, once in each experimental block, in different conditions. Each condition appeared for overall eight trials per list. Experimental items were interspersed with 15 fillers presenting scenarios similar to the experimental items, but offering different question types as response options. Five fillers gave participants the

choice between a simple positive polar question and a positive polar question including *really*, five between a simple positive polar question and a declarative question, and five between a positive polar question including *really* and a declarative question. Since these question types cannot be used to check the speaker's own proposition in contradictions scenarios, filler items only occurred in two conditions, similar to conditions C and D for experimental items. Like the target items, fillers were counterbalanced across lists and participants encountered each filler in each block, i.e. twice per experimental session in different conditions. Thus, an experimental session contained 16 x 2 experimental trials and 15 x 2 filler trials, totalling in 62 trials, preceded by four practice trials. One of the practice trials was similar to an experimental trial and three were similar to filler trials. Within experimental blocks, items were presented in a different random order for each participant. The experiment was conducted in conjunction with a second study, counterbalancing the order of the experiments between participants, so that a complete experimental session took about 45 minutes.

To keep participants engaged and attentive and to encourage them to read the contexts properly, one third of the trials were followed by questions on whose proposition they were checking or on their level of surprise or by comprehension questions on details about the context. The English participants scored 89% correct on comprehension questions, the German participants 81%, which suggests that they were reading the contexts closely.

English- and German-speaking participants were tested in a sound-attenuated booth at the Centre for Comparative Psycholinguistics at the University of Alberta and at the PhonLab

at the University of Konstanz, respectively. Stimulus presentation and data collection were controlled with the software Presentation (Neurobehavioral Systems, 2014). The productions were directly recorded on the computer (44.1kHz, 16Bit) using a head-mounted unidirectional dynamic Shure SM10A microphone with an audio interface audient iD14. Acoustic measurements were performed with Praat (Boersma & Weenink, 2016).

3. Data treatment and analysis

Of the 768 target trials for each language (16 items × 2 repetitions × 24 participants), three were discarded in the English version of the experiment because the participants did not choose either of the two questions, but instead responded with a declarative question. A further trial was lost because the participant pressed the button to move on to the next trial too early, so the programme did not record their response. We thus evaluated responses to 764 trials. For the German data, one trial had to be discarded because the participant spoke so fast and silently that her response could not be analyzed. Thus 767 target trials were included in the analysis.

3.1. Morphosyntactic and prosodic analysis

The productions were first coded as HiNQ or LowNQ. Then, the first author, a native speaker of German who had lived in Canada for over two years at that point, annotated the pitch contours of all English recordings according to the Tone and Break Indices guidelines for American English, MAE-ToBI (Silverman et al., 1992; Veilleux et al.,

2006), marking presence and type of accents and boundary tones (for the autosegmental-metrical model of prosody on which ToBI annotations are based, see e.g. Pierrehumbert, 1980; Ladd 1996). Two native German speaking research assistants performed corresponding annotations for the German data following the German GToBI guidelines (Grice et al., 2005; Grice, Baumann, Ritter, & Röhr, 2017). We checked the reliability of the annotation by pseudo-randomly selecting 15% of the data (115 utterances per language, including utterances from all participants, utterances with all items and conditions) for a new annotation by a different annotator, the first author for the German utterances and a native English-speaking research assistant for the English ones. The new annotators were first trained with a similarly selected set of 24 utterances. The agreement on the 115 test utterances of these new annotations with respect to the prosodic characteristics analyzed statistically below was 74% for English and 75% for German, corresponding to Cohen's Kappa (κ) values of 0.67 and 0.70, respectively (Cohen, 1960; Carletta, 1996; calculated with R package irr by Gamer, Lemon, Fellows, & Singh, 2019).[4] Evaluating only agreement in whether each annotator had assigned the most frequent prosodic category or not, i.e. the classification that formed the basis for statistical analyses reported below, yielded an agreement of 87% (κ = 0.73) for English and 83% (κ = 0.65) for the German annotations. The overall values obtained with either method of evaluation, with κ ranging between 0.60 and 0.80, represent a "[r]eliable distinction with substantial agreement" (Breen, Dilley, Kraemer, & Gibson, 2012, p. 293, and references therein).

3.2 Statistical analysis

We analyzed the data with (generalized) linear mixed-effects models in R (R Core Team, 2017), using the package lme4 (Bates, Mächler, Bolker, & Walker, 2015; Bates et al., 2016). We additionally employed the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016) to obtain p-values based on Satterthwaite's approximation of degrees of freedom. Linear mixed-effects models are a type of linear regression modelling that can for example take into account associations between several data points coming from the same participant by fitting by-participant random slopes and/or intercepts in addition to modelling effects of predictor variables as fixed effects (Baayen, Davidson, & Bates, 2008). As dependent measures, we analyzed the following binary response variables by fitting separate binomial generalized linear mixed-effects models (with a binomial linking function) to them: participants' choice of morpho-syntactic question types (HiNQ vs. LowNQ), the final boundary tone (final high rise vs. no final high rise), and the realization of accents on all constituents (most frequent vs. other realization). We used the default coding in R for these two-level factors (dummy coding). To validate our categorization of utterance final boundary tones (see footnotes 5 and 6 below), we additionally fit linear mixed-effects models to the numeric response variable utterance-final f0 height in semitones (st). Model fitting always started with a model with random intercepts for participant and item to account for inherent differences between participants and items. The fixed effects of the first model included an interaction between proposition checked (own proposition/p vs. addressee's proposition/$\neg p$) and certainty (high/90% vs. low/60%), as well as the factor trial number (centered by subtracting the mean) to account for possible changes in behavior over the course of the experiment. For models of prosodic dependent variables (final boundary tone, accents

and f0 measures), we additionally included the morpho-syntactic form of the question (HiNQ vs. LowNQ) as a predictor (below: 'question type'), as well as a three-way interaction between question type, proposition and certainty. Starting from this model, we then added random slopes for the predictors question type (for prosodic response variables only), trial, checked proposition, certainty and the interaction between them and compared the resulting models to the original one using the ANOVA function, until we found the most complex random effects structure providing an improved fit to the data (forward fitting). If no random by-item slopes were found to improve model fit, we additionally checked whether the by-item intercept could be removed altogether. Most of our best models included random intercepts to account for effects of participant and item, but no random slopes. Where models had a more or less complex random structure, we will report this below.

Then we backward-fitted the fixed effects by removing interactions and predictors not associated with significant effects, as long as this did not significantly decrease model fit. Finally, we performed model criticism on the resulting best model by inspecting the residuals. Where we found outlier residuals deviating more than ±2.5 standard deviations from 0 (the regression line), we trimmed the model excluding these data points and refitting, to make sure reported effects were not driven by outliers. We report the resulting best models below, stating the number of removed data points where trimming was necessary. Where the random effects structure is not explicitly described, it consists of random intercepts for participants and items without random slopes (see Baayen, 2008, ch. 7; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017, on model fitting procedures).

4. Results

3.1 Morpho-syntax

As predicted based on Ladd's ambiguity, participants in the English version of the experiment showed an overall preference for HiNQ, choosing it over LowNQ in 62% of target trials. As Figure 1 illustrates, this preference appeared for all conditions and was influenced neither by the speaker's certainty nor by whether they asked the question to check their own proposition or that of the addressee. Statistical modelling indicated no significant effects of proposition checked, certainty or an interaction between them. The best linear mixed-effects model of the participants' choice of question type contained an insignificant effect of checked proposition, $\beta = 0.006$, $SE = 0.314$, $z = 0.02$, $p = 0.985$, kept in the model since random effects included a by-participant effect slope for checked proposition. Including certainty as a predictor or an interaction between the two factors did not significantly improve model fit (the random by-item intercept was removed for the same reason). The fixed effects only showed a significant effect of trial, $\beta = -0.026$, $SE = 0.004$, $z = -5.59$, $p < 0.001$, suggesting that participants were overall more likely to choose LowNQs later in the experimental session.

To confirm that participants indeed showed an overall preference for HiNQs, we performed two separate one-sided one-sample t-tests over participant and item averages for the percentage of HiNQ choices. These tests indicated that the percentage of uttered HiNQs was indeed significantly above the chance level of 50%, $t(23) = 3.8$, $p < 0.001$ by subjects; $t(15) = 5.8$, $p < 0.001$ by items.

[insert Figure 1]

[insert Figure 2]

For the German participants, by contrast, a significant interaction appeared between checked proposition and certainty, β = 0.767, *SE* = 0.333, *z* = 2.31, *p* = 0.021 (see Figure 2). The best linear mixed-effects model additionally contained a significant main effect of checked proposition, β = -1.957, *SE* = 0.390, *z* = -5.01, *p* < 0.001, indicating that participants responded with HiNQs less often when checking the addressee's proposition ¬*p* than when checking their own proposition *p*. The main effect of certainty was not significant, β = -0.161, *SE* = 0.232, *z* = -0.69, *p* = 0.487, but the interaction showed that participants' preference for checking *p* with a HiNQ and ¬*p* with a LowNQ was stronger when they were 90% certain than when they were only 60% certain (the best model again included a by-participant effect of checked proposition).

To assess directly whether the effects of the manipulated variables differed between the languages, we combined both data sets and tested whether the three-way interaction between checked proposition, certainty and language was significant. However, this interaction did not significantly contribute to model fit, and the best model included only a significant interaction between language and proposition, β = -1.473, *SE* = 0.244, *z* = 6.03, *p* < 0.001, suggesting that the effect of proposition – participants chose less HiNQ when checking the addressee's proposition than when checking their own – was larger in the German than in the English data (in addition, the model included insignificant main effects of proposition, β = -0.020, SE = 0.244, z = -0.08, p = 0.934, and language, β = -0.186, SE = 0.166, z = 1.12, p = 0.263). This statistically corroborates the differences in the effect of checked proposition for German and English participants suggested by the

separate models reported above: German speakers marked the checked proposition morpho-syntactically, but English speakers did not.

3.2 Prosody

3.2.1 Final boundary tone

Participants uttered overall 85% of the English questions with a final f0 rise annotated as H-H% boundary tones (N = 648), while only 10% of utterances showed the high plateau H-L% (N = 80), 3% the low L-L% (N = 20) and 2% the low-rising L-H% (N = 16). The dominance of H-H% endings was similar for both question types (82% for HiNQs; 89% for LowNQ), for both checked propositions (82% for speaker's/p vs. 88% for addressee's/$\neg p$) and both levels of certainty (85% for both levels). However, it was apparent that some of these final H-H% rises were clearly higher, moving from a low point to the top of the participant's range (see Figure 3). Other utterances were annotated with H-H% boundary tones, but still did not reach the same level (Figure 4). Therefore, we systematically categorized the difference between high rises and smaller rises during the annotation (this distinction was also used by the second annotator and evaluated in the reliability check reported in section 3.1 above). Overall, 420 H-H% rises were classified as high rises and 228 as smaller rises (55% vs. 30% of the total data). We leave open whether this distinction truly reflects a further categorical bifurcation within the MAE-ToBI category H-H%. An alternative explanation could be that we underutilized the H-L% category, though small H-H% rises generally showed final rises and were auditorily quite distinct from the final plateaus annotated as H-L% (compare Figures 4 and 5). Crucially, a clear difference appeared between utterances annotated as high-rising H-H%,

which were the overall most frequent realization in the data, and all other boundary realizations. For statistical analysis, we therefore employed a binary classification into utterances with and without high final f0 rises, where the latter category collapsed H-H% boundary tones resulting in smaller rises with the categories H-L%, L-L% and L-H%.[65]

[insert Figure 3]

[insert Figure 4]

[insert Figure 5]

The best linear mixed-effects model of this binary variable included a significant interaction between checked proposition and certainty, β = -0.895, *SE* =0.325, *z* = -2.75, *p* = 0.006, see Figure 6. Overall, high final rises were significantly more frequent when participants checked the addressee's proposition ¬*p* than when checking their own proposition *p*, β = 1.584, *SE* = 0.238, *z* = 6.64, *p* < 0.001, but this difference was smaller for low certainty than for high certainty. By contrast, the main effect of certainty was marginal, β = 0.394, *SE* = 0.223, *z* = 1.77, *p* = 0.077. The factor morpho-syntactic question type did not significantly improve model fit and was therefore not included in the best model. Evaluating only utterances annotated with H-H% boundary tones yielded essentially the same results, with the best model showing the same interaction between proposition checked and certainty, β = -0.916, *SE* =0.371, *z* = -2.47, *p* = 0.014, and the same significant effect of proposition checked, β = 1.716, *SE* =0.278, *z* = 6.17, *p* < 0.001 (as well as an insignificant main effects of certainty, β = 0.327, *SE* = 0.246, *z* = 1.33, *p* = 0.183, question type, β = -0.351, *SE* = 0.205, *z* = -0.247, *p* = 0.805, and trial, β = -0.006, *SE* = 0.007, *z* = -0.795, *p* = 0.427; question type was retained because removing this

factor resulted in convergence issues, while trial was included because by-item random slope of trial improved model fit).

[insert Figure 6]

[insert Figure 7]

In annotating the German data, we directly distinguished high final f0 rises from less high rises. In the GToBI system, this difference is reflected in the presence of two distinct boundary tone combinations: H-^H%, a rise followed by a high plateau (H-) and a further rise perceived as high rising (^H%), contrasts with H-%, a simple rise ending in a plateau. Overall 53% of the utterances were annotated with the H-^H% tones indicating a final high rise, 21% with H-%, 11% with L-H%, 8% with L-% and for 7% of the utterances, the annotators were unsure about the final boundary tone. We evaluated a binary distinction between final high rises (H-^H%) and all other contours (see Figure 7).[67]

In contrast to the English results, the best binomial linear mixed-effects model only included a significant effect of syntactic question type, $\beta = 1.111$, $SE = 0.287$, $z = 3.87$, $p = 0.002$. It indicated that participants produced more final high rises when uttering a HiNQ than when uttering a LowNQ (recall that HiNQs, in turn, were selected more often when participants checked their own proposition *p* than when they checked the addressee's proposition $\neg p$). The best model additionally included a random by-participant slope of question type.

To directly assess the differences between the languages, we tested for a four-way interaction between proposition, certainty, question type and language. However, this model did not converge, indicating that there were not enough data points to model this

kind of complexity. Therefore, we fitted a simpler model, which showed a significant three-way interaction between proposition, certainty and language, β = 1.196, *SE* = 0.475, *z* = 2.52, *p* = 0.012, suggesting that the interaction between proposition and certainty, which appeared as significant for the intercept language English, β = -0.894, *SE* = 0.326, *z* = -2.75, *p* = 0.006, was significantly weaker for the German data, in line with the findings reported above. The model also contained an interaction between question type and language, β = 1.209, *SE* = 0.259, *z* = 4.66, *p* < 0.001, while the main effect of question type was not significant, β = -0.174, *SE* = 0.175, *z* = -0.99, *p* = 0.322, matching the finding of a significant effect of question type for German, but not for English.

In sum, this section found clear differences between English and German in the realization of utterance-final boundary tones: English participants showed an effect of proposition checked – realizing higher final f0 rises when checking their addressee's proposition *¬p* than for their own proposition *p* –, which was further modulated by their certainty of that proposition, whereas the occurrence of high final f0 rises in German was only predicted by morpho-syntactic question type, with more high final rises for HiNQs than LowNQs.

3.2.2 Accents

Auxiliaries in the English utterances (*are*; note that only English items contained auxiliaries, so that this category was absent from the German data) were the only constituents to show an effect of one the manipulated variables. Auxiliaries were most frequently realized with a rising accent with a delayed peak (L+H*<, *N* = 383, i.e. 50% overall). For statistical analyses, the dependent variable coded whether auxiliaries were

realized with the most frequent accent category L+H*< or any of the other categories. Binomial linear mixed-effects modelling indicated that this accent choice was influenced by the factor proposition checked: Participants used L+H*< accents slightly more often when checking their addressee's proposition $\neg p$ than when checking their own proposition p, $\beta = 0.686$, $SE = 0.296$, $z = 2.32$, $p = 0.020$ (see Figure 8). The best model additionally contained a significant main effect of question type, associated with a much larger estimate, $\beta = 8.873$ $SE = 1.289$, $z = 6.89$, $p < 0.001$, showing that L+H*< appeared more often in HiNQs than in LowNQs, which may be connected to the accentuation of negations (24 data points or 3% trimmed, the best model did not contain a random intercept for item). Less frequent realizations consisted mostly of unaccented auxiliaries ($N = 268$, i.e. 35% overall), while all other accents together accounted for only 15% ($N = 113$).

[insert Figure 8]

In addition to auxiliaries, we evaluated the accents on the sentence-final compound nouns (e.g. *water lily*), main verbs (e.g. *planting*) and negations (*n't* and *not* in English, *kein* 'none' and *nicht* 'not' in German). Below, we present short summaries of the accentuation of these constituents, see Appendix B for further details. The subject constituent (*they* in English, *sie* 'they' in German) was never accented and is thus not further considered here. The realization of the accents on the other constituents did not show an effect of proposition, certainty or an interaction between them in either the English or the German data set. Instead, accentuation of all other constituents (except finite verbs, which did not show any significant differences), was determined by morpho-syntactic question type for both languages, though reasons for this seem more trivial for

English (see Appendix B for statistical analyses using a binary dependent variable that coded whether each word was produced with the most frequent realization or not). Figures 9 and 10 illustrate utterances where all constituents were realized with their most frequent realizations for English and German, respectively.

[insert Figure 9]

[insert Figure 10]

English negations were frequently accented in LowNQs, but never in HiNQs (overall 65% unaccented, *N* = 493). This is expected, as negations in HiNQs are cliticized to the auxiliary (*Aren't*) and thus do not contain a vowel of their own that could carry an accent. Among the overall 271 negations realized with an accent, the most common accent category was L* (*N* = 119, i.e. 16% overall), while all other accents constituted overall 20% of the realizations (*N* = 152).

Compound nouns in the English data carried L* accents most of the time (*N* = 683, 89% overall), but even more so in HiNQs than in LowNQs. Among the compound nouns not realized with the L* accent, most were unaccented (*N* = 61 or 8% of all realizations), while other accent categories constituted about 3% of all realizations (*N* = 20).

Among the main verbs, 34% had a low accent (L*, *N* = 258), 26% a high accent (H*, *N* = 200), 18% a rising accent with a delayed peak (L+H*<, *N* = 140), 14% no accent (*N* = 108), while all other accents together comprised 8% of the data (*N* = 58). No significant difference appeared between HiNQs and LowNQs.

Turning to the German data, overall 33% of the negations were unaccented (*N* = 253). However, while unaccented negations formed the largest category in HiNQs (54%, *N* = 216), they made up only overall 10% of productions with LowNQs (*N* = 37). Instead,

negations in LowNQs were most often realized with the rising accents L+H* (34%, *N* = 126) and L*+H (24%, *N* = 90), which in turn appeared for HiNQs in 22% (*N* = 87) and 4% (*N* = 16) of the cases, respectively. Note that this difference between HiNQs and LowNQs cannot be explained in the same way as for English, since both negation words are accentable in terms of their segmental make-up.

For German sentence-final compound nouns, the accent L*+H was the overall most frequent realization for final compound nouns in both question types (*N* = 435, i.e. 57%), though it accounted for 73% in HiNQs and only for 39% in LowNQs (*N* = 290 vs. *N* = 145). Less frequent realizations of final compound nouns were L* accents (overall 20%, *N* = 157), L+H* accents (overall 10%, *N* = 74), no accent (overall 7%, *N* = 55) and H* accents (overall 6%, *N* = 45).

German verbs appeared most often with a high H* accent (*N* = 349, i.e. 46%) or a low L* accent (*N* = 203, i.e. 26%), while unaccented realizations made up 12% (*N* = 90) and all other accents combined 16% (*N* = 125).

In summary, accentuation of most constituents was influenced by morpho-syntactic question type in English and German. In contrast to the results for German, the effect was either small (utterance-final compounds) or had a straightforward explanation (negations and auxiliaries) for the English data. The only significant effect of proposition checked – albeit small – appeared for English auxiliaries, which were realized with L+H*< accents (vs. unaccented realizations) more often when checking the addressee's proposition. No effects of certainty and no interaction between the two factors appeared in either language.

4. Summary and Discussion

In this section, we first examine the hypotheses in light of the results of our experimental study that orthogonally manipulated checked proposition and speaker certainty for negative polar questions in English and German (section 4.1). Next, we compare our participants' productions to previous descriptions of the prosody of negative polar questions in these languages (section 4.2). Finally, we relate our results to accounts of intonational meaning (section 4.3) before ending with a conclusion (section 4.4).

4.1 Evaluation of hypotheses

Our study was designed to test the following hypotheses:

1. Speakers mark which proposition is being checked by their question. German uses question type, English uses prosody.
    a. English speakers use HiNQs both to check the speaker's proposition *p* and to check the addressee's proposition $\neg p$ (e.g. Ladd, 1981).
    b. English speakers use prosody, in particular final intonation, to mark the checked proposition (suggestion in Domaneschi et al., 2017).
    c. German speakers use HiNQs to check their own proposition *p* and LowNQs to check the addressee's proposition $\neg p$ (Büring & Gunlogson, 2000)
2. The speaker's level of certainty regarding the checked proposition influences the prosodic realization of the question (e.g. Pheby, 1975, for German; for Romance languages Armstrong, 2017; Armstrong & Prieto, 2015; Prieto & Borràs-Comes, 2018; Michelas et al., 2016).

In line with hypothesis 1, we found experimental support for 'Ladd's ambiguity': Speakers of English use negative questions with the negation in a high position (HiNQ, e.g. *Aren't they adding a menu item?*) both to check their own belief that the proposition *p* is true (e.g., *p* = they are adding a menu item) and their addressee's belief that it is not (Ladd, 1981). In our data, speakers of Western Canadian English most frequently chose this question type over the alternative with the negation in a lower position (LowNQ, e.g. *Are they not adding a menu item?*) in both cases, supporting hypothesis 1a. Note that while HiNQs constituted the majority of responses (62% overall), LowNQs still appeared frequently (38% overall). This may to some degree be related to the fact that participants made a binary choice between these two options. However, two facts speak against this interpretation: First, according to Romero and Han (2004), while HiNQs obligatorily convey a question bias, LowNQs may be used as neutral questions, but can also be used in a biased context. Second, Domaneschi et al. (2017), who gave their participants five different response options, still found 26% LowNQ choices and observed a similar percentage of HiNQ responses (67%) in the same context, i.e. where the speaker had a previous bias towards *p* and the addressee expressed a bias for ¬*p*. Note also that all the other response options in Domaneschi et al. (2017), which included the possibility for speakers to freely formulate their own question response, together constituted less than 10% of responses in this context, suggesting that providing a binary choice between HiNQ and LowNQ in the present study was appropriate. A search of the 560-million-words Corpus of Contemporary American English (COCA; Davies, 2008, 2009) for text strings indicating the use of HiNQs and LowNQs suggests that in spontaneous

conversations and naturally-occurring written texts, HiNQs are relatively more frequent than in these experimental contexts, which were specifically constructed to elicit biased questions, but LowNQs also constitute between 8% and 43% of negative polar questions, depending on the subject pronoun (Table 2; note that sentences with full noun phrases, demonstratives etc. as subjects were not counted; also, the search was restricted to strings including present tense forms of the auxiliary/verb *to be*).

Table 2. String frequencies for LowNQs and HiNQs similar to those occurring in the present studies with various subject pronouns in COCA (Davies, 2008, 2009), with searched strings given in brackets, and relative percentage of HiNQs, calculated as number of HiNQs * 100 / (number of HiNQs + number of LowNQs), rounded.

Subject pronoun	Number of LowNQs	Number of HiNQs	% HiNQs
I	323 (am I not)	352 (aren't I) + 69 (ain't I)	57
You	447 (are you not)	5103 (aren't you)	92
He	204 (is he not)	1208 (isn't he)	86
She	77 (is she not)	658 (isn't she)	90
It	1372 (is it not)	10974 (isn't it)	89
We	311 (are we not)	1158 (aren't we)	79
They	300 (are they not)	1445 (aren't they)	83

SUM	3034	20967	87

Supporting hypothesis 1b, we further found that speakers of English prosodically disambiguated the checked proposition, expanding a suggestion by Domaneschi et al. (2017). Participants ended their questions with higher f0 rises more often when checking the addressee's proposition than when checking their own. In addition, an effect of proposition checked also appeared for the accentuation of English auxiliaries (namely, more L+H*> accents when checking ¬*p* than when *p*), although it was small (see below for more detailed information on the prosodic marking). The importance of the difference in accentuation should therefore not be overstated, but see below for some discussion of how this finding could be related to accounts of intonational meaning.

In line with hypothesis 1c, we found that German differs from English with respect to the morpho-syntactic marking of the checked proposition: German speakers chose a majority of HiNQs when checking their own proposition *p*, but uttered LowNQs more frequently to check their addressee's proposition that ¬*p*, in line with the previous literature (Büring & Gunlogson, 2000). The prosodic realization, however, did not directly encode proposition checked in German. Instead, it was best explained by the morpho-syntactic question form: Unaccented negations constituted the largest group in HiNQs, whereas negations in LowNQs were most frequently realized with L*+H accents. L*+H accents on the sentence-final compound nouns were more frequent in HiNQs than in LowNQs. Additionally, HiNQs were more often realized with a high final f0 rise (boundary tone H-^H%) than LowNQs. Thus, while Figure 10 above shows the typical intonational realization of a HiNQ, Figure 11 illustrates the accents and boundary tone found most

frequently for German LowNQs. Note, however, that German LowNQs were intonationally more variable than their HiNQ counterparts: While there were only three LowNQs that had the intonational pattern shown in Figure 11 (about 1% of German LowNQs), Figure 10 is representative of 99 realizations that had the most frequent intonational realizations for all constituents (25% of German HiNQs). Still, the picture that emerges is that for the German participants, their choice of morpho-syntactic question type largely determined, or at least restricted, options for the prosodic realization. This contrasts with English, where accentuation was also affected by the choice of question type, but these effects were either very small or can simply be explained through the availability of syllable nuclei for accentuation, as discussed below.

[insert Figure 11]

Summing up the evaluation of hypothesis 1, our results suggest that in German, the distinction between checking the speaker's own epistemic bias for the proposition *p* and checking the addressee's bias for ¬*p* is encoded syntactically via the choice of HiNQ vs. LowNQ and this morpho-syntactic choice then largely determines the prosodic realization. In English, by contrast, the same morpho-syntactic form (HiNQ) is uniformly preferred, and the checked proposition is directly prosodically encoded via accentuation of the auxiliary and the final boundary tone (rising accent on auxiliary and high boundary tone when checking the addressee's proposition, unaccented auxiliary and less high boundary tone when checking the speaker's own). Thus, both languages reflect the distinction between double-checking the speaker's own bias and double-checking the

addressee's bias. Prosody is influenced directly by proposition checked in English, and follows from the choice of morpho-syntactic question type in German.

In hypothesis 2, we hypothesized that certainty would play a role. In our data, there was no main effect of certainty in either language. However, certainty played a role in that it interacted with the marking of proposition checked in some cases: In particular, it affected only the choice of morpho-syntactic form in German and only the prosodic form in English. Specifically, regarding morpho-syntactic marking in German, speakers' choice of HiNQ vs. LowNQ further interacted with their level of certainty regarding the proposition they checked with their question: When participants were 90% certain (that their own belief or that of their addressee was true), they chose their preferred question type – HiNQ for checking *p*, LowNQ for checking ¬*p* – more often than when they were only 60% certain.

In English, where checked proposition did not affect the choice of question type, certainty did not have an effect on choice of question type, either. Instead, the interaction between proposition and certainty appeared in the *prosodic* realization of the questions in our English data: The effect of checked proposition on use of utterance-final high-rising boundary tones got boosted when participants were 90% certain compared to when they were 60% certain. However, statistical modelling did not confirm a significant interaction with certainty for the effect of proposition on the accentuation of auxiliaries, though Figure 8 shows that for HiNQs, the difference between the two propositions was numerically larger for 90% certainty than for 60% certainty.

Thus, for both languages, certainty was not itself encoded prosodically or in the choice of morpho-syntactic question type, but it strengthened the preference for encoding the

checked proposition. Thereby, our results contrast with the observed prosodic marking of speaker certainty, confidence or commitment reported in many Romance languages, e.g. for Bari Italian (Grice & Savino, 1995, 2003), Majorcan Catalan (Vanrell et al., 2013), French (Michelas et al., 2016), Puerto Rican Spanish (Armstrong, 2017), Central Catalan (Prieto & Borràs-Comes, 2018). Whether or not these results indicate a contrast between the West-Germanic languages English and German on the one hand and the Romance languages on the other hand is presently unclear due to differences in study design. Whereas we treated checked proposition and certainty as two separate factors crossed in our study design, previous studies often treated speaker certainty as situated on a gradient (explicitly, e.g. Armstrong, 2017), with the checking of the speaker's or addressee's proposition following from the speaker's certainty of their own proposition.[7,8] In this context, the results of the third study reported in Romero, Arnhold, Braun, and Domaneschi (2017) might be illuminating: In a modified version of the present study investigating only the choice of morpho-syntactic question type, the speaker's and addressee's original beliefs were flipped such that the speaker originally held a bias for $\neg p$, whereas the addressee's bias was for p. As in the present study, participants' choice of question type was significantly influenced by the checked proposition, however, there was no effect of certainty and no interaction. These results do not support an epistemic gradient ranging from the speaker being totally sure of their own proposition to the speaker being totally convinced of their addressee's proposition. Instead, they seem to suggest that the primary factor, at least for English and German, is the checked proposition – as well as the presence of biases for specific propositions in the first place (e.g. Domaneschi et al., 2017). Thus, the combined results of the present study, the third

study in Romero et al. (2017) and Domaneschi et al. (2017) suggest that, in line with the previous literature, the prosody of polar questions is influenced by (i) syntactic form or question type (e.g. HiNQ vs. positive polar question), (ii) pragmatic function (neutral information-seeking use vs. biased questions seeking reassurance/confirmation or expecting a given answer) and (iii) the propositional content – *p* vs. ¬*p* – checked by the question and/or whether the propositional content is 'owned' by the speaker or the addressee. Various combinations of these factors are signalled with different morpho-syntactic and/or prosodic polar question forms. By contrast, at least for the West-Germanic languages investigated here, speaker certainty or commitment does not seem to be marked per se, but can only increase the marking of one of the other factors.Should future studies confirm the apparent contrast between West-Germanic and Romance languages, a possible explanation could follow from the fact that Romance languages tend to be pro-drop languages: While interrogativity can be marked by subject inversion/dislocation, this is not possible in sentences without overt expression of subjects. As a result, sentences without overt subjects, which are frequent, must be marked as statements or questions by prosody (Escandell-Vidal, 1998, 1990; Grice & Savino, 2003; Vanrell et al. 2013), and the same can be argued to apply to question bias and certainty. This could be the reason for intonation playing a more important role in this language family.[98] However, the opposite could also be speculated to be true, so that in a language where intonation is already used to mark several distinctions, prosody may not be available to mark a further distinction. In this respect, it might be noteworthy that interactions with certainty in our data were clear for German, which allows morpho-syntactic marking of the checked proposition, and somewhat less pervasive for English,

which relies exclusively on prosody for this purpose. A third alternative would be that Romance and Germanic languages simply differ in the semantic/pragmatic distinctions they 'choose' to encode grammatically, in the same way that some languages morphologically mark aspect or evidentiality and others do not.

4.2 Prosody of negative polar question in English and German

Before we address the wider questions of the semantics of intonation and its interaction with syntax, a comment is due regarding the exact prosodic form of the English polar questions in our data. Recall that hypothesis 1b remained vague with respect to how proposition checked would be marked prosodically. The prosodic realizations in our data only partially corresponded to descriptions in the literature. Most importantly, while the discussion of utterance-final pitch movements in the literature has centered on rises vs. falls (see introduction), final pitch falls were virtually absent from our data. Instead, English negative polar questions in our data overwhelmingly ended in pitch rises, with 85% of utterances having final high H-H% boundary tones and 89% of preceding final accents being low L*. A likely explanation for the differences in final pitch movement are regional differences, in particular the fact that Canadian English is a so-called 'uptalk' variety in which many statement utterances end in a high boundary tone (James, Mahut, & Latkiewicz, 1989; Shokeir, 2008; Talla Sando, 2009): It is possible that the final rises in Domaneschi et al.'s (2016) British English data (and potentially those in the American English data evaluated by Hedberg & Sosa, 2002, and Nilsenová, 2006) correspond to high rises in our Western Canadian data, while the absence of a high rise in Western Canadian English may be equivalent to a fall when marking the checked proposition in a

negative polar question in British (and American) English. This would mean that Canadian English preserves the distinction between final rises and falls reported for question intonation in other varieties as a distinction in final pitch height: In the relative absence of pitch falls due to uptalk, even utterances with final rises may be perceived as prototypical 'statement intonation' (in its 'uptalked' variant) as long as the rise is not too high, while only high final rises would thus constitute prototypical 'question intonation' appropriate for checking the addressee's proposition (future perception experiments to determine the boundary between the two categories are desirable). Thus, we predict that our findings can be extended to British and non-uptalk varieties of American English: We hypothesize that in these varieties, HiNQs checking the speaker's proposition *p* are realized with final f0 falls and HiNQs checking the addressee's proposition ¬*p* are realized with a final f0 rises.

Alternatively, it is possible that our observation of a prosodic disambiguation of Ladd's ambiguity is a peculiarity of Western Canada that does not extend to other varieties of English. This does, however, seem less likely both in view of the extended discussion regarding the potential functions of final rises and falls in British and American English summarized in the introduction, and based on Ladd's (1981) original article, in which he discusses the role of prosody in disambiguating the scope of negation and interrogation in tag questions. Furthermore, with the exception of the final boundary tone, the prosodic form of HiNQs in our data corresponded quite closely to previous descriptions of American English. As described by Hedberg and Sosa (2002) and Romero and Han (2001, 2002), we mostly found a rising accent on the auxiliary (L+H*< in our data, L+H* in theirs). As described by Romero and Han (2001, 2002) this was most frequently

followed by a final low L* combined with an H-H% high boundary tone, forming what is generally described as the unmarked question contour (but note the variety of nuclear contours reported by Hedberg & Sosa, 2002).

In a similar vein, it is interesting to note that even though both final pitch rises and final falls have been described as frequent realizations of German polar questions, with various interpretations suggested for the difference (cf. introduction), final falls were in fact rare in our German data, as well. Altogether 74% of the utterances carried high H-^H% or H-% boundary tones and the most frequent accent category for the sentence-final compound noun was the L*+H rise, followed in frequency by the low L* accent. The final fall-rise contour described for biased polar questions by Pheby (1975) was largely absent: This fall-rise contour would either correspond to a GToBI annotation with a combination of a final high H* or rising L+H* accent and a rising boundary tone L-H% or to a combination of a final falling H+L* or H+!H* accent and a high H-^H% or H-% boundary tone. However, L-H% boundary tones occurred in only 11% of the utterances, while final compound nouns were never realized with H+L* accents and only very rarely with H+!H* accents (less than 1% of the data).[109] In fact, the prevalence of final rises in our data corresponds very well with their frequency in Kohler's (2005b) corpus of spontaneous spoken appointment-making dialogues in German: Of 121 polar questions, he found 39% to have a final high rising contour, 30% to have a final low rising contour, and only 21% to end in falls and 10% in other contours. Thus, it seems likely that the intonation contours in our data set are representative of German prosody more generally, so that the present findings shed light on the previously under-described prosodic differences between German HiNQs and LowNQs.

4.3 Intonational meaning

Relating our findings to proposals of intonational meaning, we are first going to discuss results regarding utterance-final pitch movements in English and German before we turn to the realization of accentuation in both languages.

The utterance-final pitch movements in our English data can be quite straightforwardly related to the literature if we are correct that, as suggested above, only the high final rises in our Western Canadian data correspond to rises in other English varieties, whereas the lower final rises are equivalent to falls in other varieties discussed in the literature. If this equivalence is correct, the high rises our participants used preferably to check the addressee's proposition $\neg p$ correspond to H-H% in other varieties, a boundary tone combination that according to Steedman (2008) directly attributes the responsibility for the propositional content to the addressee. By contrast, the less high or low boundary tones that participants used most often to check their own proposition p (annotated as lower H-H% or the less frequent H-L% or L-L%) would correspond to L-L% in other varieties and thus attribute the responsibility for the propositional content to the speaker according to Steedman (we disregard L-H% boundary tones here and below, as they constituted only 2% of our data overall). Thus, our findings show very good agreement with his account (and related accounts by Gunlogson, 2001, and Escandell-Vidal, 1998, 1999, 2017), as the contrast between double-checking the speaker's vs. the addressee's proposition dovetails with attributing responsibility for the propositional content to the speaker vs. the addressee.

The correspondence with Pierrehumbert and Hirschberg's (1990) account is slightly less clear. According to them, as well as Bartels (1997b, 2013), H-H% points towards the

relevance of an upcoming utterance for interpreting the current utterance (in the case of questions like in our data, the upcoming utterance is the expected answer of the addressee). By contrast, L-L% (which would correspond to the lower final rises in our Western Canadian data) does not signal a dependence on an upcoming utterance according to their proposal. At first blush, it is not easy to see why negative polar questions checking the speaker's proposition *p* and the addressee's proposition ¬*p* should differ in this respect: Both seem to convey a similar 'forward reference', to use Pierrehumbert and Hirschberg's term, to the upcoming expected response of the addressee.[1~~0~~]

However, an alternative interpretation could be constructed based on Bartels' (1997b, 2013) observation that polar questions with final falling intonation are similar to alternative questions in signaling that the speaker is only interested in a *yes* or *no* answer, whereas other answers might be more acceptable when polar questions bear final rising intonation. For example, *Do you like it?* with a fall is similar to *Do you like it or not?*, whereas a realization with a rise seems to be more encouraging of the addressee presenting other information. In the semantic-pragmatic literature, Bartels' description has recently been further developed by Westera (2017, also cf. Bäuerle, 1979, Biezma & Rawlins, 2012). He proposes that polar questions with final falling pitch signal 'thematic exhaustivity' whereas the same questions with rising pitch signal that something else, another answer besides *yes* or *no*, may be relevant. When combined with Roberts' (2012) discourse structure framework using questions under discussion (QUDs), this means that the function of rising intonation is to signal that the current QUD has a sibling in the QUD-tree. For example, asking *Are you American?* with a rise signals that other parallel

questions like *Are you Canadian?* are sub-parts of the same mother-QUD (e.g. *What is your nationality?*), meaning that the addressee is invited to already answer such questions and that the speaker may in fact ask them next if receiving only a *yes* or *no* answer. Such a question could for example be asked to start a conversation at a party. By contrast, a US immigration officer would typically ask the same question with falling pitch, which signals that only a *yes* or *no* answer is to be given and announces that other nationalities are irrelevant (i.e. sibling QUDs like *Are you Canadian?* are not relevant). Bringing this back to our results, one could suggest that our English-speaking participants preferably used final high rises when checking the addressees' proposition $\neg p$ to signal that they were not only interested in double-checking the proposition itself, but also in further information, for example exactly how much credence the addressees put in their proposition $\neg p$. By contrast, when participants double-checked their own proposition p, they may have used the lower final rises (which we assume correspond to falls in other varieties of English) to signal that they were indeed only interested in the truth of the proposition itself, and ideally in the addressee confirming that the speaker was right all along.

Turning now to sentence-final pitch movements in the German data, a match with accounts of intonational meaning is not completely straightforward. As summarized in the introduction, various, partially overlapping interpretations have been proposed for the contrast between rises and falls (as well as specific meanings for fall-rises and rise-falls). However, there is little explicit discussion of the contrast between high rises and less-high rising final plateaus, annotated in GToBI as H-^H% vs. H-% boundary tones combined with a L*(+H) nuclear accent, which was the most salient contrast in our data.

However, Grice and Baumann's (2002) and Grice et al.'s (2005) tables illustrating common nuclear contours for German show the combination L* H-^H%, which likely corresponds to the final high rises annotated as L*+H H-^H% in our data, as typical for neutral polar questions (or echo questions), see also Braun, Dehé, Neitsch, Wochner and Zahner (2018). While they caution against interpreting their examples as directly illustrating intonational meaning and do not show examples of questions with a L*(+H) H-% contour, it could maybe be concluded that polar questions with less high rising plateaus, simply by virtue of not being neutral high rises, would convey some kind of bias (in Braun et al. 2018, these contours were used to convey a rhetorical question meaning).

In our German data, the distinction between H-^H% vs. H-% boundary tones when following L*+H nuclear accents correlated with the proposition checked by the question. Although the primary marking of this contrast was morpho-syntactic, HiNQs (typically checking the speaker's proposition *p*) were more frequently realized with H-^H% than LowNQs (checking ¬*p*), so that each checked proposition was indirectly associated with a prosodic form. We are unable to say with certainty whether proposition checked was a factor that independently affected morphosyntactic question form and prosody or whether proposition checked affected morphosyntactic form, which in turn affected prosodic realization. However, since morphosyntax and prosody are not independent of each other, it is almost impossible to discuss intonational meaning for our negative question forms . What is interesting, though, is that in contrast to the English data, higher final rises were more strongly associated with checking the speaker's proposition *p* than with checking addressee's proposition ¬*p*. The reason for this difference could be that in contrast to

English, proposition checked is already marked by the morpho-syntactic question type in German. Therefore, prosody can be used for other purposes in German. When checking their own proposition *p* with a HiNQ, speakers mostly used a clear final high rise (L*(+H) H-^H%) signaling neutrality and thus openness towards the addressee's position (strongly *fragehaltig* in Batliner's, 1989, terms). By contrast, when checking the addressee's proposition ¬*p* with a LowNQ, speakers mostly used the less high rising H-% boundary tones to signal lack of neutrality and that they were ready to accommodate the addressee's position. This could be related to previous research describing question prosody in terms of politeness as summarized in the introduction.

Next, our results regarding accent realizations can be related to compositional accounts of intonational meaning. Starting again with the English data, the only constituent that showed a significant effect of one of our manipulated variables, proposition checked, was the auxiliary. Here, the most frequent realization was an L+H*> accent, which contrasted with the next-most frequent category, unaccented realizations. In addition to the small, but significant effect of proposition checked (more L+H*> when checking the addressee's proposition ¬*p* than when checking the speaker's proposition *p*), we observed larger effect of question type (more L+H*> in HiNQs than in LowNQs). Before discussing these effects, some comments are due on the meaning conveyed by the choice of L+H*(>) as the most frequent accent type.

Pierrehumbert and Hirschberg (1990) state that rising accents (L+H* and L*+H) convey the salience of a scale, with L+H* marking that the accented item is to be believed instead of believing a related item on a scale (i.e. a related item in a set of alternatives in

the sense of Rooth, 1985, 1992). This explains why these accents are typically used in corrections or to mark contrasts, as in their example of this naturally occurring exchange:

(6) A: It's awfully warm for January.

B: It's even warm for December.

L+H* L H%

(from Pierrehumbert & Hirschberg, 1990, p. 296)

By realizing the word December with a L+H* rising accent, B invokes the scale contrasting months of the year and signals that she is making a predication (i.e. not January, but December). Since in our study, the truth of the proposition as a whole was under discussion, it makes sense that the L+H*< accent frequently appeared on the auxiliary and negation complex. Prominent accents on the finite verb are typically taken to mark either polarity focus (Goodhue, 2018; Rooth, 1992) or verum focus (see Höhle, 1992; Turco, Dimroth, & Braun, 2013). Following Pierrehumbert and Hirschberg's account, the use of L+H*(<) on auxiliaries in our data could thus be described as the speaker making salient the scale {*p*, ¬*p*} (for polarity focus) or the scale {'you are totally sure that ¬*p*', 'you are quite sure that ¬*p*', …} (for verum focus).

Taking this meaning of the L+H*< accent as a starting point, two explanations can be given for the effect of question type (L+H*< accents on the auxiliary were more frequent in HiNQs than in LowNQs). First, in both HiNQ and LowNQ, this accent marks focus on the polarity, namely on the negative polarity, of the sentence. Thus, both the auxiliary complex (i.e. Aux + *n't*) and the negation could be accented to mark this polarity focus. However, segmental phonology favours the auxiliary over the negation as an accent location for HiNQs, but not for LowNQs: As already mentioned above, the negation is

reduced to a clitic in English HiNQs, whereas it is a full prosodic word in LowNQs (*Aren't they…?* vs. *Are they not…?*). Consequently, negations were never accented, but their host auxiliaries regularly carried accents in HiNQs, whereas in LowNQs, negations almost always carried an accent, but the preceding auxiliaries did not.

Second, Romero and Han (2002) suggest that preposing the auxiliary introduces a VERUM operator, i.e. that English HiNQs always express a verum focus, whereas English LowNQs do not (which also explains why they are overall preferred in situation like those in our study, where speakers' and addressees' biases conflict). According to this interpretation, the observed difference in accent realization between HiNQs and LowNQs would actually be meaningful in that the L+H*< accents preferred on auxiliaries would mark verum focus in HiNQs, whereas LowNQs rarely showed rising accents on auxiliaries and thus would not mark verum focus. In line with this idea, negations in LowNQs, which were preferably accented instead of auxiliaries, carried various types of accents, including some realizations as L+H*, but the most frequent category in this context was the non-contrastive L* accent.

Regarding the effect of proposition checked, the fact that L+H*> accents were slightly more frequent when checking the addressee's proposition ¬*p* than when checking their own proposition *p* could receive various explanations, but at this point, we do not see one that is clearly superior.[1] However, it is important to remember that the effect was small and did not interact with question type. Thus, participants overwhelmingly responded by using HiNQs, in which they mostly realized L+H*> accents on the auxiliaries.

Turning to the accentuation of other constituents, our data fit well with Pierrehumbert and Hirschberg's (1990) suggestion for the compositional meaning of accents: Simple L*

accents were most frequently used on verbs and sentence-final compound nouns in our data. L* accents, according to Pierrehumbert and Hirschberg, serve to highlight a constituent, but explicitly exclude it from the speaker's predication, which is typical of canonical questions, where the speaker does not make a statement but asks the addressee to make one instead.

In contrast to Pierrehumbert and Hirschberg's account, the accentuation of HiNQs in our English data may be harder to reconcile with Steedman's (2000, 2008) account. He suggests that L+H* accents, which were the most frequent realization of auxiliaries, mark themes that are mutually believed by the speaker and the hearer, i.e. uncontentious, whereas L*+H accents appear on themes that are not marked as agreed between the speaker and the hearer. As there was clearly disagreement in the dialogues in our study, the occurrence of L+H*(<) accents would be unexpected, apart from the fact that it is questionable whether auxiliaries can be themes. His statement that L* accents mark rhemes that are not marked as agreed between speaker and hearer could fit with their occurrence as the most frequent realization of compound nouns, if the accent on this sentence-final constituent is assumed to project a larger rhematic constituent.

Finally, regarding accentuation in the German data, it seems fitting that our German participants mostly used LowNQs to check the addressee's proposition and in turn mostly realized the negation in these LowNQs with the rising accents L+H* or L*+H, accents that according to Pierrehumbert and Hirschberg (1990) invoke a scale, and frequently a contrast. As the negation is the part of the utterance that was in question in these short dialogues, it makes sense that speakers chose to mark it as contrastive. While both types of rising accents can be used to accomplish this, German speakers employed the L+H*

accent slightly more frequently than the L*+H accent (note that we did not statistically evaluate the relative frequency of these two accent types since the main contrast in the realization of German negations was between accented and unaccented). Thus, they used the same accent type that English speakers used primarily on auxiliaries in the same condition, i.e. when checking the addressee's proposition ¬*p*. Interestingly, in contrast to the English data, speakers also used a sizable amount of L*+H accents, which according to Pierrehumbert and Hirschberg in contrast to L+H* "convey a lack of predication" (1990, p. 296). While Pierrehumbert and Hirschberg's examples for both accent types come only from statements, they discuss the contrast between H* vs. L* accents in polar questions, which likewise contrast with respect to the speaker making a predication in statements: The H* accent typically involves predication – the speaker marks the constituent realized with H* as new and signals that the proposition resulting from inserting this constituent is to be added as a mutual belief–, while the L* exempts the constituent bearing it from the predication. Here, they state that in morpho-syntactic polar questions, is less striking than for statements. However, they remark that H* accents frequently appear in confirmation questions, i.e. when the speaker believes the answer will affirm the semantic nucleus of the question (in their example, the expectation is of a *yes* answer to a positive polar question). Both marking and not marking such an expectation seem appropriate in the contexts we provided to our participants. Importantly, both rising accents highlight the polarity contrast (*p* vs. ¬*p*) as expected in the context of our study.

To sum up this discussion, we found several matches between our results and compositional accounts of meaning suggested in the literature – taking into account the

morpho-syntactic marking of our data as questions where appropriate. In particular, the effect of proposition checked on the use of high-rising utterance-final pitch in our English data fits well with the meaning suggested for H-H% boundary tones in the literature (attribution of responsibility according to Steedman, 2008, or marking an interest in further information according to Bartels, 1997b, 2013, and others). In contrast, further research is required on the meaning of the contrast between high and less high utterance-final rises in German (H-^H% vs. H-% in GToBI). Regarding the meaning of accents, we again found good agreement between the frequent use of contrast-marking rising accents on auxiliaries or negations in both our English and German and the literature (evocation of a scale according to Pierrehumbert & Hirschberg, 1990).

4.4 Conclusion

To conclude, our study of English and German negative polar questions confirmed experimentally that English HiNQs can be used to check both the speaker's bias for the proposition *p* and the addressee's proposition ¬*p*, whereas German marks this distinction morpho-syntactically (Ladd, 1981; Büring & Gunlogson, 2000). Moreover, we found that speakers of English frequently resolved this ambiguity within HiNQs prosodically by marking the checked proposition through sentence-final pitch movements and accentuation of the auxiliary (high rise boundary tone and rising accent on the auxiliary for checking the addressee's proposition, normal-high or low boundary tones and unaccented auxiliary for checking the speaker's proposition). By contrast, the prosody of the German negative polar questions was largely dependent on the morpho-syntactic choice, with HiNQs including the negation *nicht* 'not' generally carrying different

intonation contours than LowNQs containing the negation *kein* 'no'. Speaker certainty was not directly encoded, but affected the marking of the checked proposition, with higher certainty making the marking more consistent where it occurred – prosodically in English, morpho-syntactically in German.

Acknowledgements

We would like to thank Clara Huttenlauch, Katharina Zahner and Devon Gozjolko for help with data annotation, Clara Huttenlauch for help with data collection, Keely Morrow for checking the English items and Oleksiy Bobrov and Achim Kleinmann for technical support. We further thank Juhani Järvikivi for hosting our data collection at the Centre for Comparative Psycholinguistics at the University of Alberta, as well as lab members for practical support.

Funding statement

This work was supported by the German Research Foundation (DFG) [Priority Programme 1727 "XPrag.de: New Pragmatic Theories based on Experimental Evidence", project "BiasQ: Bias in Polar Questions"].

Notes

[1] As already pointed out by Ladd, the ambiguity between English HiNQs checking *p* vs. ¬*p* is resolved when polarity items are present: Positive polarity items like *some* and *too* indicate that the speaker is checking *p*, whereas negative polarity items like *any* and

either indicate that they are checking ¬*p* (e.g. *Isn't there some/any vegetarian restaurant around here?*). However, since disambiguating polarity items are frequently absent, HiNQs often remain string-wise ambiguous. Additionally, Romero and Han (2004) note that LowNQs can occasionally express original speaker bias and show the checking *p*/¬*p* ambiguity.

[2] Note that while we have translated the negative question with an English LowNQ, the position of the negation cannot be determined in the German original since the object is a pronoun.

[3] Pheby elsewhere describes a similar effect of pitch range also for the first case, i.e. the final rise normally marking neutral questions. He states that when the final rise has a large pitch range it can carry a 'negative' expectation, expressing speaker surprise and potentially disbelief regarding the checked proposition (Pheby, 1975, p. 60; examples are positive polar questions; also note that we have translated his use of the German *Implikation* with 'expectation' here, since it is not clear whether he means 'implicature' in the semantic sense).

[4] As pointed out by a reviewer, two context descriptions in Grice and Baumann's (2002) table of common nuclear contours, which is repeated in Grice et al. (2005), seem related to bias: contour 2, the rise-fall (late peak) L*+H L-% described as "selbstverständliche Feststellung" (Grice and Baumann, 2002, p. 19) / "self-evident assertion" (Grice et al., 2005, p. 71) and contour 6a, the early peak H+!H* L-% described as "Bestätigung einer bekannten Tatsache" (Grice and Baumann, 2002, p. 20) / "established fact" (Grice et al., 2005, p. 72), although Grice and colleagues state that "[t]he contexts provided in the table contain pragmatic interpretations referring to specific examples; they should not be taken

as abstract meanings for given contours" (Grice et al., 2005, p. 70). The example utterances for both contours, which are taken from previous literature, are statements in both cases. As these contours seem to have been described almost exclusively for statements, we do not provide an exhaustive review here, but for the former contour see Pheby (1975, p. 147; 1984, p. 847, 881-882), who describes the rise-fall as infrequent, but familiar and more obliging than a simple fall, and Féry (1993:93-96), who transcribes it as L*HL and states that it has a meaning like 'of course'; on the latter contour, the early peak, see Féry (1993:103-106), who transcribes it as H H*L and notes that it is frequently used by television reporters, as well as several publications of perception studies by Kohler and colleagues (e.g. Kohler, 1991, 2005a; Niebuhr, 2003). The only mention regarding questions that we could find comes from Uhmann (1991:171-174), who transcribes the rise-fall as T*+H T% (note that she consistently uses T instead of L to mark low targets) and remarks that it was rare in her data, appearing exclusively on a few precursor questions (*Vorläuferfragen*), although she remarks that a use on certain declaratives is maybe possible. Neither of these two nuclear contours appeared in our data.

[5] These overall numbers are based on evaluating boundary tones and pitch accent types on all constituents together, with the absence of a pitch accent treated as a pitch accent type ("none"). Agreement regarding the presence vs. absence of a pitch accent was 92% ($\kappa = 0.84$) for English and 91% ($\kappa = 0.81$) for German, while agreement regarding pitch accent type for constituents where both annotators had marked the presence of a pitch accent was 72% ($\kappa = 0.53$) for English and 70% ($\kappa = 0.58$) for German. Agreement

regarding boundary tones was 58% (κ = 0.37) for English and 82% (κ = 0.70) for German.

[65] To test the validity of this distinction, we measured f0 height at the end of the utterance (in semitones relative to 100Hz, measured 50ms from the end of the utterance to avoid microprosodic effects; note that f0 was not measurable for 15 utterances, i.e. 2% of the data). We then fitted a linear mixed-effects model to this data, using participant gender and the binary classification into high rises and other final boundary tones as a predictor. The resulting model (trimmed, removing 37 data points or 4%) contained a fixed effect of the binary classification into final high rises vs. other boundary tones showing that the former were associated with significantly higher f0 measurements, β = 3.672, *SE* = 0.467, *t* = 7.87, *p* < 0.001 (mean: 16.2st for high final rises vs. 11.6st for others). The best model additionally contained a by-participant random slope of boundary tone classification. The fixed effect of gender was also significant, indicating that compared to the female participants, male participants produced overall significantly lower f0 values, β = -12.114, *SE* = 1.205, *t* = -10.05, *p* < 0.001, as did the participant who identified their gender as 'other', β = -8.530, *SE* = 2.942, *t* = -2.90, *p* = 0.008 (excluding data from this participant produced the same results). Finally, restricting the analysis to only utterances annotated with H-H% boundary tones resulted in a nearly identical best model, indicating that those H-H% tones classified as high final rises indeed had significantly higher f0 than those that were not, β = 3.118, *SE* = 0.462, *t* = 6.75, *p* < 0.001 (mean: 16.2st for high rises vs. 11.7st for non-high rises; trimming removed 33 data points, i.e. 5% of the 648 utterances with H-H% tones). This model also again contained a by-participant random slope of the classification of high rises, as well as a significant effect of gender,

indicating overall lower pitch for male participants, β = -12.306, *SE* = 1.241, *t* = -9.92, *p* < 0.001, and the participant identifying with the third gender category 'other', β = -8.842, *SE* = 2.970, *t* = -2.98, *p* = 0.007, compared to female speakers.

Note also the difference between this classification, which is parallel to our classification of the German data below, and another use of the term 'high-rise', which contrasts rises from a final high accent (H* H-H%) with those from a final low accent (L* H-H%), i.e. is based on accent category, not the height of the f0 realizing the final boundary tone (e.g. Bartels, 1997b, 2013; Hirschberg & Ward, 1995; Levis, 1999; Pierrehumbert & Hirschberg, 1990).

[67] We again validated this choice by performing f0 measurements and linear mixed-effects modelling as for the English data (final f0 not measurable for 16 utterances, i.e. 2% of the data; best model trimmed, removing 33 data points or 4%). Modelling indicated that utterances classified as having a final high rise ended in significantly higher f0, β = 2.490, *SE* = 0.576, *t* = 4.32, *p* < 0.001 (mean: 17.7st for high rises vs. 14.5st for others). Also, the model contained a by-participant effect suggesting that the absolute difference differed between speakers and a significant main effect of gender, β = -11.310, *SE* = 0.8317, *t* = -13.60, *p* < 0.001. Comparing only H-% and H-^H% yielded a model with the same effect, importantly indicating that final f0 was higher for H-^H%, β = 1.518, *SE* = 0.438, *t* = 3.46, *p* = 0.005 (37 data points or 5% of 569 trimmed; mean: 17.7st for H-^H% vs. 15.4st for H-%).

[87] A slightly different approach is taken by Prieto and Borràs-Comes (2018), who treat speaker commitment (to their own proposition) and speaker agreement (with their addressee's proposition) as two separate factors, which were however not crossed in their

study design. The stimuli of their perception study consisted of twelve different contexts each followed by a matching question (with all questions presented with four different intonational contours, i.e. each participant rated 12 x 4 = 56 stimuli). The twelve contexts represented six different conditions (i.e. two items per condition), three of which manipulated only speaker commitment (low, mid, high), while the other three manipulated speaker agreement (low, mid, high). Importantly, the condition with high speaker agreement at the same time had low speaker commitment, while both the mid and low speaker agreement conditions had mid speaker commitment.

[9] We thank an anonymous reviewer for suggesting this possible explanation.

[10] However, it is a bit difficult to match up Pheby's descriptions with GToBI annotations. Importantly, all of his fall-rise contours stretch across several words, with a high or falling nuclear accent ('Tonsilbe') followed by a tail ('Nachlauf') consisting of a low plateau and a final rise (see Pheby, 1975, p. 54-55), which would correspond to the GToBI annotations given here. However, if his accents on the 'Tonsilbe' were interpreted as pre-nuclear instead of nuclear, the fall-rise could correspond to a GToBI annotation of H* L*(+H) H-(^H)%, i.e. a pre-nuclear high accent, followed by a nuclear low or rising accent and a high boundary tone, i.e. the most common realization of HiNQs in our data as illustrated in Figure 10.

[11] Note that we slightly simplify in the main text here and above by referring to the meaning of the boundary tone combinations H-H% vs. L-L% as a whole, since the contrast between these two combinations is most relevant to our data and most discussed in the literature. Pierrehumbert and Hirschberg (1990) attribute very similar meanings to H- and H% (forming part of a larger utterance/dependence on upcoming speech) and to

L- and L% (emphasizing the separation of the current unit from a subsequent phrase/absence of dependence on upcoming speech). According to Bartels' interpretation, "to P[ierrehumbert] & H[irschberg], the difference in function between phrase accents and boundary tones lies merely in the size of the prosodic unit for which they cue discourse integration–intermediate vs. intonational phrase" (Bartels 2013, p. 59). Bartels (1997b, 2013) assumes that L% and H- are inherently meaningless, so that the meaning of L-L% is entirely determined by the meaning of L- (assertiveness), whereas the meaning of H-H% corresponds to the meaning of H% (dependence on upcoming units of speech). In contrast to H-L% and L-H%, the simplification regarding the compositionality of meaning is thus justified for the H-H% and L-L% cases we discuss here.

[11~~2~~] For example, in line with the high rise of HiNQs double-checking $\neg p$, evocation of a scale – and thus the use of the accent L+H*> – is more important/natural if the speaker cares about other alternatives on the scale beyond the current one. Alternatively, highlighting the polarity contrast is additionally important when the proposition checked conflicts with the speaker's beliefs.

References

Altmann, H. (1984). Linguistische Aspekte der Intonation am Beispiel Satzmodus (*Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation der Universität München (FIPKM)* No. 19). Institut für Phonetik und sprachliche Kommunikation der Universität München: München.

AnderBois, S. (2011). *Issues and Alternatives* (PhD thesis). UC Santa Cruz.

Asher, N., & Reese, B. (2007). Intonation and Discourse: Biased Questions. (S. Ishihara, S. Jannedy, & A. Schwarz, Eds.), *Interdisciplinary Studies on Information Structure (ISIS). Working Papers of the SFB 632* (Vol. 8). Potsdam: Universitätsverlag Potsdam, 1–38.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Bartels, C. (1997a). The Pragmatics of Wh-Question Intonation in English. *U. Penn Working Papers in Linguistics*, 4(2), 1–17.

Bartels, C. (1997b). Towards a compositional interpretation of English statement and question intonation. PhD thesis, University of Massachusetts Amherst.

Bartels, C. (2013). *The intonation of English statements and Questions. A compositional interpretation*. New York and London: Routledge (first published 1999 by Garland Publishing, New York).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. http://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Green, P. (2016). lme4: Linear mixed-effects models using “Eigen” and S4. R package version 1.1-13. Retrieved from https://CRAN.R-project.org/package=lme4

Bäuerle, Rainer. (1979). Questions and Answers. In R. Bäuerle, U. Egli & A. von Stechow (Eds.), *Semantics from different points of view*. Berlin, Heidelberg, New York: Springer, 61–74.Bierwisch, M. (1966). Regeln für die Intonation deutscher Sätze. In M. Bierwisch (Ed.), *Untersuchungen über Akzent und Intonation im Deutschen* (pp. 99–201). Berlin: Akademie-Verlag.

Biezma, M., & Rawlins, K. (2012). Responding to Alternative and Polar Questions. *Linguistics and Philosophy*, 35(5): 361–406. http://doi.org/10.1007/s10988-012-9123-z

Boersma, P., & Weenink, D. (2016). Praat. Doing phonetics by computer (Version 6.0.18). Retrieved from www.praat.org.

Braun, B., N. Dehé, J. Neitsch, D. Wochner & Zahner, K. (2018). The prosody of rhetorical and information-seeking questions in German. Advance Online Publication. *Language and Speech*. https://doi.org/10.1177/0023830918816351.

Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory*, 8(2), 277–312. https://doi.org/10.1515/cllt-2012-0011

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

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 249–254.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, XX(1), 37–46.

Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990-present. Retrieved April 18, 2019, from https://www.english-corpora.org/coca/

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990-2008+). *International Journal of Corpus Linguistics*, 14(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Dehé, N. & Braun, B. (2019). The prosody of rhetorical questions in English. *English Language and Linguistics*. Advance Online Publication. https://doi.org/10.1017/S1360674319000157

Domaneschi, F., Romero, M., & Braun, B. (2017). Bias in polar questions: Evidence from English and German production experiments. *Glossa*, 2(1), 1–28. http://doi.org/10.5334/gjgl.27

Escandell-Vidal, M. V. (1998). Intonation and procedural encoding: The case of Spanish interrogatives. In V. Rouchota & A. H. Jucker (Eds.), *Current issues in relevance theory* (pp. 169–203). Amsterdam/Philadelphia: John Benjamins.

Escandell-Vidal, M. V. (1999). Los enunciados interrogativos. Aspectos semánticos y pragmáticos.

Escandell-Vidal, V. (2017). Intonation and evidentiality in Spanish polar interrogatives. *Language and Speech*, 60(2), 224–241. https://doi.org/10.1177/0023830917698178

von Essen, O. (1964). *Grundzüge der hochdeutschen Satzintonation*. Ratingen / Düsseldorf: Alois Henn.

Féry, C. (1993). *German intonational patterns*. Tübingen: Max Niemeyer Verlag.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019) irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. Retrieved from https://CRAN.R-project.org/package=irr

Goodhue, Daniel. (2018). *On asking and answering biased polar questions* (PhD thesis). McGill University.

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 181(1993), 267–298.

Grice, M., Baumann, S., & Benzmüller, S. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press.

Grice, M., Baumann, S., Ritter, S., & Röhr, C. (2017). GToBI. Übungsmaterialien zur deutschen Intonation. Retrieved November 15, 2017, from http://www.gtobi.uni-koeln.de/index.html

Grice, M., & Savino, M. (1995). Intonation and communicative function in a regional variety of Italian. *Phonus* (Vol. 1). Saarbrücken: Saarland University, 19-32.

Grice, M., & Savino, M. (2003). Map Tasks in Italian: Asking questions about given, accessible and new information. *Catalan Journal of Linguistics, 2*, 153–180.

Gunlogson, C. (2001). True to form: Rising and falling declaratives as questions in English. PhD thesis, University of California, Santa Cruz.

Gussenhoven, C. (1983). *On the grammar and semantics of sentence accents*. Dordrecht: Foris Publications.

Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the grammar of English*. London: Equinox.

Hedberg, N., & Sosa, J. M. (2002). The Prosody of Questions in Natural Discourse. In *Proceedings of Speech Prosody 2002: The first international conference on speech prosody* (pp. 375–378). Aix-en-Provence: ISCA.

Höhle, T. N. (2013). Über Verum-Fokus im Deutschen. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (pp. 112–141). Opladen: Westdeutscher Verlag. https://doi.org/10.1007/978-3-663-12176-3_5

Isačenko, A., & Schädlich, H.-J. (1966). Untersuchungen über die deutsche Satzintonation. In M. Bierwisch (Ed.), *Untersuchungen über Akzent und Intonation im Deutschen* (pp. 7–67). Berlin: Akademie-Verlag.

James, E., Mahut, C., & Latkiewicz, G. (1989). The investigation of an apparently new intonation pattern in Toronto English. *Information Communication*, 10, 11–17.

Kügler, F. (2003). Do we know the answer? – Variation in yes-no question intonation. In S. Fischer & R. van de Vijver (Eds.), *Experimental Studies in Linguistics* 1 (pp. 9–29). Potsdam: Universitätsverlag Potsdam.

Kügler, F., Smolibocki, B., Arnold, D., Baumann, S., Braun, B., Grice, M., … Röhr, C. T. (2015). DIMA – Annotation Guidelines for German Intonation. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015) (Paper number 317). Glasgow, UK: University of Glasgow. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0317.pdfKohler, K. J. (1991). Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. In K. J. Kohler (Ed.), *Studies in German intonation. Arbeitsbericht Nr. 25 des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* (Vol. 25, pp. 115–185). Kiel: Universität Kiel.

Kohler, K. J. (2005a). Timing and communicative functions of pitch contours. *Phonetica*, 62(2–4), 88–105. https://doi.org/10.1159/000090091

Kohler, K. J. (2005b). Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) No. 35a). (K. J. Kohler, F. Kleber, & B. Peters, Eds.), *Prosodic Structures in German Spontaneous Speech*. Kiel: Institut für Phonetik und digitale Sprachverarbeitung (IPDS) der Universität Kiel, 125–142.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-33. Retrieved from https://cran.r-project.org/package=lmerTest

Ladd, D. R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In *Proceedings of Chicago linguistic society* vol. 17 (pp. 164–171).

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I Error and Power in Linear Mixed Models. *Journal of Memory and Language*, 94, 305–315. http://doi.org/10.1016/j.jml.2017.01.001

Michalsky, J. (2017). *Frageintonation im Deutschen: Zur intonatorischen Markierung von Interrogativität und Fragehaltigkeit*. Berlin/Boston: Walter de Gruyter.

Michelas, A., Portes, C., & Champagne-Lavau, M. (2016). When pitch accents encode speaker commitment: Evidence from French intonation. *Language and Speech*, 59, 266–293. http://doi.org/10.1177/0023830915587337

Neurobehavioral Systems (2014). Presentation. Version 17.0.

Niebuhr, O. (2003). Perceptual study of timing variables in F0 peaks. In *Proceedings of the 15th ICPhS, Barcelona, Spain* (pp. 1225–1228).

Nilsenová, M. (2006). Rises and falls: Studies in the semantics and pragmatics of intonation. PhD thesis, Amsterdam University.

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of Colloquial English* (2nd edition). London: Longman.

Peters, J. (2006). Intonation. In M. Wermke, K. Kunzel-Razum, & W. Scholze-Stubenrecht (Eds.), *Duden. Die Grammatik* (revised reprint of the 7th edition, pp. 95–128). Mannheim, Leipzig, Wien, Zürich: Dudenverlag.

Petrone, C., & Niebuhr, O. (2014). On the Intonation of German Intonation Questions: The Role of the Prenuclear Region. *Language and Speech*, 57(1), 108–146. http://doi.org/10.1177/0023830913495651

Pheby, J. (1975). *Intonation und Grammatik im Deutschen*. Berlin: Akademie-Verlag.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Prieto, P., & Borràs-Comes, J. (2018). Question intonation contours as dynamic epistemic operators. *Natural Language and Linguistic Theory*, 36, 563–586. http://doi.org/10.1007/s11049-017-9382-z

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Version 3.4.0 (2017-04-21). Retrieved from https://www.r-project.org/

Roberts, C. (2012). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics*, 6: 1–69. http://doi.org/10.3765/sp.5.6

Romero, M., & Han, C. (2002). Verum Focus in Negative Yes/No Questions and Ladd's p/-p Ambiguity. *Proceedings of SALT XII*, 1(4), 204–224.

Romero, M., & Han, C.-H. (2004). On Negative "Yes / No" Questions. *Linguistics and Philosophy*, 27(5), 609–658.

Romero, M., Arnhold, A., Braun, B., & Domaneschi, F. (2017). Negative Polar Question Types in English. In A. Lamont & K. A. Tetzloff (Eds.), *NELS 47: Proceedings of the Forty-Seventh Annual Meeting of the North East Linguistic Society: Volume 3*. Amherst, MA: GLSA: Graduate Linguistics Student Association, 35–48.

Rooth, M. (1985). Association with Focus. PhD thesis, University of Massachusetts, Amherst.

Rooth, M. (1992). A Theory of Focus Interpretation. *Natural Language Semantics*, 1(1): 75–116. http://doi.org/10.1007/BF02342617

van Rooy, R., & Šafářová, M. (2003). On polar questions. *Proceedings of SALT XIII*, 292–309.

Schubiger, M. (1958). *English intonation. Its form and function*. Tübingen: Max Niemeyer Verlag.

Shokeir, V. (2008). Evidence for the stable use of uptalk in South Ontario English. University of Pennsylvania Working Papers in Linguistics, 14(2), 16–24. Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., … Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In *Second International Conference on Spoken Language Processing* (pp. 867–870). http://doi.org/10.1515/cllt-2012-0011

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4), 649–689.

Steedman, M. (2008). Information-Structural Semantics for English Intonation. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus. Cross-linguistic perspectives on meaning and intonation* (pp. 245–264). Dordrecht: Springer. http://doi.org/10.1007/978-1-4020-4796-1_13

Stock, E. (1999). *Deutsche Intonation* (4th edition). Berlin, München, Leipzig: Langenscheidt - Verlag Enzyklopädie.

Stock, E., & Zacharias, C. (1973). *Deutsche Satzintonation* (2nd revised edition). Leipzig: VEB Verlag Enzyklopädie.

Sudo, Y. (2013). Biased Polar Questions in English and Japanese. In D. Gutzmann & H.-M. Gärtner (Eds.), *Beyond expressives: Explorations in use-conditional meaning* (pp. 275–295). Leiden: Brill. http://doi.org/10.1163/9789004183988_009

Talla Sando, Y. (2009). Upspeak across Canadian English accents: Acoustic and sociophonetic evidence. In *Proceedings of the 2009 annual conference of the Canadian Linguistic Association* (pp. 1–12).

Turco, G., Dimroth, C., & Braun, B. (2013). Intonational means to mark verum focus in German and French. *Language and Speech*, 56(564), 461–491. https://doi.org/10.1177/0023830912460506

Vanrell, M. D. M., Mascaro, I., Torres-Tamarit, F., & Prieto, P. (2013). Intonation as an Encoder of Speaker Certainty: Information and Confirmation Yes-No Questions in

Catalan. *Language and Speech*, *56*(2), 163–191. http://doi.org/10.1177/0023830912443942

Vanrell, M., Ballone, F., Schirru, C., & Prieto, P. (2014). Intonation and its interfaces in Sardinian polar questions, *Loquens* 1(2), e014. http://doi.org/10.3989/loquens.2014.014

Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). 6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI. January IAP 2006. Massachusetts Institute of Technology: MIT OpenCourseWare. Retrieved January 25, 2016, from https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/

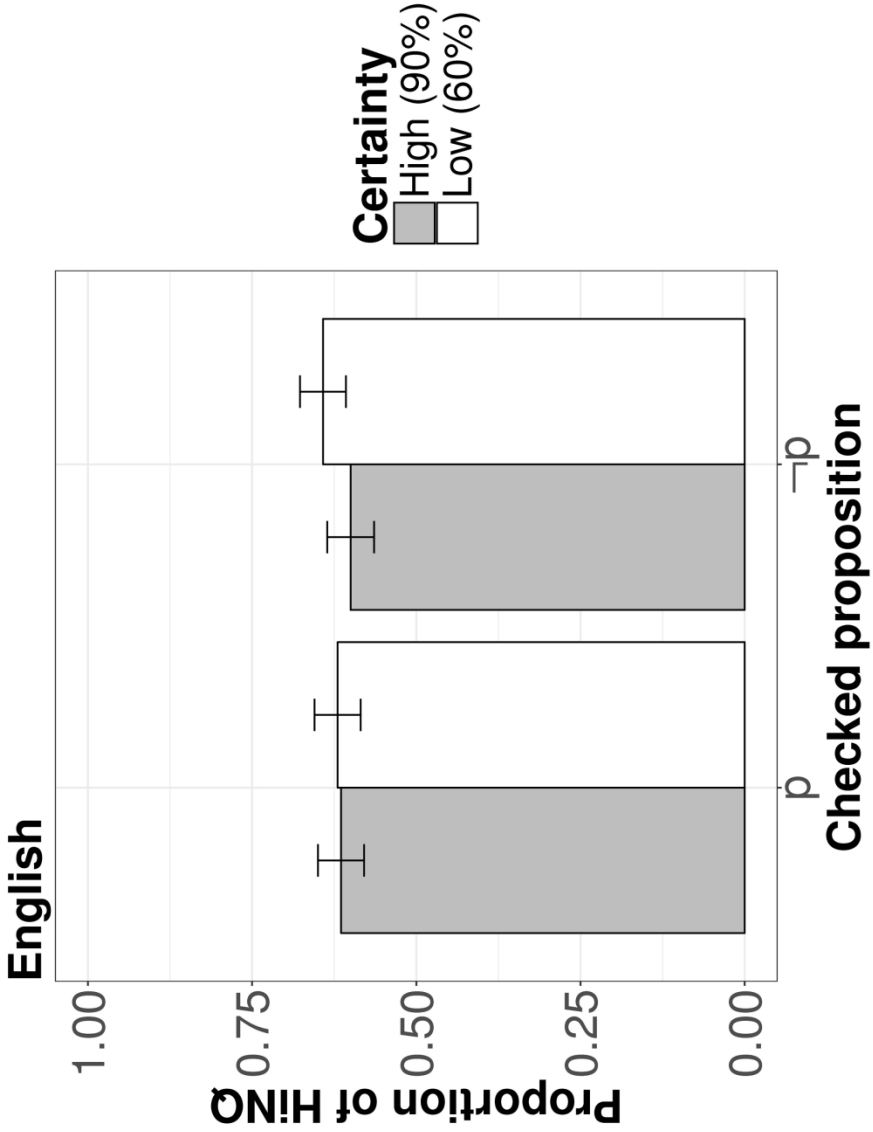

Figure 1. Participant's choice of HiNQ vs. LowNQ responses by certainty and checked proposition for English data. Error bars indicate standard error of the mean in all figures.

215x279mm (300 x 300 DPI)

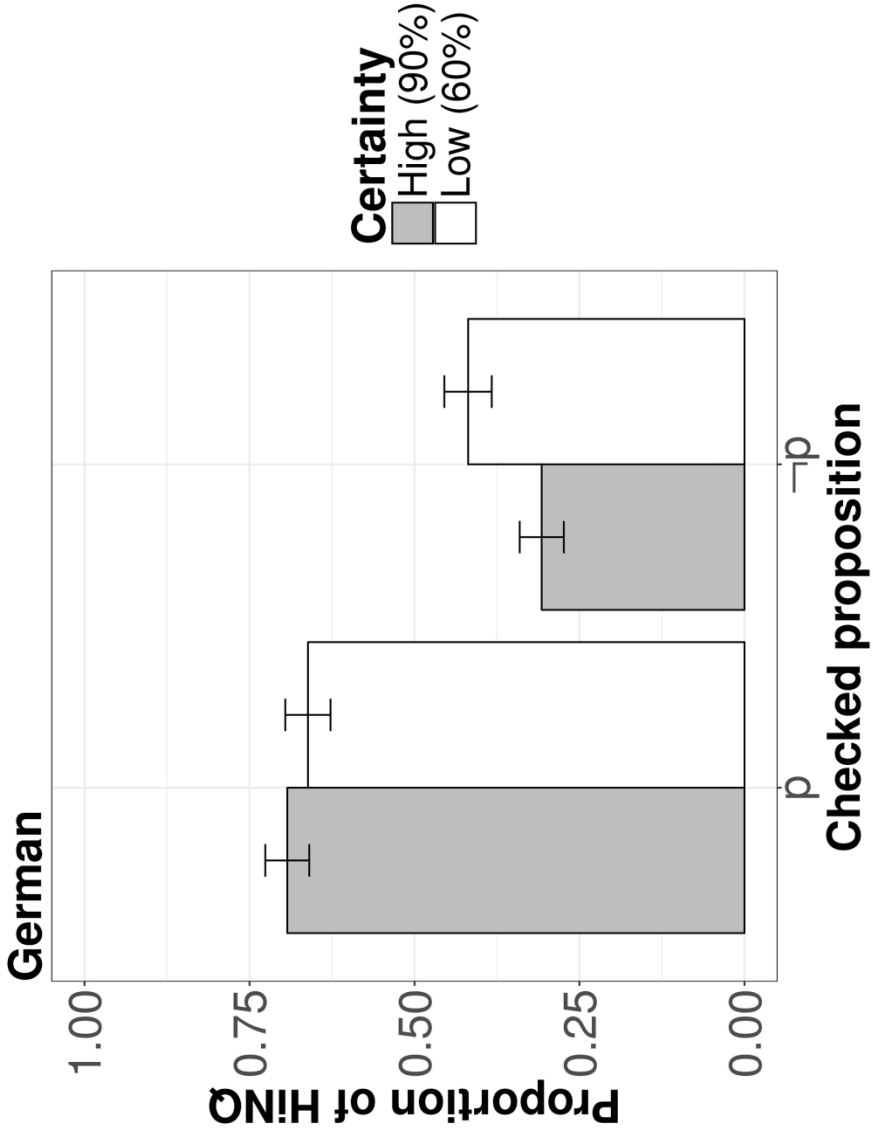

Figure 2. Participant's choice of HiNQ vs. LowNQ responses by certainty and checked proposition for German data.

215x279mm (300 x 300 DPI)

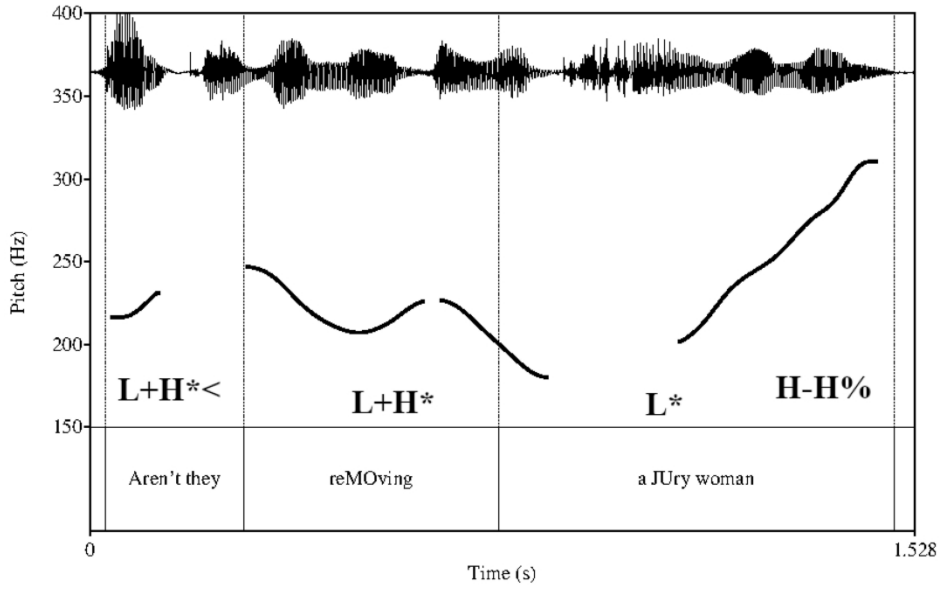

Figure 3. High final rise annotated with H-H% boundary tones. Utterance *Aren't they removing a jury woman?* (HiNQ) realized by participant 28. Accented syllables are indicated by capital letters.

241x152mm (300 x 300 DPI)

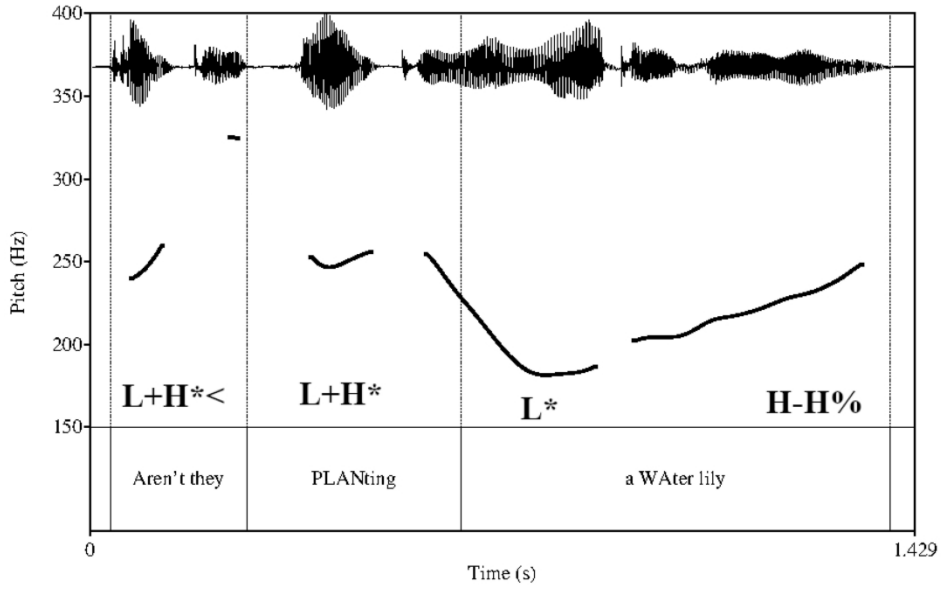

Figure 4. Less high final rise annotated with H-H% boundary tones. Utterance *Aren't they planting a water lily?* (HiNQ) realized by participant 28. Accented syllables are indicated by capital letters.

241x152mm (300 x 300 DPI)

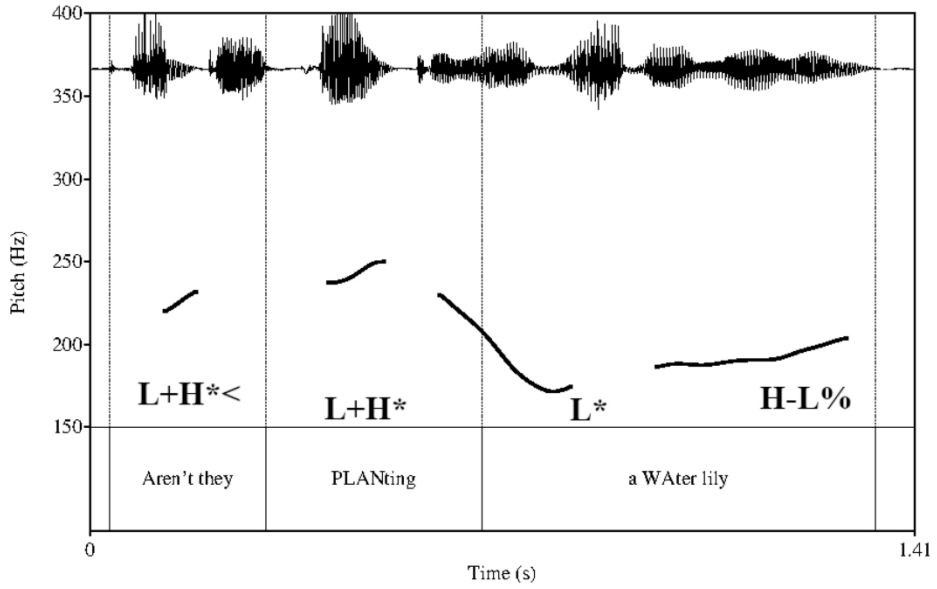

Figure 5. Final plateau annotated with H-L% boundary tones. Utterance *Aren't they planting a water lily?* (HiNQ) realized by participant 28. Accented syllables are indicated by capital letters.

241x152mm (300 x 300 DPI)

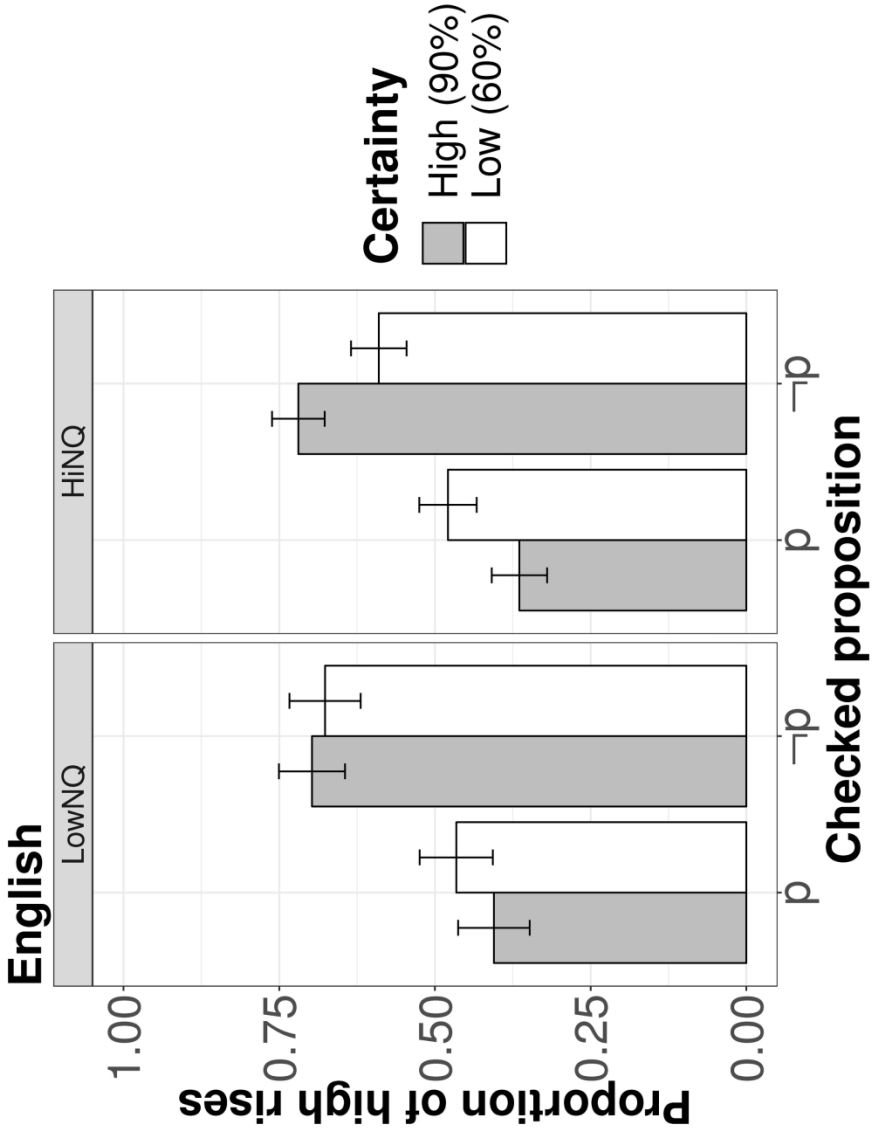

Figure 6. Proportion of question-final high f0 rises for English data.

215x279mm (300 x 300 DPI)

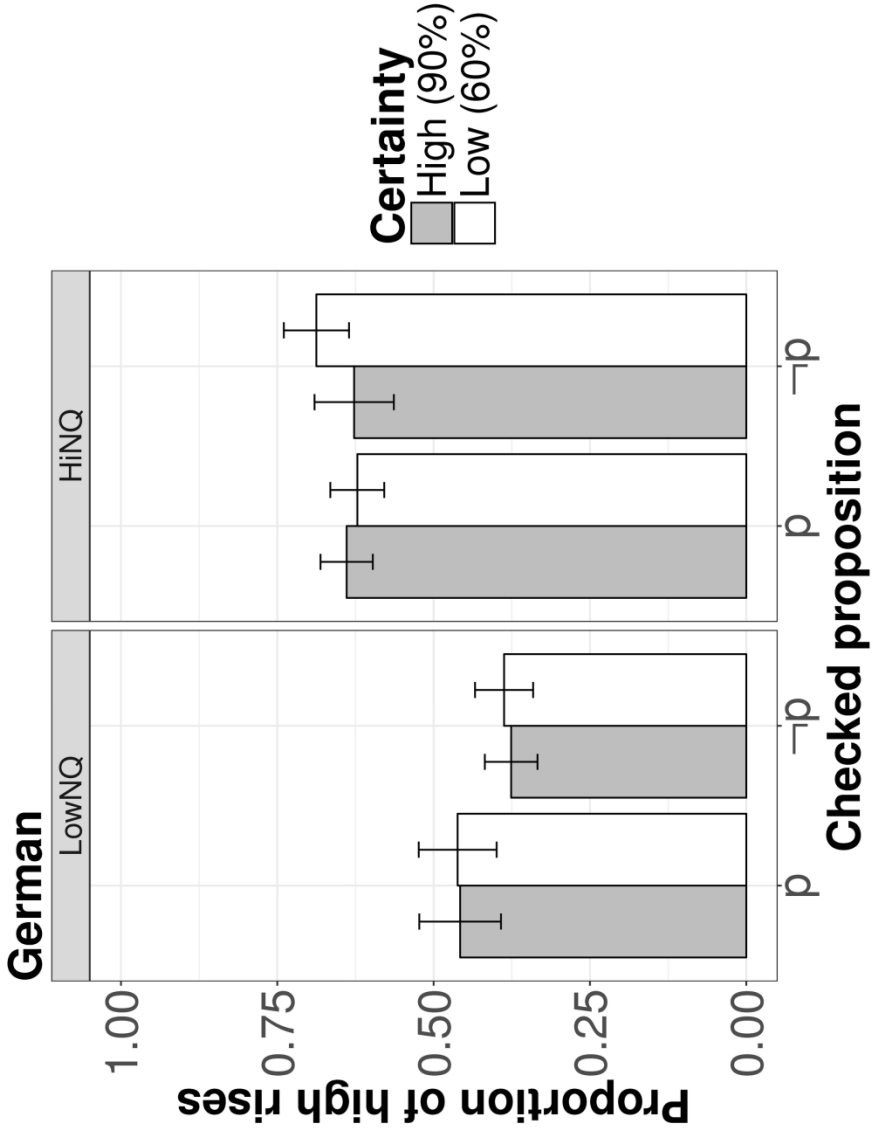

Figure 7. Proportion of question-final high f0 rises for German data.

215x279mm (300 x 300 DPI)

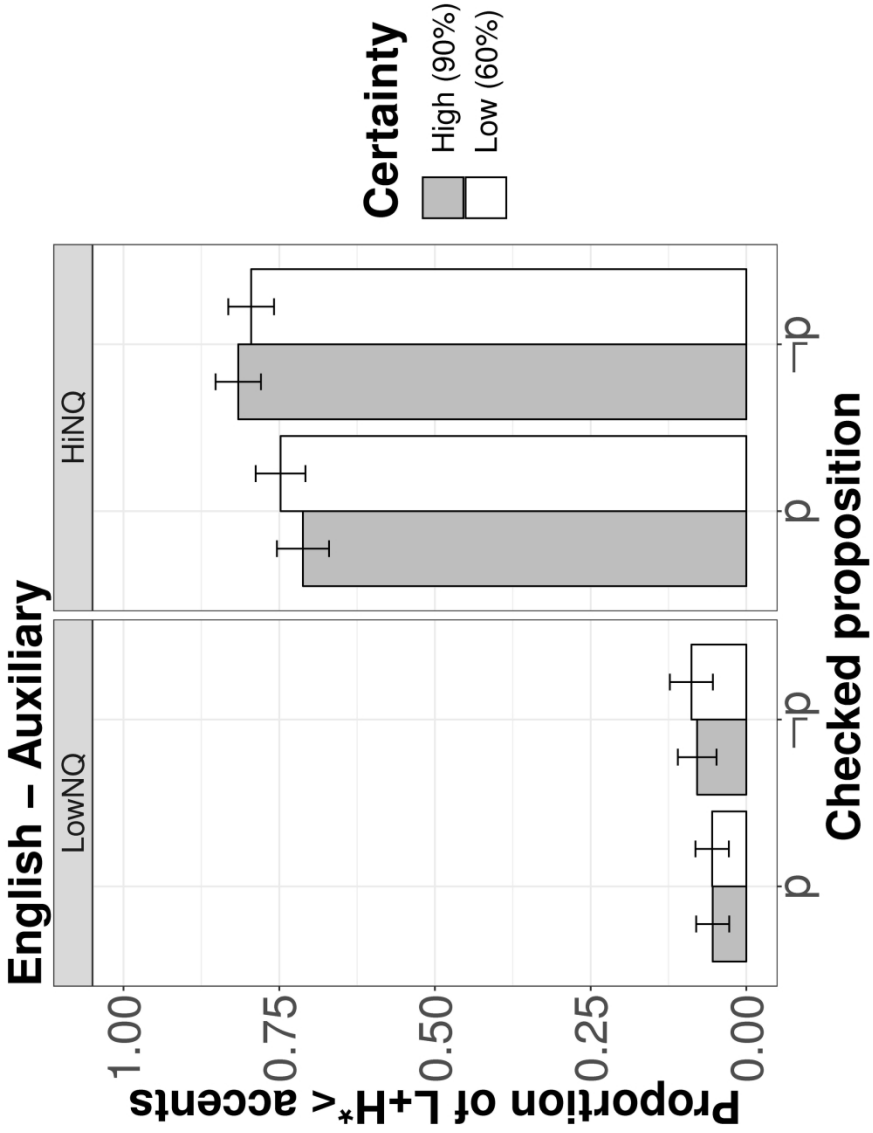

Figure 8. Proportion of auxiliaries realized with L+H*< accents in English data.

215x279mm (300 x 300 DPI)

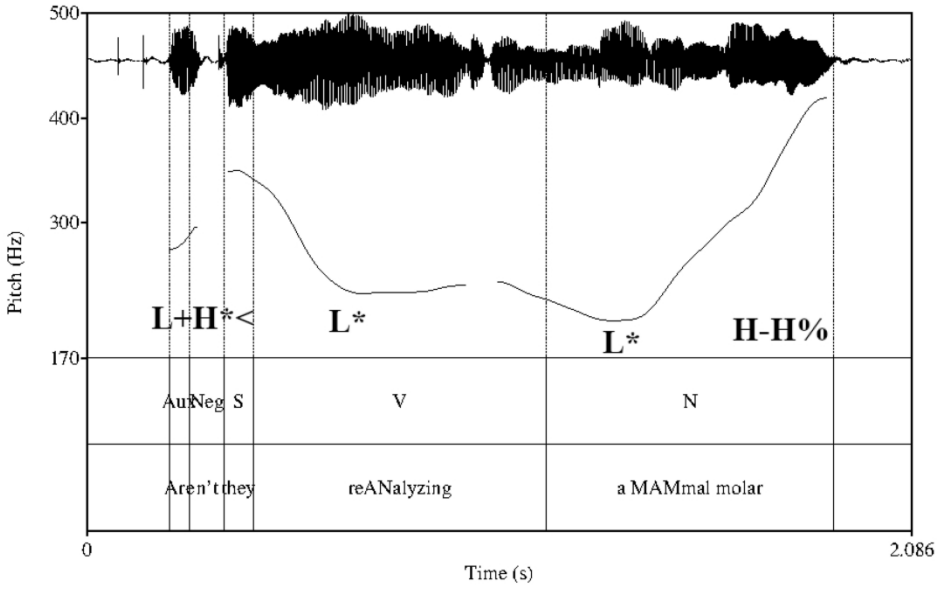

Figure 9. Sentence *Aren't they reanalyzing a mammal molar?* (HiNQ) realized with each constituent carrying the most frequent accent realization (Aux = auxiliary, Neg = negation, S = subject, V = main verb, N = sentence-final compound noun). Accented syllables are indicated by capital letters.

241x152mm (300 x 300 DPI)

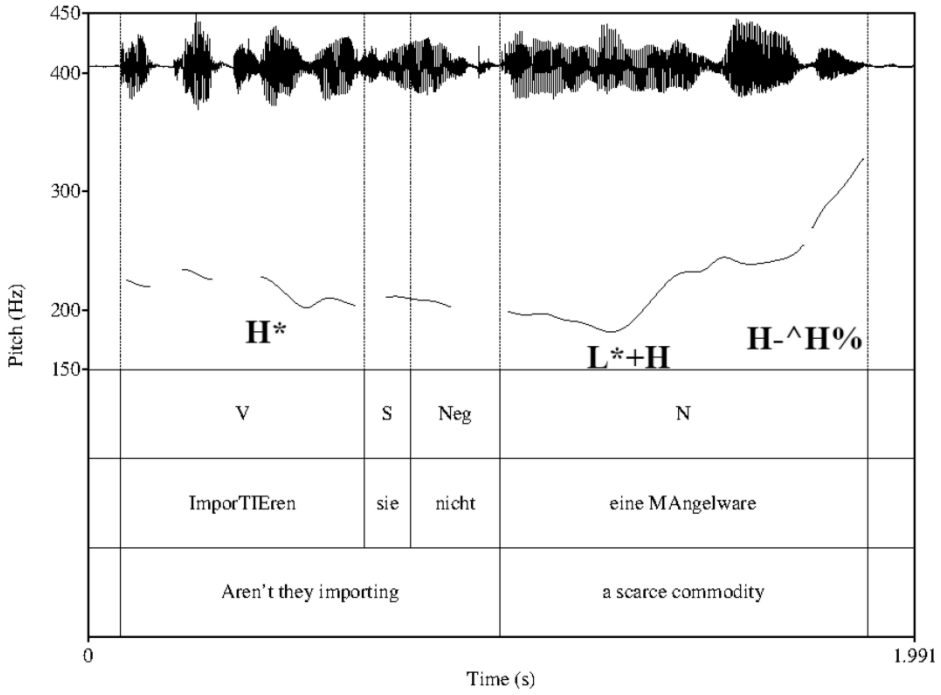

Figure 10. Sentence *Importieren sie nicht eine Mangelware?* 'Aren't they importing a scarce commodity?' (HiNQ) realized with each constituent carrying the most frequent accent realization (V = main verb, S = subject, Neg = negation, N = sentence-final compound noun). Accented syllables are indicated by capital letters.

241x178mm (300 x 300 DPI)

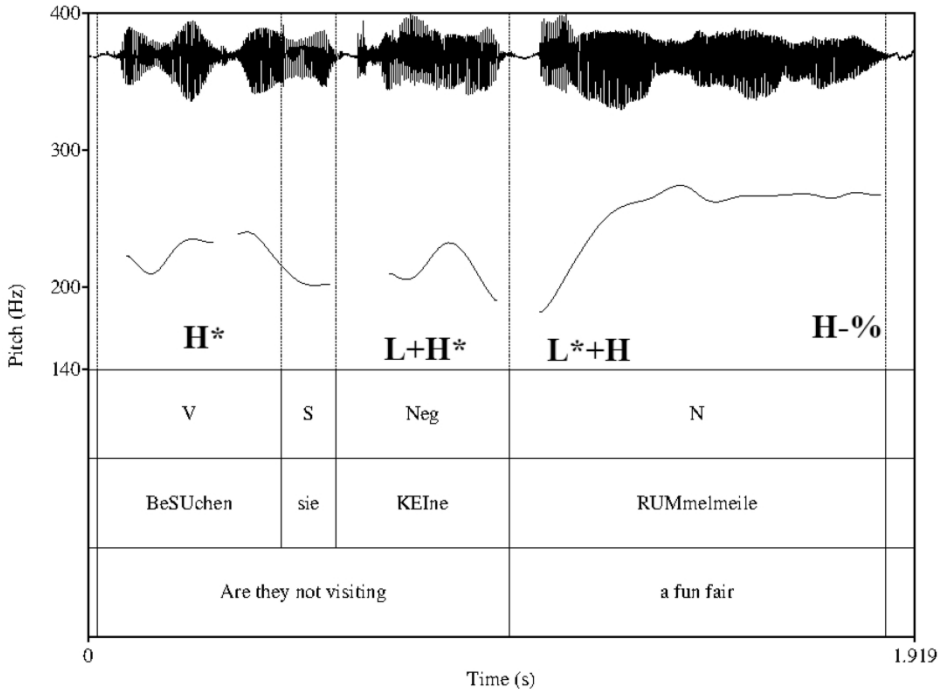

Figure 11. Sentence *Besuchen sie keine Rummelmeile?* 'Are they not visiting a fun fair?' (LowNQ) realized with each constituent carrying the accent realization most frequent in German LowNQs (V = main verb, S = subject, Neg = negation, N = sentence-final compound noun). Accented syllables are indicated by capital letters.

241x178mm (300 x 300 DPI)

Appendix A: List of items

Table A. List of target questions for all English items.

Item	HiNQ	LowNQ
1.	Aren't they assigning a learner level?	Are they not assigning a learner level?
2.	Aren't they adding a menu item?	Are they not adding a menu item?
3.	Aren't they cutting a mini melon?	Are they not cutting a mini melon?
4.	Aren't they inviting a novel writer?	Are they not inviting a novel writer?
5.	Aren't they releasing a vinyl volume?	Are they not releasing a vinyl volume?
6.	Aren't they invoking a motor neuron?	Are they not invoking a motor neuron?
7.	Aren't they celebrating a rally winner?	Are they not celebrating a rally winner?
8.	Aren't they installing an angle-iron?	Are they not installing an angle-iron?
9.	Aren't they modelling an atom layer?	Are they not modelling an atom layer?
10.	Aren't they stealing a venom vial?	Are they not stealing a venom vial?
11.	Aren't they deploying a NATO army?	Are they not deploying a NATO army?
12.	Aren't they analyzing a movie villain?	Are they not analyzing a movie villain?

13.	Aren't they photographing a river valley?	Are they not photographing a river valley?
14.	Aren't they reanalyzing a mammal molar?	Are they reanalyzing a mammal molar?
15.	Aren't they planting a water lily?	Are they not planting a water lily?
16.	Aren't they removing a jury woman?	Are they not removing a jury woman?

Table B. List of target questions for all German items.

Item	HiNQ	LowNQ
1.	Berücksichtigen sie nicht eine Lehrermeinung? 'Aren't they considering a teacher's opinion?'	Berücksichtigen sie keine Lehrermeinung? 'Are they not considering a teacher's opinion?'
2.	Sanieren sie nicht eine Römervilla? 'Aren't they reconstructing a Roman villa?'	Sanieren sie keine Römervilla? 'Are they not reconstructing a Roman villa?'
3.	Verwenden sie nicht eine Nylonrolle? 'Aren't they using a nylon paint roller?'	Verwenden sie keine Nylonrolle? 'Are they not using a nylon paint roller?'

4.	Beschäftigen sie nicht einen Maurerlehrling? 'Aren't they employing a bricklayer apprentice?'	Beschäftigen sie keinen Maurerlehrling? 'Are they not employing a bricklayer apprentice?'
5.	Veranstalten sie nicht ein Rallyrennen? 'Aren't they organising a rally race?'	Veranstalten sie kein Rallyrennen? 'Are they not organising a rally race?'
6.	Empfehlen sie nicht einen Nierenwärmer? 'Aren't they recommending sashes?'	Empfehlen sie keinen Nierenwärmer? 'Are they not recommending sashes?'
7.	Erwarten sie nicht einen Juraneuling? 'Aren't they expecting a law novice?'	Erwarten sie keinen Juraneuling? 'Are they not expecting a law novice?'
8.	Ersetzen sie nicht eine Neonröhre? 'Aren't they replacing a neon lamp?'	Ersetzen sie keine Neonröhre? 'Are they not replacing a neon lamp?'
9.	Vertreten sie nicht eine Malerinnung? 'Aren't they representing a painter's guild?'	Vertreten sie keine Malerinnung? 'Are they not representing a painter's guild?'

10.	Prognostizieren sie nicht eine Mehreinnahme? 'Aren't they predicting additional revenue?'	Prognostizieren sie keine Mehreinnahme? 'Are they not predicting additional revenue?'
11.	Repräsentieren sie nicht einen Wählerwillen? 'Aren't they representing [a] voters' intention?'	Repräsentieren sie keinen Wählerwillen? 'Are they not representing [a] voters' intention?'
12.	Importieren sie nicht eine Mangelware? 'Aren't they importing a scarce commodity?'	Importieren sie keine Mangelware? 'Are they not importing a scarce commodity?'
13.	Besuchen sie nicht eine Rummelmeile? 'Aren't they visiting a fun fair / amusement mile?'	Besuchen sie keine Rummelmeile? 'Are they not visiting a fun fair / amusement mile?'
14.	Planen sie nicht einen Angelweiher? 'Aren't they planning a fishing pond?'	Planen sie keinen Angelweiher? 'Are they not planning a fishing pond?'
15.	Tanzen sie nicht einen Ringelreihen?	Tanzen sie keinen Ringelreihen?

	'Aren't they dancing ring-a-ring-a-roses?'	'Are they not dancing ring-a-ring-a-roses?'
16.	Errichten sie nicht ein Marmormahnmal? 'Aren't they erecting a marble memorial?'	Errichten sie kein Marmormahnmal? 'Are they not erecting a marble memorial?'

Appendix B: Accents

Figure AppendixB1 shows the raw number of all accent type realizations by condition (proposition and certainty) for the separate constituents and sentence types. Table C lists the number and percentage of realizations with the most frequent accent realization for all constituents in the English data.

[insert Figure AppendixB1]

Table C. Proportion of most frequent accent realizations in the English data by constituent, checked proposition, certainty and morpho-syntactic question type, given in percent, rounded, and absolute number of realizations (in brackets).

Constituent	Realization	Checked proposition	Certainty	In HiNQs	In LowNQs
Auxiliary	L+H*<	Speaker's/*p*	High/90%	71% (84)	5% (4)
		Speaker's/*p*	Low/60%	75% (89)	5% (4)

		Addressee's/¬*p*	High/90%	82% (93)	8% (6)
		Addressee's/¬*p*	Low/60%	80% (97)	9% (6)
Negation	No accent	Speaker's/*p*	High/90%	100% (118)	4% (3)
		Speaker's/*p*	Low/60%	100% (119)	4% (3)
		Addressee's/¬*p*	High/90%	100% (114)	9% (7)
		Addressee's/¬*p*	Low/60%	100% (122)	10% (7)
Compound noun	L*	Speaker's/*p*	High/90%	94% (111)	82% (61)
		Speaker's/*p*	Low/60%	92% (110)	81% (59)
		Addressee's/¬*p*	High/90%	91% (104)	89% (68)
		Addressee's/¬*p*	Low/60%	93% (113)	84% (57)
Verb	L*	Speaker's/*p*	High/90%	36% (43)	28% (21)
		Speaker's/*p*	Low/60%	30% (36)	32% (23)
		Addressee's/¬*p*	High/90%	39% (44)	28% (21)
		Addressee's/¬*p*	Low/60%	36% (44)	38% (26)

For negations in the English data, Pearson's chi-squared test with Yates' continuity correction indicated that the presence of an accent on the negation was significantly influenced by question type, χ^2 (1) = 678.57, $p < 0.001$, i.e. there were significantly fewer accented negations in HiNQ than in LowNQ. However, this could not be confirmed with linear mixed-effects modelling due to the lack of variation in the HiNQ productions, which constituted the majority of the data. Therefore, we did not use question type as a predictor in these models. The resulting best binomial linear mixed-effects model did not indicate any significant effects of checked proposition or certainty, either. The main effect of proposition included in the model was not significant, β = -0.110, *SE* = 0.319, *z* = -0.34, *p* = 0.731. The model further contained a significant fixed effect of trial, β = 0.026, *SE* = 0.005, *z* = 5.49, *p* < 0.001, suggesting that participants accented negations (in LowNQs) more often over the course of the experimental session, as well as a by-participant slope of checked proposition in the random structure.

For utterance-final compound nouns, the best linear fixed-effects model indicated that the most frequent accent category, L*, appeared more often in HiNQs than in LowNQs, β = 15.778, *SE* = 5.403, *z* = 2.92, *p* = 0.004, but no other effects were significant (trimming removed 49 data points or 6%).

In evaluating accentuation of verbs in the English data, we found no significant effect of any of the predictors. We contrasted the most frequent realization, a low L* accent, with a second category comprising all other realizations. Note that there were not enough data points to test a three-way interaction between proposition, certainty and question type.

[insert Figure AppendixB2]

Table D. Proportion of most frequent accent realizations in the German data by constituent, checked proposition, certainty and morpho-syntactic question type, given in percent, rounded, and absolute number of realizations (in brackets).

Constituent	Realization	Checked proposition	Certainty	In HiNQ	In LowNQ
Negation	No accent	Speaker's/p	High/90%	59% (79)	12% (7)
		Speaker's/p	Low/60%	52% (66)	14% (9)
		Addressee's/¬p	High/90%	44% (26)	9% (12)
		Addressee's/¬p	Low/60%	56% (45)	8% (9)
Compound noun	L*+H	Speaker's/p	High/90%	74% (98)	49% (29)
		Speaker's/p	Low/60%	72% (92)	43% (28)
		Addressee's/¬p	High/90%	69% (41)	36% (48)
		Addressee's/¬p	Low/60%	74% (59)	36% (40)
Verb	H*	Speaker's/p	High/90%	53% (70)	44% (26)
		Speaker's/p	Low/60%	49% (62)	38% (25)
		Addressee's/¬p	High/90%	47% (28)	41% (55)
		Addressee's/¬p	Low/60%	41% (33)	45% (50)

For German (see Figure AppendixB2, Table D), linear mixed-effects modelling of accentuation of negation words indicated that checked proposition and certainty did not have a significant effect. Instead, the best model only contained an effect of question type, β = -3.932, *SE* = 0.332, *z* = -11.85, *p* < 0.001 (trimming removed 16 data points or 2% or the data). This effect suggested that participants accented negations overall less often in HiNQs than in LowNQs, i.e. HiNQs had a higher proportion of unaccented realizations.

German sentence-final compound nouns were most often realized with rising L*+H accents. This realization was overall more frequent in HiNQs than in LowNQs, as shown by the best binomial linear mixed-effects model, β = 1.818, *SE* = 0.367, *z* = 4.95, *p* < 0.001 (the model additionally contained a by-participant effect of question type). Thus, the accent L*+H was the overall most frequent realization for final compound nouns in both question types, though it accounted for 73% in HiNQs and only for 39% in LowNQs (*N* = 290 vs. *N* = 145). Finally, the model included a fixed effect of trial, indicating fewer L*+H accents later in the experimental sessions, β = -0.011, *SE* = 0.005, *z* = -2.17, *p* = 0.030.

For the accentuation of verbs in the German data, we evaluated a binary dependent variable distinguishing H* accents from all other realizations. Again, no significant effect of any of the predictors emerged. However, an insignificant fixed effect of question type, β = 0.365, *SE* = 0.273, *z* = 1.33, *p* = 0.182, was retained in the final model due to the presence of a by-participant effect of question type. In addition to the most frequent realization with H* accents (*N* = 349 or 46%), German verbs also appeared with low L*

accents (N = 203 or 26%), unaccented (N = 90 or 12%), with rising L+H* accents (N = 87 or 11%), with rising L*+H accents (N = 36 or 5%), and with H+!H* and H+L* accents (one realization each).

Appendix A: List of items

Table A. List of target questions for all English items.

Item	HiNQ	LowNQ
1.	Aren't they assigning a learner level?	Are they not assigning a learner level?
2.	Aren't they adding a menu item?	Are they not adding a menu item?
3.	Aren't they cutting a mini melon?	Are they not cutting a mini melon?
4.	Aren't they inviting a novel writer?	Are they not inviting a novel writer?
5.	Aren't they releasing a vinyl volume?	Are they not releasing a vinyl volume?
6.	Aren't they invoking a motor neuron?	Are they not invoking a motor neuron?
7.	Aren't they celebrating a rally winner?	Are they not celebrating a rally winner?
8.	Aren't they installing an angle-iron?	Are they not installing an angle-iron?
9.	Aren't they modelling an atom layer?	Are they not modelling an atom layer?
10.	Aren't they stealing a venom vial?	Are they not stealing a venom vial?
11.	Aren't they deploying a NATO army?	Are they not deploying a NATO army?
12.	Aren't they analyzing a movie villain?	Are they not analyzing a movie villain?

13.	Aren't they photographing a river valley?	Are they not photographing a river valley?
14.	Aren't they reanalyzing a mammal molar?	Are they reanalyzing a mammal molar?
15.	Aren't they planting a water lily?	Are they not planting a water lily?
16.	Aren't they removing a jury woman?	Are they not removing a jury woman?

Table B. List of target questions for all German items.

Item	HiNQ	LowNQ
1.	Berücksichtigen sie nicht eine Lehrermeinung? 'Aren't they considering a teacher's opinion?'	Berücksichtigen sie keine Lehrermeinung? 'Are they not considering a teacher's opinion?'
2.	Sanieren sie nicht eine Römervilla? 'Aren't they reconstructing a Roman villa?'	Sanieren sie keine Römervilla? 'Are they not reconstructing a Roman villa?'
3.	Verwenden sie nicht eine Nylonrolle? 'Aren't they using a nylon paint roller?'	Verwenden sie keine Nylonrolle? 'Are they not using a nylon paint roller?'

4.	Beschäftigen sie nicht einen Maurerlehrling? 'Aren't they employing a bricklayer apprentice?'	Beschäftigen sie keinen Maurerlehrling? 'Are they not employing a bricklayer apprentice?'
5.	Veranstalten sie nicht ein Rallyrennen? 'Aren't they organising a rally race?'	Veranstalten sie kein Rallyrennen? 'Are they not organising a rally race?'
6.	Empfehlen sie nicht einen Nierenwärmer? 'Aren't they recommending sashes?'	Empfehlen sie keinen Nierenwärmer? 'Are they not recommending sashes?'
7.	Erwarten sie nicht einen Juraneuling? 'Aren't they expecting a law novice?'	Erwarten sie keinen Juraneuling? 'Are they not expecting a law novice?'
8.	Ersetzen sie nicht eine Neonröhre? 'Aren't they replacing a neon lamp?'	Ersetzen sie keine Neonröhre? 'Are they not replacing a neon lamp?'
9.	Vertreten sie nicht eine Malerinnung? 'Aren't they representing a painter's guild?'	Vertreten sie keine Malerinnung? 'Are they not representing a painter's guild?'

10.	Prognostizieren sie nicht eine Mehreinnahme? 'Aren't they predicting additional revenue?'	Prognostizieren sie keine Mehreinnahme? 'Are they not predicting additional revenue?'
11.	Repräsentieren sie nicht einen Wählerwillen? 'Aren't they representing [a] voters' intention?'	Repräsentieren sie keinen Wählerwillen? 'Are they not representing [a] voters' intention?'
12.	Importieren sie nicht eine Mangelware? 'Aren't they importing a scarce commodity?'	Importieren sie keine Mangelware? 'Are they not importing a scarce commodity?'
13.	Besuchen sie nicht eine Rummelmeile? 'Aren't they visiting a fun fair / amusement mile?'	Besuchen sie keine Rummelmeile? 'Are they not visiting a fun fair / amusement mile?'
14.	Planen sie nicht einen Angelweiher? 'Aren't they planning a fishing pond?'	Planen sie keinen Angelweiher? 'Are they not planning a fishing pond?'
15.	Tanzen sie nicht einen Ringelreihen?	Tanzen sie keinen Ringelreihen?

	'Aren't they dancing ring-a-ring-a-roses?'	'Are they not dancing ring-a-ring-a-roses?'
16.	Errichten sie nicht ein Marmormahnmal? 'Aren't they erecting a marble memorial?'	Errichten sie kein Marmormahnmal? 'Are they not erecting a marble memorial?'

Appendix B: Accents

Figure AppendixB1 shows the raw number of all accent type realizations by condition (proposition and certainty) for the separate constituents and sentence types. Table C lists the number and percentage of realizations with the most frequent accent realization for all constituents in the English data.

[insert Figure AppendixB1]

Table C. Proportion of most frequent accent realizations in the English data by constituent, checked proposition, certainty and morpho-syntactic question type, given in percent, rounded, and absolute number of realizations (in brackets).

Constituent	Realization	Checked proposition	Certainty	In HiNQs	In LowNQs
Auxiliary	L+H*<	Speaker's/*p*	High/90%	71% (84)	5% (4)
		Speaker's/*p*	Low/60%	75% (89)	5% (4)

		Addressee's/¬*p*	High/90%	82% (93)	8% (6)
		Addressee's/¬*p*	Low/60%	80% (97)	9% (6)
Negation	No accent	Speaker's/*p*	High/90%	100% (118)	4% (3)
		Speaker's/*p*	Low/60%	100% (119)	4% (3)
		Addressee's/¬*p*	High/90%	100% (114)	9% (7)
		Addressee's/¬*p*	Low/60%	100% (122)	10% (7)
Compound noun	L*	Speaker's/*p*	High/90%	94% (111)	82% (61)
		Speaker's/*p*	Low/60%	92% (110)	81% (59)
		Addressee's/¬*p*	High/90%	91% (104)	89% (68)
		Addressee's/¬*p*	Low/60%	93% (113)	84% (57)
Verb	L*	Speaker's/*p*	High/90%	36% (43)	28% (21)
		Speaker's/*p*	Low/60%	30% (36)	32% (23)
		Addressee's/¬*p*	High/90%	39% (44)	28% (21)
		Addressee's/¬*p*	Low/60%	36% (44)	38% (26)

For negations in the English data, Pearson's chi-squared test with Yates' continuity correction indicated that the presence of an accent on the negation was significantly influenced by question type, χ^2 (1) = 678.57, $p < 0.001$, i.e. there were significantly fewer accented negations in HiNQ than in LowNQ. However, this could not be confirmed with linear mixed-effects modelling due to the lack of variation in the HiNQ productions, which constituted the majority of the data. Therefore, we did not use question type as a predictor in these models. The resulting best binomial linear mixed-effects model did not indicate any significant effects of checked proposition or certainty, either. The main effect of proposition included in the model was not significant, β = -0.110, *SE* = 0.319, *z* = -0.34, *p* = 0.731. The model further contained a significant fixed effect of trial, β = 0.026, *SE* = 0.005, *z* = 5.49, *p* < 0.001, suggesting that participants accented negations (in LowNQs) more often over the course of the experimental session, as well as a by-participant slope of checked proposition in the random structure.

For utterance-final compound nouns, the best linear fixed-effects model indicated that the most frequent accent category, L*, appeared more often in HiNQs than in LowNQs, β = 15.778, *SE* = 5.403, *z* = 2.92, *p* = 0.004, but no other effects were significant (trimming removed 49 data points or 6%).

In evaluating accentuation of verbs in the English data, we found no significant effect of any of the predictors. We contrasted the most frequent realization, a low L* accent, with a second category comprising all other realizations. Note that there were not enough data points to test a three-way interaction between proposition, certainty and question type.

[insert Figure AppendixB2]

Table D. Proportion of most frequent accent realizations in the German data by constituent, checked proposition, certainty and morpho-syntactic question type, given in percent, rounded, and absolute number of realizations (in brackets).

Constituent	Realization	Checked proposition	Certainty	In HiNQ	In LowNQ
Negation	No accent	Speaker's/p	High/90%	59% (79)	12% (7)
		Speaker's/p	Low/60%	52% (66)	14% (9)
		Addressee's/¬p	High/90%	44% (26)	9% (12)
		Addressee's/¬p	Low/60%	56% (45)	8% (9)
Compound noun	L*+H	Speaker's/p	High/90%	74% (98)	49% (29)
		Speaker's/p	Low/60%	72% (92)	43% (28)
		Addressee's/¬p	High/90%	69% (41)	36% (48)
		Addressee's/¬p	Low/60%	74% (59)	36% (40)
Verb	H*	Speaker's/p	High/90%	53% (70)	44% (26)
		Speaker's/p	Low/60%	49% (62)	38% (25)
		Addressee's/¬p	High/90%	47% (28)	41% (55)
		Addressee's/¬p	Low/60%	41% (33)	45% (50)

For German (see Figure AppendixB2, Table D), linear mixed-effects modelling of accentuation of negation words indicated that checked proposition and certainty did not have a significant effect. Instead, the best model only contained an effect of question type, β = -3.932, *SE* = 0.332, *z* = -11.85, *p* < 0.001 (trimming removed 16 data points or 2% or the data). This effect suggested that participants accented negations overall less often in HiNQs than in LowNQs, i.e. HiNQs had a higher proportion of unaccented realizations.

German sentence-final compound nouns were most often realized with rising L*+H accents. This realization was overall more frequent in HiNQs than in LowNQs, as shown by the best binomial linear mixed-effects model, β = 1.818, *SE* = 0.367, *z* = 4.95, *p* < 0.001 (the model additionally contained a by-participant effect of question type). Thus, the accent L*+H was the overall most frequent realization for final compound nouns in both question types, though it accounted for 73% in HiNQs and only for 39% in LowNQs (*N* = 290 vs. *N* = 145). Finally, the model included a fixed effect of trial, indicating fewer L*+H accents later in the experimental sessions, β = -0.011, *SE* = 0.005, *z* = -2.17, *p* = 0.030.

For the accentuation of verbs in the German data, we evaluated a binary dependent variable distinguishing H* accents from all other realizations. Again, no significant effect of any of the predictors emerged. However, an insignificant fixed effect of question type, β = 0.365, *SE* = 0.273, *z* = 1.33, *p* = 0.182, was retained in the final model due to the presence of a by-participant effect of question type. In addition to the most frequent realization with H* accents (*N* = 349 or 46%), German verbs also appeared with low L*

accents (*N* = 203 or 26%), unaccented (*N* = 90 or 12%), with rising L+H* accents (*N* = 87 or 11%), with rising L*+H accents (*N* = 36 or 5%), and with H+!H* and H+L* accents (one realization each).

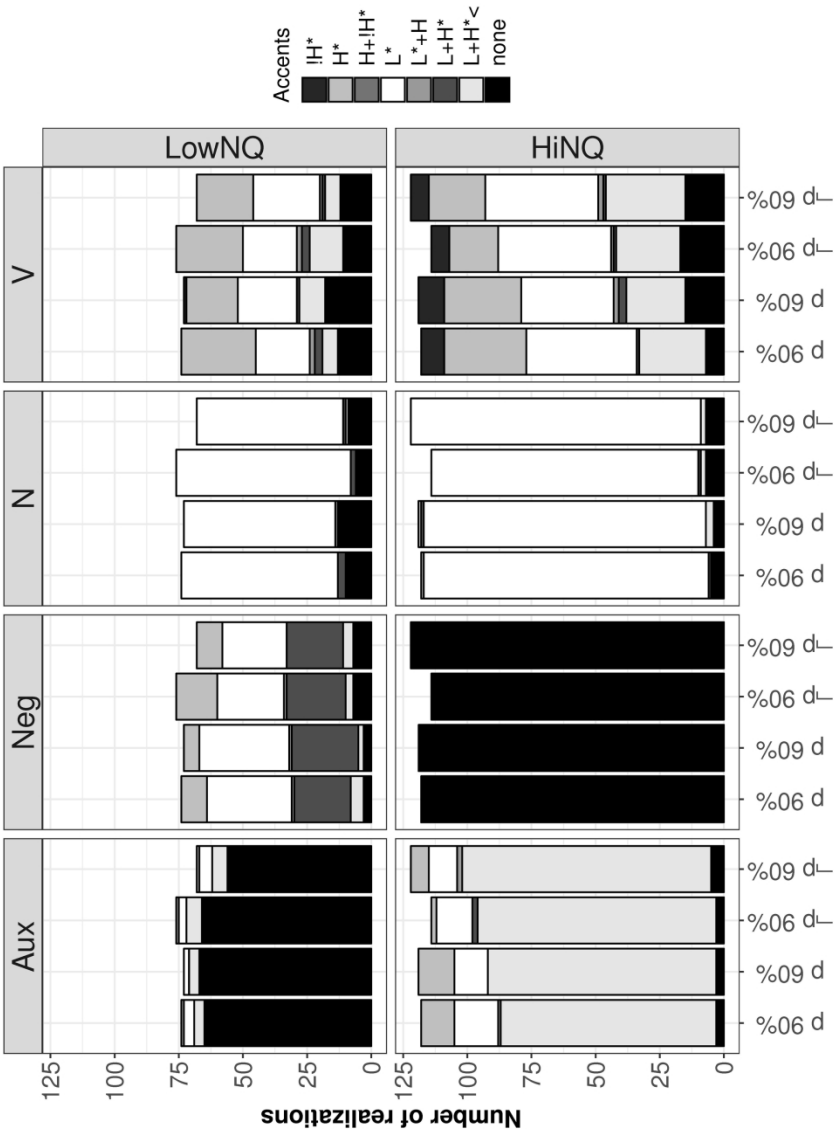

Figure AppendixB1. Number of realization of different accent categories by constituent, morpho-syntactic question type and condition crossing proposition checked (speaker's/p vs. addressee's/¬p) and certainty regarding the proposition (high 90% certain vs. low 60% certain) in the English data.

215x279mm (300 x 300 DPI)

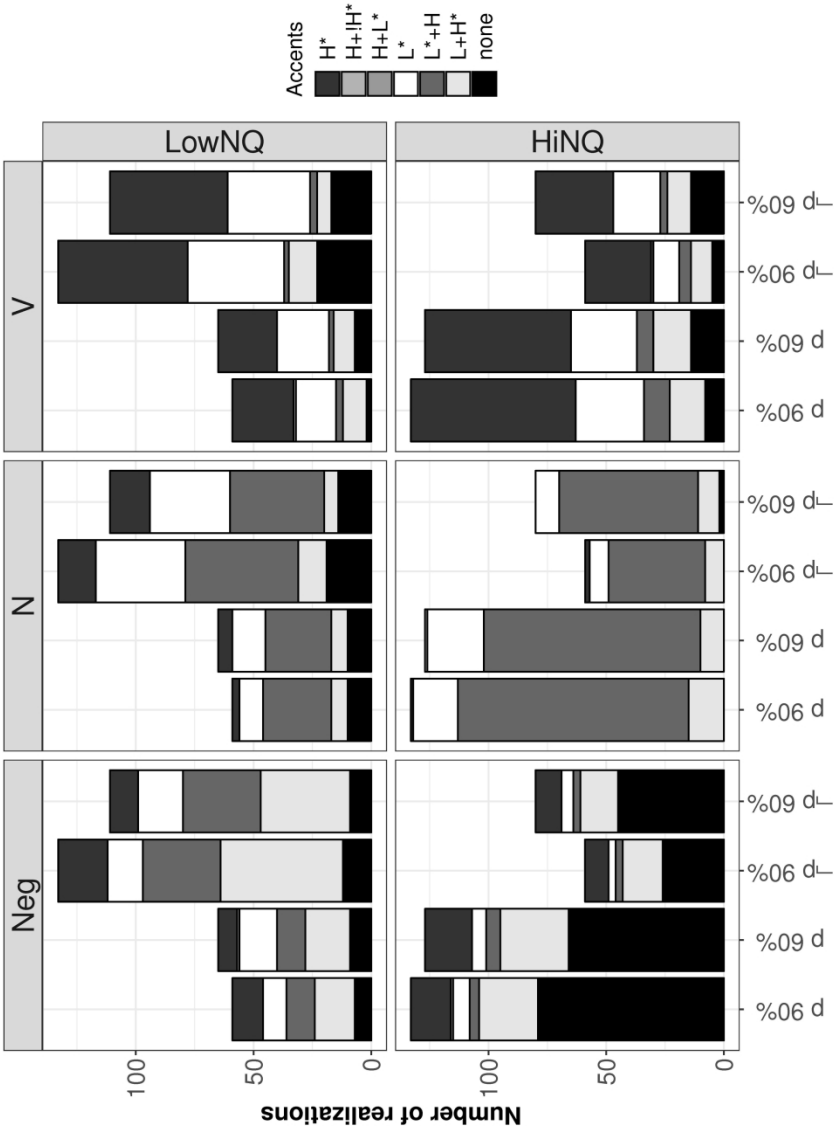

Figure AppendixB2. Number of realization of different accent categories by constituent, morpho-syntactic question type and condition crossing proposition checked (speaker's/p vs. addressee's/¬p) and certainty regarding the proposition (high 90% certain vs. low 60% certain) in the German data.

215x279mm (300 x 300 DPI)