

Polar Express: Polar Question Forms Expressing Bias-Evidence Conflicts in Italian

Maria Di Maro (Corresponding author)

University of Naples 'Federico II', Naples, Italy

E-mail: maria.dimaro2@unina.it

Antonio Origlia

University of Naples 'Federico II', Naples, Italy

Francesco Cutugno

University of Naples 'Federico II', Naples, Italy

Received: April 9, 2021

Accepted: May 19, 2021

Published: July 19, 2021

doi:10.5296/ijl.v13i4.18871

URL: <https://doi.org/10.5296/ijl.v13i4.18871>

Abstract

Past research has concentrated on the use of different forms of polar questions in specific contexts, defined in terms of the relationship between original bias and contextual evidence. It has been showed that, for English and German, people tend to prefer specific forms given the pragmatic context. Based on previous experiments, in this work, we observe that the same tendencies occur in Italian. Also, we adopt a more refined experimental setup with three different tasks and a more natural evaluation scale to better capture nuances in appropriateness evaluations, provided by human subjects, which therefore reflects the more realistic one-to-many relationship among forms and functions. In fact, the results show how specific forms of polar questions are especially typical of situations where the bias has the opposite value with respect to the evidence, i.e., in positive bias versus negative evidence, for which a high negative polar question in the past tense was more frequently selected by the subjects (Note 1).

Keywords: Polar questions, Common ground inconsistencies, Bias

1. Introduction

Questions are utterances that seek for verbal or other semiotic responses [Hayano, 2013, Sidnell and Stivers, 2012, 395]. They can lead to different interactional outcomes: i) imposing presuppositions, agendas, and preferences; ii) generating various actions that might be potentially face-threatening; iii) causing interlocutors to elaborate an answer [Brown and Levinson, 1978]. Studies on questions demonstrate that there are three different aspects that are to be studied when dealing with interrogatives, namely: grammar, prosody, and epistemic asymmetry. With epistemic asymmetry, it is intended the different degree of knowledge of the interlocutors toward any specific topic [Sidnell, 2012]. According to the type of question - content, polar, or alternative questions - these levels of analysis interconnect with each other to express specific functions, as it will be shown in this work specifically as far as grammar and epistemic asymmetry are concerned.

In this work, the relationship between the grammar and the epistemic asymmetry of polar questions (PQs) is, indeed, described, as this type of questions appears to be mostly preferred when *Common Ground Inconsistencies* occur. With *Common Ground Inconsistencies* we refer to conflicts arising when grounded information clashes with new contextual evidence. In general, PQs are defined as questions that make “relevant affirmation/confirmation or disconfirmation” [Stivers and Enfield, 2010]. PQs can have two possible binary answers: true or false. Many languages have grammatical markings which distinguish PQs from declarative sentences (word order, question particles, etc.). Each language, moreover, can also have many ways to ask PQs [Hayano, 2013, Sidnell and Stivers, 2012, 396]. In English, for example, we can generally have the following classes:

- i positive polar questions [PPQs] (i.e., *Did he bring food?*)
- ii high negation polar questions [HNPNQ] (i.e., *Didn't he bring food?*)
- iii low negation polar questions [LNPNQ] (i.e., *Did he bring no food?*)
- iv tag questions [Tag] (i.e., *He brought food, didn't he?*)

Nevertheless, it is important to remember that these standard grammatical types are not specific for questions in any context, since an utterance in one of these interrogative forms does not necessarily do questioning, and, on the other hand, a non-interrogative utterance can also function as a question [Hayano, 2013, Sidnell and Stivers, 2012, 396].

Interrogative prosody is another linguistic criterion which is used in most languages. In Italian, Arabic, and Romanian, for example, rising intonation is described as a conventionalised mean to ask PQs [Dryer, 2013]. More in detail, Italian is considered one of the 173 languages which have been mapped in *The World Atlas of Language Structure Online* as a language with interrogative intonation as the only interrogative marker (Note 2) [Dryer, 2013]. However, it is misleading to consider just the intonation as a strongly indicative mark of PQs. In fact, PQs are not necessarily marked by intonational movements, and intonational movements are not necessarily used exclusively with questions [Hayano, 2013, Sidnell and Stivers, 2012, 396]. Furthermore, there are languages that do not have specific grammatical or intonational resources to mark PQs, as, for instance, in a documented Papuan language, Yáí Dnye

[Levinson, 2010]. The way speakers interpret a question as a PQ depends, therefore, on other factors as well. An important criterion is, indeed, the domain of knowledge that is the semantic domain of expertise of a speaker. For instance, [Labov and Fanshel, 1977] stated that “when a speaker makes a statement about an event that falls into the recipient knowledge domain (B-event statement), it functions as a polar question and elicits confirmation or disconfirmation”. The authors made a distinction between *A-events*, which are known to *A* but not to *B*, and *B-events*, which are known to *B* but not to *A*. When *a* makes a statement, which is not part of her domain of knowledge, the statement will be interpreted as a question. More specifically, in Levinson’s words [Levinson, 2010], “[...] when an utterance addresses information that the speaker does not know but a recipient is likely to know, it is treated and responded to as a PQ or a confirmation request” [Hayano, 2013, Sidnell and Stivers, 2012, 397]. The *epistemic asymmetry* plays, therefore, a crucial role in the interpretation of questions. However, the epistemic stance of the questioner can correspond to diverse gradients which can correspond to different syntactic forms: Q1) *who did you talk to?* - content questions suggest that the questioner has little knowledge about the topic (higher K-); Q2) *you talked to Steve?* - declarative interrogatives suggest that the speaker expects a positive answer, as they know more about the topic (lower K-); Q3) *you were talking to Steve, weren’t you?* - Positive statements followed by question tags suggests that the speaker strongly believes in their presupposition for which they just need a confirmation, that is a positive answer (higher K+) [Hayano, 2013, Sidnell and Stivers, 2012, 399]. For this reason, the epistemic gradient and, therefore, the bias of the speaker towards a presupposition and a consequent expected answer also determine the grammar and prosody of the question. In fact, some studies also showed that falling intonation in PQs is associated with higher certainty, whereas rising intonation with lower certainty [Couper-Kuhlen, 2012]. Since the epistemic stance can be encoded in the grammatical and prosodic form of PQs, this means that these questions can function as an open door to the mental state of the questioner, who has a specific opinion about certain information. According to [Oshima, 2017], PQs convey an epistemic bias toward a positive or negative answer. In fact, as [Bolinger, 1978] stated, PQs advance a *hypothesis for confirmation*, where the hypothesis can be positive (i.e., *I strongly believe that this is true*), neutral (i.e., *I don’t know a lot about it, therefore I need confirmation on that*), or negative (i.e., *I strongly believe that this is false*). This means that questions do not only serve to request information, as they can also be employed as a powerful tool to control the answerer’s reactions.

PQs can, therefore, convey three major constraints:

- i Presuppositions: defined as the background beliefs of the speaker encoded in a statement whose validity is taken for granted [Stalnaker, 1977], presuppositions are usually used in questions unproblematically; nevertheless, questioners can also embed hostile presuppositions in questions [Hayano, 2013, Bolinger and Stivers, 2012, 401], in that they convey and impose questioners’ beliefs on recipients. Specifically, in this work, the positive bias has been interpreted as the expectation of a positive answer depending on the positive epistemic stance towards a previous presupposition. Different works have focused on the study of bias in polar questions in different languages and/or varieties [Malamud et al., 2015, Frana et al., 2019, Orrico et al., 2020, Arnhold et al. 2021].

- ii Agendas: questions set agendas concerning the topic (what the questioner is talking about) and the action (what the questioner is doing with the question, i.e., suggesting an answer), which both can be biased [Hayano, 2013, Sidnell and Stivers, 2012, 402].
- iii Preferences: when questioners pose questions, they can set preferences, such as answers over non-answer responses, or affirmation over disaffirmation; concerning this last point, PQ forms typically display a preference for:
- affirmation
 - positive polar questions: *Have you heard from her?*
 - positive statements combined with a question tag: *You've heard from her, haven't you?*
 - positive declarative questions: *You heard from her?*
 - negative polar questions: *Haven't you heard from her?*
 - disaffirmation
 - negative declarative: *You haven't heard from her/You never heard from her*
 - negative statements combined with a question tag: *You haven't heard from her, have you?*
 - positive polar questions combined with a negative polarity item: *Have you heard from her yet?*
 - positive interrogatives combined with negatively tiled adverbs: *Have you really heard from her?*

[Hayano, 2013, Sidnell and Stivers, 2012, 405]

As previously pointed out, when a speaker has a K- position, their utterance is recognised as a question. Besides this crucial factor, it can be noticed that a questioner has a position closer to the K+, in the case they already possess that specific knowledge. In this paper, this is the case of a previous grounded knowledge. When this happens, the question tends to be interpreted as a criticism or challenge [Hayano, 2013, Sidnell and Stivers, 2012, 410]. As [Steensig and Drew, 2008] acknowledged, “asking a question is not an innocent thing to do; when a question is asked about what its recipient has said or done, it carries a possible implication of disaffiliation”. In the terms of this work, when a Common Ground Inconsistency occurs, the questioner, referring to the knowledge that is already stored in the common ground, challenges the answerer, who states something contradicting the questioner's K+. In this context, a PQ is therefore uttered. This, in turn, expects a positive answer because of a strong belief towards that stored presupposition.

In this paper, we firstly investigate the motivations guiding this research in section 2, then we present the experiment we carried out (section 3) along with the results (section 4). To

conclude, we discuss the collected results whose importance is highlighted with respect to application scenarios.

2. Motivations

As pointed out in the previous section, PQs internally embed not only a mere request but also presuppositions, agendas and preferences. Furthermore, when the questioner is closer to a K+ position, the use of a PQ can also implicate a disaffiliation. In this case, we refer to epistemically biased questions. According to the literature, one way of expressing disaffiliation is through the use of *Reversed Polarity Questions* that are questions that convey bias towards the opposite valence than the utterance [Koshik, 2002, 2005]. For example, negative interrogatives can also function as positive assertions challenging the recipient's position [Heritage, 2002]. Criticisms and challenges can also be expressed through declaratives (i.e., *You shouldn't have done that*), imperatives (i.e., *Don't do that to me again*), or exclamations (i.e., *How dare you?*), which are perceived more confrontational and explicit and can be therefore face-threatening [Hayano, 2013, Sidnell and Stivers, 2012, 411]. Among non-standard communications, conflicting representations [Huang, 2017] are listed as interactions taking place when a discrepancy between what is communicated and what is believed by the agent occurs. In these scenarios, PQs can, therefore, serve as a knowledge challenging tool.

Different authors pointed out how either the original bias of the speaker or the contextual evidence bias could influence the syntactic form of PQs. These two types of bias are defined as follows:

Original speaker bias (B): “[...] belief or expectation of the speaker that p is true, based on his epistemic state prior to the current situational context and conversational exchange” [Ladd, 1981, 166].

Contextual evidence bias (E): “[...] expectation that p is true (possibly contradicting a prior belief of the speaker) induced by evidence that has just become mutually available to the participants in the current discourse situation” [Buring and Gunlogson, 2000, 7].

Following Domaneschi et al. [2017] whose study will be illustrated in the next section, possible combinations of the original bias of the speaker (where $B(p)$ is positive, $B(-)$ is neutral, and $B(\neg p)$ is negative) and the contextual evidence (where $E(p)$ is positive, $E(-)$ is neutral, and $E(\neg p)$ is negative) were investigated, in order to point out the influence they may have on the choice of PQ forms. This contrast represents, indeed, the conflict existing between the presupposed knowledge of the questioner and the one of the answerer. The experiment carried out by the authors for English and German is considered as an inspiration for the study presented in this work, whose aim is to check whether a pragmatic influence on PQs' syntax also occurs in Italian. Two different hypotheses guided the design of the experiment:

H1 The bias-evidence conflict requires specific superficial PQ forms not only in not only in previously studied Germanic languages but also in the case of a Romance language, and more specifically in Italian, as in this study.

H2 Using a specific PQ form results in improved communication efficiency, as the nature of the conflict can be, therefore, better signalled.

These hypotheses will be proved through the experiment described in the next section.

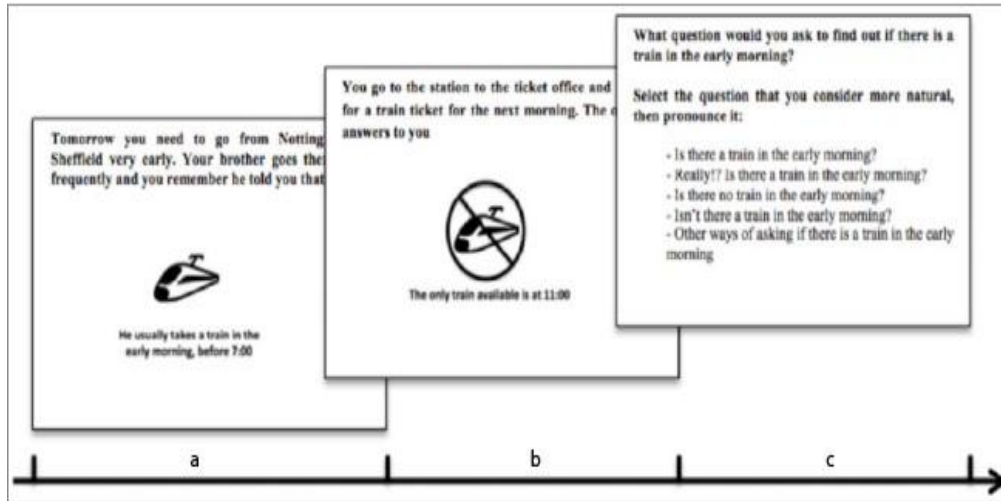


Figure 1. Example of a trial for condition B(p)_E(-p), where ‘a’ builds the original bias, ‘b’ the contextual evidence, and ‘c’ shows the questions the participants had to choose.

Source: Domaneschi et al. [2017]

Table 1. Results for preferred PQ form per pragmatic cell in English and German [Domaneschi et al., 2017]

		<i>Original Bias</i>		
		positive	neutral	negative
<i>Contextual Evidence</i>	positive		PPQ/RPQ	RPQ
	neutral	HNPQ (<i>outer</i>)	PPQ	
	negative	HNPQ (<i>outer/inner</i>)	LNPQ	

3. Polar Express: Experimental Setup

In Domaneschi et al. [2017], the experiment described consisted in a series of scenarios with six different types of conflicts randomly presented to participants. The scenarios presented ordinary fictional conversations, in form of dialogues made up of one or two turns (i.e., two friends preparing dinner, two students looking for the library). Each story was composed of two caption/picture pairs (‘a’ and ‘b’ in Figure 1), followed by the selection of the most appropriate PQ (‘c’ in Figure 1). Participants, therefore, had to choose one and only one appropriate question to pronounce. The choice was among five options: i) positive polar question (PPQ), ii) really-positive polar question (RPQ), iii) low negation polar question (LNPQ), iv) high negation polar question (HNPQ), v) and other.

The first picture (‘a’ in Figure 1, on the left) aims at manipulating the original bias of the speaker; specifically, the utterance *He usually takes a train in the early morning before 7:00* is

meant to generate a bias for the proposition *p*. On the other hand, the second illustration (‘b’ in Figure 1, in the middle) manipulates the bias triggered by the contextual evidence.



Figure 2. Free production task

Original bias: You and a friend of yours are visiting Germany and have decided you want to go eating in a traditional pub. You remember that your friend loves beer (positive); Contextual Evidence: When the waiter comes to take your orders, your friend orders wine. (neutral)

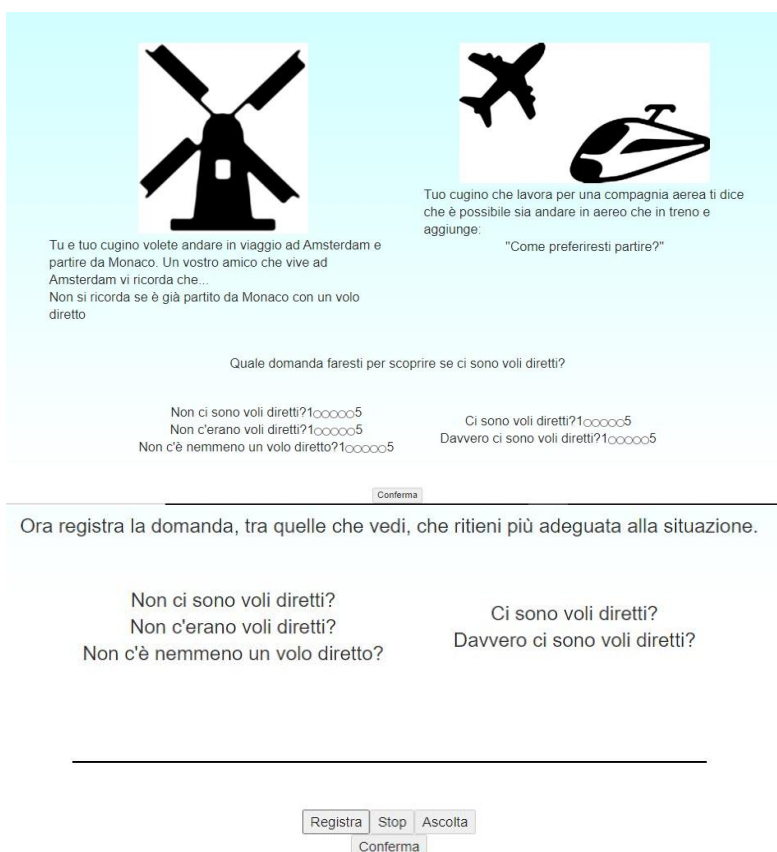


Figure 3. Guided production task

Original bias: You and your cousin want to travel from Munich to Amsterdam. A friend of yours who lives in Amsterdam tells you that he does not remember if there is a direct flight

(neutral); Contextual Evidence: Your cousin who works for a flight company tells you that it is possible to travel both by flight and by train and ask you "How do you want to travel?" (Positive).

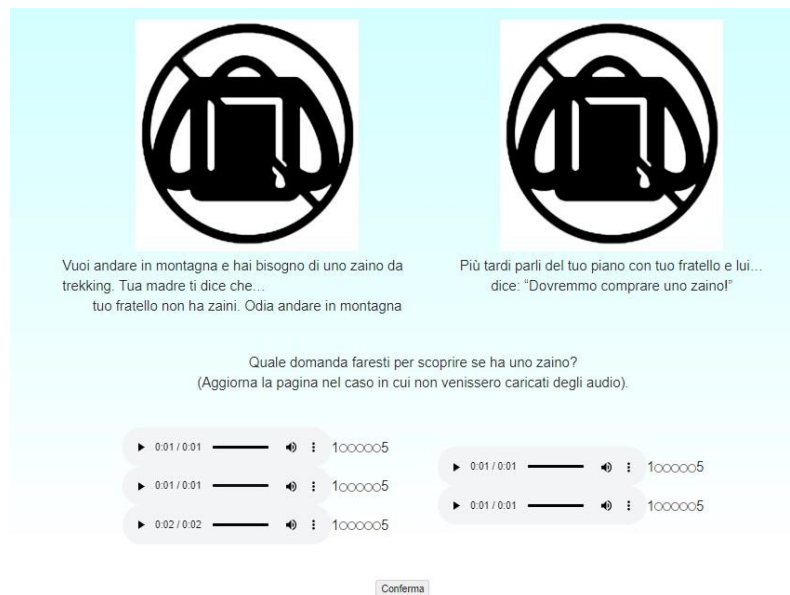


Figure 4. Synthesis scoring task

Original bias: You want to go to the mountains, and you need a hiking backpack. Your mother tells you that your brother does not have any backpack. He hates to go to the mountains (negative); Contextual evidence: Later, you talk with your brother about your plan and he says: "We should buy a backpack" (negative)

The utterance *The only train available is at 11:00* represents a negative evidence of the proposition p . The result of the reference study, in Table 1, shows that both the original bias and the bias derived from the contextual evidence interact in the selection of the appropriate question: in both languages positive PQs are typically selected when there is no original speaker belief and positive or non-informative contextual evidence is provided; low negation questions (i.e., *Do you not...?*) are most frequently chosen when no original belief meets negative contextual evidence; high negation questions (i.e., *Don't you...?*) are prompted when positive original speaker belief is followed by negative or non-informative contextual evidence; positive questions with *really* are produced most frequently when a negative original bias is combined with positive contextual evidence. Regarding HNPQ, we can distinguish two readings in the column with positive bias and neutral or negative evidence. [Ladd, 1981] referred to *outer negation reading* when the speaker wants to double check p , and *inner negation reading* with which the speaker wants to double check $\neg p$. In the inner reading, negation is part of the proposition being checked, whereas in the outer reading it is not. The two readings can be distinguished by the presence of positive polarity items (i.e., *some, already* or *too*), and negative polarity items (i.e., *any, yet, either*) [Domaneschi et al., 2017].

Starting from these data, an extended version of the experiment was developed to collect similar tendencies in Italian. The collection was carried out online using software specifically

designed to administer the test. The 30 scenarios composing the data collection were selected from the English and German drafts used for the aforementioned experiment carried out in Domaneschi et al. [2017] and translated into Italian. The German data were preferred instead of the English ones, since the syntactic structures used in German were similar to the ones documented for the Italian language, specifically as far as the distinction between inner and outer reading for the high negation PQs. This will be specified in detail in section 4.2. The pragmatic situations driven by the combination of original bias and contextual evidence which were collected are $B(p)_E(p)$, $B(p)_E(-)$, $B(p)_E(\neg p)$, $B(-)_E(p)$, $B(-)_E(-)$, $B(-)_E(\neg p)$, $B(\neg p)_E(p)$, $B(\neg p)_E(-)$, $B(\neg p)_E(\neg p)$. Nevertheless, $B(p)_E(p)$, $B(\neg p)_E(-)$, and $B(\neg p)_E(\neg p)$ were left out from the analysis. In fact, as pointed out in Domaneschi et al. [2017], speakers with an original bias for p that receive contextual evidence for p will assume that p is true and will not question further about its truth. Similarly, the same happens for the $B(\neg p)_E(\neg p)$ condition. In Roelofsen et al. [2013], PQs were rated as not natural in the aforementioned conditions. As far as the $B(\neg p)_E(-)$ condition is concerned, the appropriate PQs, as described in Romero and Han [2004] and AnderBois [2011], are a combination of high and low negation. In fact, these two forms also resulted to be frequently selected as appropriate in the present study. Nevertheless, these three conditions were left to future analysis, in order to focus on conditions which were more suitable for the description of conflicting scenarios. An exception is made as far as the first task of this study is concerned (Free Production), for which also other forms were collected, such as wh-questions, which might be considered as more appropriate than PQs in these conditions. Further details are given in section 4.

The target subjects were limited to the Campania region (Southern Italy), in order to avoid the diatopic variation to influence the choice. In fact, to control the regional variety and to ensure the gender balance, each participant had to firstly answer a sociolinguistic questionnaire, concerning age, gender, geographical origin, other places where they lived more than 12 months, and other spoken languages. To ensure that each possible bias-evidence combination for each task occurred, 81 participants were needed. The resulting sample comprises 42 females, 39 males, with an average age of 32,37. Each participant was provided with 10 different scenarios. For each of them, they were asked to perform one, randomly selected, of the three different planned tasks. In fact, contrary to what established in Domaneschi et al. [2017], three different tasks were here randomly shown. Furthermore, for two of the three tasks, instead of asking them to just select one form, as in Domaneschi et al. [2017], they could evaluate questions' appropriateness reflecting the natural tendency of speakers to use more than one form to express the same function. The tasks are described in detail below:

- i Free Production (FP): participants were asked to spontaneously record a question in order to acquire a specific piece of information for that particular situation (Figure 2). This additional type of task is useful to collect information concerning the spontaneous choice of question types depending on pragmatic needs. Furthermore, the intonational patterns that could be extracted from such spontaneous choices can be adopted for the definition of prosody-pragmatics interface schema.
- ii Guided Production (GP): participants were provided with a set of different written forms of PQ, for each of which they must give a score from 1 to 5, according to their

appropriateness in that determined situation (where 1 corresponds to a question completely inappropriate and 5 to completely appropriate). Once having rated the questions, participants also had to record the one they considered to be the most appropriate (Figure 3). In this way, the spoken production of the selected questions is also collected.

- iii Synthesis Scoring (SS): five synthesised PQs were reproduced, for each of which the participants have to give a score from 1 to 5, according to their appropriateness (Figure 4). The questions were synthesised via neural text-to-speech services provided by Microsoft (Note 2), whose intonation is based on statistical patterns extracted from training data. This is important considering the lack of described intonational schema for bias-evidence contrast in Italian PQs. In fact, the selected patterns are here considered as a starting point with the aim of understanding if some frequent patterns are generally adequate to express a particular type of conflict.

For GP and SS tasks, the question forms provided were based on the ones selected in Domaneschi et al. [2017]. Five stimuli were therefore presented, as in the following examples:

HNPQ: *Non ci sono voli diretti?* - ‘Aren’t there direct flights?’

HNPQ_P: *Non c'erano voli diretti?* - ‘Weren’t there direct flights?’

LNPQ: *Non c'è nemmeno un volo diretto?* - ‘Are there no direct flights?’

PPQ: *Ci sono voli diretti?* - ‘Are there direct flights?’

RPQ: *Davvero ci sono voli diretti?* - ‘Are there really direct flights?’

Contrary to the previous experiment, the option *other* was left out, as the participants had the possibility to assign low scores to all the proposed items, if none was considered appropriate. Since no stimulus is considered appropriate in a situation, others might be a better option for the user. Furthermore, the possibility to consider other syntactic forms rather than PQs as appropriate in some situations was also inferred by the FP task. The option *other* was, therefore, substituted with a HNPQ in the past tense. This choice lied on empirical considerations. In fact, this form seems to be more frequently adopted and seems to convey a stronger degree of the speakers’ bias. Note that in Domaneschi et al. [2017], changes in tense, word order, and addition of particles were ignored if they did not affect the biases at issue.

4. Analysis and Results

In this section, the data gathered during the experiment are presented and analysed for each of the tasks carried out.

4.1 Free Production

The FP task was aimed at collecting spontaneous productions from the participants. They were, therefore, asked to record the most appropriate question in the presented situations without giving them possible options among which to choose. As reported in Figure 5, HNPQs and HNPQ_Ps were more frequently chosen in $B(p)_E(\neg p)$ and $B(-)_E(\neg p)$ situations. On the other

hand, LNPQs were also more frequently selected in $B(-)E(\neg p)$ situations, but in smaller numbers compared to HNPQs and HNPQ_Ps. HNPQs were also frequent in $B(-)E(-)$ situations but not as much as PPQs. In fact, PPQs, for their versatility, were produced in $B(p)E(-)$, $B(-)E(p)$, $B(-)E(-)$, and $B(\neg p)E(p)$ situations. RPQs were produced exclusively in $B(-)E(p)$ and $B(\neg p)E(p)$ situations, as in Domaneschi et al. [2017].

Since participants were free to record any stimulus, they considered appropriate, wh-questions were also produced. Interestingly, these forms mostly appear in pragmatic conditions, where the speaker has no original bias against positive 1, neutral 2, or negative evidence 3. One possible interpretation for this choice can refer to the fact that in some cases the bias had a major impact on the speakers, bringing them to collect additional information in case of lack of knowledge.

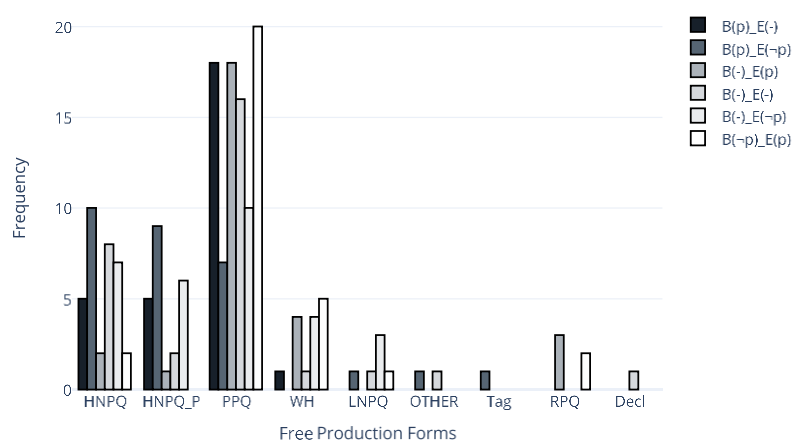


Figure 5. Free production results

On the other hand, the frequent selection of wh-questions in $B(\neg p)E(p)$ scenarios might be due to a major impact of the evidence on the speaker. In fact, instead of asking confirmation with an epistemic adverb like *really*, as expected, speakers might rely on the negative evidence and inquire more about it. Another interesting results is given by the use of such questions in $B(p)E(p)$ and $E(\neg p)E(\neg p)$ situations, which were left out in the resulting analysis. This choice can explain the alleged inappropriateness of PQs in those scenarios.

- (1) a. *Come faccio a fare la tessera?*
b. ‘How can I get the badge?’
- (2) a. *A quale negozio stai pensando?*
b. ‘What shop are you thinking about?’
- (3) a. *Quanto è lontano il supermercato?*
b. ‘How far is the supermarket?’

Furthermore, the standard PQ forms considered in the other tasks of the experiment were in few cases also enriched with other linguistic markers used to convey different degrees of bias,

as shown in Table 2. In fact, as also reported in Mal á[2007], there are different types of bias which are linked to their illocutionary force. Specifically, we can differentiate between:

i) epistemic bias, reflecting what the speaker thinks, expects, or knows the right answer is; ii) deontic bias, reflecting what the speaker judges the right answer ought to be; iii) desiderative bias, what the speaker wants the right answer to be. For example, it is interesting to point out that HNPQs, especially in the past tense, which are mostly used in $B(p)_E(\neg p)$ situations, can be preceded by the adversative conjunction marker *ma* (En. *but*). This marker is, indeed, used to *question the correctness of a new, adversative or contrasting referent, circumstance, or situation* [Metslang et al., 2017]. Facing this contrasting contextual evidence, the speaker needs, therefore, to strongly express its *hope*, as defined in Mal á[2007], toward the correctness of their presupposition. In this case, the conjunction is used to express an epistemic bias. Interestingly, the strength of this epistemic marker is used exclusively in combination with HNPQ_Ps (44% of the HNPQ_Ps were preceded by *ma* ‘but’) whose adequateness in $B(p)_E(\neg p)$ was proved to be unquestionable, as also shown in the next tasks of this experiment.

On the other hand, the adversative conjunction is less frequently used with HNPQs and not used at all with PPQs. These forms were, conversely, sometimes used with other types of epistemic markers. These can be described as part of what is called ‘epistemic modality’. Epistemic modality refers to a conjecture about the truth value of a proposition [Metslang et al., 2017]. This is used in questions expressing a supposition interpretable either as a statement or a question depending on the epistemic status of the speaker and the listener [Metslang et al., 2017]. For example, in 4, the marker *forse* (En. *maybe*) is used in combination with a PPQs in a $B(-)_E(p)$ condition to express an epistemic possibility. In 5, on the other hand, an epistemic expression introducing the HNPQ is used to express doubts towards the given evidence. Moreover, PPQs were frequently used in combination with the causal conjunction *quindi* (En. *so*), as in example 6. As also described for the Spanish language [Gómez, 1993], this conjunction is used with the conversational role corresponding to confirmation request. In fact, PPQs of this type were mostly used when this function was needed ($B(-)_E(p)$, $B(p)_E(p)$, and $B(\neg p)_E(p)$). PPQ_implicit, on the other hand, refers to PPQs which were preceded by other phrasal expressions, as in 7, where the pragmatic function is of information-seeking.

Table 2. Percentage of PQs with and without epistemic markers in the FP task

Label	Occurrence
HNPQ	80,39%
HNPQ_ma	1,92%
HNPQ_sicuro	7,69%
HNPQ_possibile	1,92%
HNPQ_quindi	7,69%
HNPQ_vero	1,92
Label	Occurrence

HNPQ_P	55.56%
HNPQ_P_ma	44,44%
Label	
PPQ	81,39%
PPQ_P	0,77%
PPQ_quindi	9,30%
PPQ_perché	0,77%
PPQ_implicit	3,87%
PPQ_sicuro	1,55%
PPQ_ancora	0,77%
PPQ_forse	1,55%

- (4) a. *Forse hai l'assicurazione?*
 b. 'Do you maybe have an insurance?'
- (5) a. *Sei sicuro che non c'è un negozio di elettronica?*
 b. 'Are you sure there isn't an electronic store?'
- (6) a. *Quindi c'è una biblioteca universitaria?*
 b. 'So is there a university library?'
- (7) a. *Sai se c'è un ristorante?*
 b. 'Do you know if there is a restaurant?'

These alternative forms, representing a lower percentage of participants' choices, were not deepened in this paper.

4.2 Guided Production

As far as the guided production task is concerned, the data analysis regards on the one hand the scores and on the other the selection of one of the items to be pronounced. The results representing the speakers' tendencies in evaluating the appropriateness of specific question forms according to the type of conflict are summarised in Figure 6. Here, the percentages of the highest scores for each question type in each conflict situation are shown. The statistical analysis was carried out with R [Stowell, 2014]. The data were firstly analysed with the Shapiro-Wilk normality test [Shapiro and Wilk, 1965] to check distributional assumptions. In all combinations of bias and evidence, at least one form had a non-normal distribution of the scores, so non-parametric tests were used. To compare the mean values of the distributions, the [Kruskal and Wallis, 1952] test was used to check the existence of significant differences. In all cases, the test indicated that at least one significant difference was present; these were further detailed using the pair-wise [Wilcoxon, 1992] test. The H_0 states that there is no statistically significant difference among the average values of the analysed distributions. More

specifically, the probability that the observed difference is due to chance is endorsed in the H_0 . The rejection of the H_0 would, therefore, mean that the difference is statistically significant. The practical interpretation in this study would be the preference for one question form in each situation. Conflict-related results are going to be described and discussed in detail in the next sections.

Positive Bias vs. Neutral Evidence For the $B(p)_E(-)$ conflicts, PPQs, HNPQs, and HNPQ_Ps show the highest scores (Figure 6), where in Domaneschi et al. [2017] HNPQs were selected. The data presented in Figure 7 and Table 3 confirm this tendency with respect to LNPQs and RPQs, as they are not perceived as appropriate in this situation: they are chosen less frequently in a statistically significant way when compared with PPQs, HNPQs, and HNPQ_Ps.

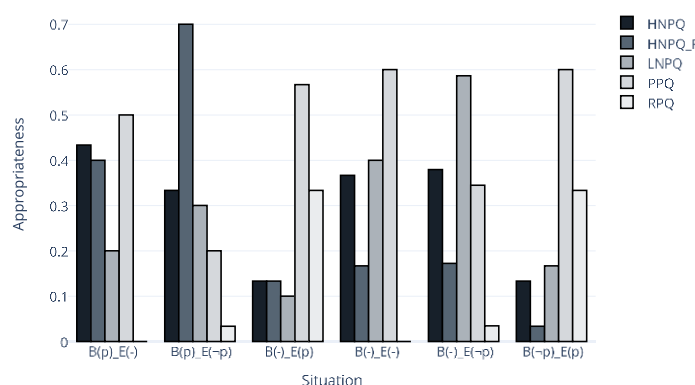


Figure 6. Percentage of highest scores for each type of PQs in different situations

Differently from Domaneschi et al. [2017], PPQs appear to be a valid choice, since no statistically significant difference is found between the three question types. This can be explained by the fact that, according to the way the question is pronounced, PPQs can actually be preferred, because they can show the same pragmatic function and, at the same time, do not damage the *face* [Goffman, 1967] of the interlocutor. In fact, the explicit reference to the conflict through the use of a negation can represent a threat, especially in a situation where the evidence is perceived to be not strong enough (i.e., *neutral*).

Positive Bias vs. Negative Evidence For the conflict arising from a strong presupposition and evidence contradicting it, HNPQ_Ps are scored as more appropriate (Figure 6). The Box plot in Figure 7 and the Table 3 show that this tendency has strong statistical significance when its appropriateness is compared with that of PPQ and RPQ. Conversely, significance is lost when compared to LNPQ ($p = 0.08$). Interestingly, this conflict type was defined as the *ambiguity cell* in Domaneschi et al. [2017], as far as the English data were concerned. This ambiguity derives from the fact that, in English, HNPQs can have an inner or outer reading. The difference between inner and outer HNPQs depends on the polarity of the proposition being checked. In fact, in the inner reading, the negation is part of the proposition being checked (question about a negative proposition), whereas in the outer reading it is not (question about an affirmative proposition [Ladd, 1981]). This means that, with the outer reading the original belief p is double-checked (i.e., *Isn't there some good restaurant around here?*), whereas with the inner reading the opposite proposition ($\neg p$) is double-checked (i.e., *Isn't there any good*

restaurant around here?). The English data in Domaneschi et al. [2017] show that for the $B(p)_E(\neg p)$ condition both inner and outer readings are possible. In German, the difference between HNPQ and LNPQ in this situation is lower, since the pragmatic meanings of inner HNPQs and LNPQs are similar. In fact, in German and in Italian, inner HNPQs have the same form as LNPQs, and both readings are possible. This can be the explanation for a lack of a statistically significant difference in the HNPQ/LNPQ situation ($p = 0.35$) and in the HNPQ_P/LNPQ situation ($p = 0.08$), for Italian. Furthermore, although HNPQ_Ps are preferred in $B(p)_E(\neg p)$ conditions, the difference between the past tense and present tense in the negation does not lead to a strong refutation of the H_0 ($p = 0.35$). This confirms what has been described in Domaneschi et al. [2017], where the high negation was preferred with a percentage of 67%, although the authors did not take into account the tense. This tendency can also be confirmed by other studies which interpret the use of HNPQs to express denegation speech acts, for which the conflict bias/evidence is strongly express [Krifka, 2017].

Table 3. Statistically significant differences in different pragmatic situations. No significance is marked with x ($p > 0.05$); weak significance is marked with * ($0.01 < p < 0.05$); strong significance is marked with ** ($p < 0.01$)

		HNPQ	HNPQ_P	LNPQ	PPQ
$B(p)_E(-)$	HNPQ_P	x	-	-	-
	LNPQ	**	x	-	-
	PPQ	x	x	*	-
	RPQ	**	**	**	**
$B(p)_E(\neg p)$	HNPQ_P	x	-	-	-
	LNPQ	x	x	-	-
	PPQ	x	**	x	-
	RPQ	**	**	**	**
$B(-)_E(p)$	HNPQ_P	x	-	-	-
	LNPQ	x	x	-	-
	PPQ	*	**	**	-
	RPQ	x	x	x	*
$B(-)_E(-)$	HNPQ_P	*	-	-	-
	LNPQ	x	x	-	-
	PPQ	x	*	x	-
	RPQ	**	**	**	**
$B(-)_E(\neg p)$	HNPQ_P	*	-	-	-
	LNPQ	x	** -	-	-
	PPQ	x	x	*	-
	RPQ	**	**	**	**
$B(\neg p)_E(p)$	HNPQ_P	x	-	-	-
	LNPQ	x	x	-	-

PPQ	**	**	**	-
RPQ	X	**	X	X

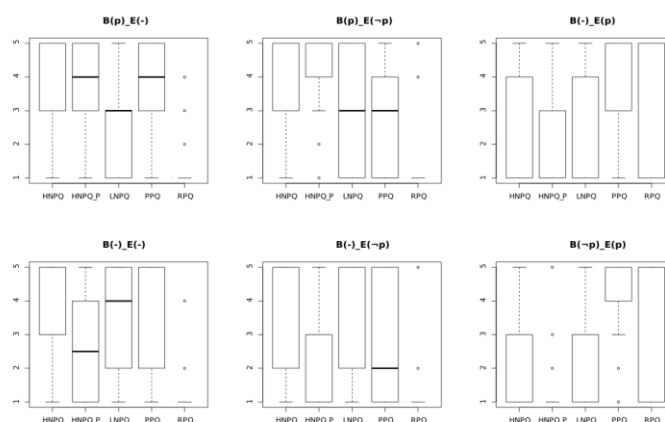


Figure 7. Boxplots showing scores for PQs forms in different pragmatic conditions

Neutral Bias vs. Positive Evidence In situations where there is no original bias combined with positive evidence, PPQs are considered to be more appropriate (Figure 6), as also demonstrated in Domaneschi et al. [2017]. In fact, their appropriateness is statistically significant when compared with that of the negative PQs (Figure 7; Table 3). The statistically significant difference with RPQs is, instead, lower ($p = 0.03$). In English and German, a similar, but slightly stronger, tendency was noted Domaneschi et al. [2017]. In fact, the preposed *really* was supposedly interpreted as a discourse particle with the function of expressing interest and engagement and not as an epistemic adverb asking for confirmation about the proposition, as expected for the negative-positive scenario.

Neutral Bias vs. Neutral Evidence When neither original bias nor contextual bias are displayed, PPQs are preferred around 60% of the time, as in English and German [Domaneschi et al., 2017]. A weak statistically significant difference is shown when PPQs are compared with HNPQ_P ($p = 0.02$) as shown in Figure 7 and Table 3. No statistically significant difference, instead, occurred between PPQs and HNPQs/LNPQs ($p = 0.9$). In fact, as hypothesised in Domaneschi et al. [2017], HNPQs can be used in this situation when only the contextual evidence is considered, whereas LNPQs are selected when only the original bias is considered.

Neutral Bias vs. Negative Evidence In $B(-)_E(-p)$ conflicts, LNPQs are preferred as for English and German [Domaneschi et al., 2017], with a statistically significant difference detected only when compared to HNPQ_Ps (Figure 7; Table 3). The comparison with the negative PQs follows the same explanation reported for the previous conflict. Furthermore, this scenario was also problematic, as the mention of the p-proposition to question about was perceived as unexpected for the participants because it was already negated by the evidence.

Negative Bias vs. Positive Evidence Contrary to what was expected and discussed in Domaneschi et al. [2017], in this conflict scenario, PPQs were considered to be more appropriate than RPQs (around 60%). As reported in Figure 7 and Table 3, PPQs are preferred with statistically significant difference with respect to HNPQs, HNPQ_Ps and LNPQs. There

is no statistically significant difference with RPQs, as the preposed *really* was supposedly interpreted as an epistemic adverb with a confirmation function, as expected. One possible explanation of the highest preference for PPQs can be found in their production. In fact, both RPQs and PPQs produced with an accent on finite verbs can be used with a negative original bias for confirmation purposes Asher and Reese [2005].

In the second part of the GP task, participants were asked to choose only one of the options to be recorded. Almost all the tendencies that were reported for the first part were confirmed, as shown in Figure 8. Only for the $B(p)_E(-)$ and the $B(\neg p)_E(p)$ situations the tendencies are slightly different. In the scoring part of the experiment, for the $B(p)_E(-)$ situation the PPQs were rated as the most appropriate, although there was no statistically significant difference from HNPQs and HNPQ_Ps. Here, HNPQs are more frequently chosen and PPQs are chosen right after them, as in Domaneschi et al. [2017]. Similarly as in Domaneschi et al. [2017], for the $B(\neg p)_E(p)$ situation, the RPQs were more frequently selected, where in the first part the PPQs were rated as more frequently as more appropriate. In conclusion, it can be stated that positive PQs are considered to be more versatile and generally more appropriate in non-conflicting scenarios, whereas negative PQs - high, low, or in the past tense - are more appropriate when different kinds of conflict occur in the contextual evidence.

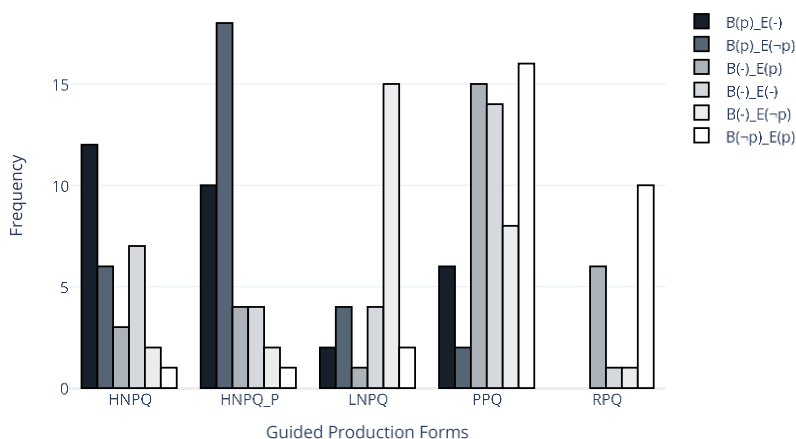


Figure 8. Guided production results – selection

4.3 Synthesis Scoring

Regarding the synthesis scoring task, the data analysis is concerned with the scores participants gave to each one of the given options. The results representing the speakers' tendencies in evaluating the appropriateness of specific question intonational forms according to the type of conflict are summarised in Figure 9. As in the GP task, HNPQ_P collected higher scores in $B(p)_E(\neg p)$ scenarios. The same form bet the others in the $B(p)_E(-)$ scenario, where in the GP task the PPQ had the highest score. This could be explained by the fact that the written form can be interpreted differently, whereas the synthesised forms have fewer perceived possible interpretations. The PPQ is generally preferred in $B(-)_E(p)$ and $B(-)_E(-)$ scenarios, whereas the LNPQ is preferred in $B(-)_E(\neg p)$ scenarios, similarly to the GP task. Differently from the previous task, the RPQ is here preferred in $B(\neg p)_E(p)$ conflicts. The collection of such results will be used as a term of comparison for the productions that seem to be far from the standard.

In fact, we are not yet able to tell which intonational patterns are typical of specific PQs in Italian. One possible future application for these results can be, in fact, the analysis of the deviating forms which could have been chosen by the participants to communicate specific pragmatic meanings.

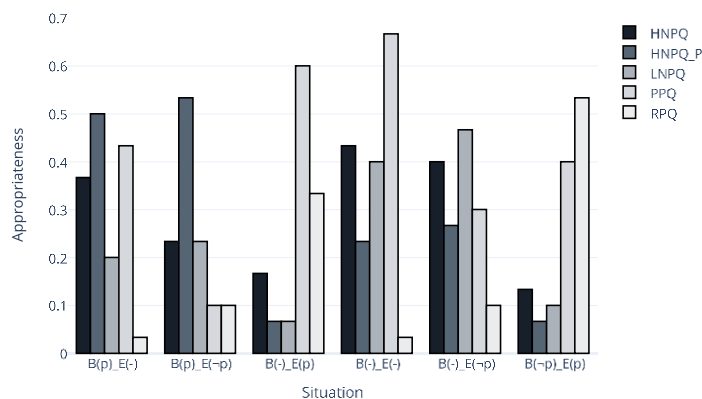


Figure 9. Synthesis scoring results

5. Conclusion

The experiment presented in this paper was aimed at testing whether specific forms of PQs were perceived as more appropriate in specific pragmatic scenarios. The experiment was built upon the one carried out in Domaneschi et al. [2017], where scenarios representing different bias-evidence combinations were presented to participants who had to choose the most appropriate question among the ones suggested. In this study, the experiment was, conversely, subdivided in three tasks: the first one (FP) left the participants free to pronounce whatever form they considered to be appropriate to express that particular pragmatic function; the second one (GP) provided the participants with a set of different forms for which they had to give a score of appropriateness; the third one (SS) provided, as the previous one, the participants with a set of synthesised forms to be given a score. In general, the combination of the three tasks of this experiment resulted in the confirmation of the tendencies reported in Domaneschi et al. [2017]. Therefore, the H1, concerned with proving whether specific forms of PQ were typically used in particular pragmatic scenarios in Italian as well as in German and English, was confirmed. Nevertheless, differently from Domaneschi et al. [2017], the differences resulted to be less sharp, as different forms have similar scores in similar scenarios. This result depends on the annotation protocol which allowed the subjects to express themselves in greater detail, enabling to capture different combinations of pragmatic function and syntactic structure. Specifically, the study shows a clear tendency for preferring HNPQs in the past tense when a positive bias clashes with negative contextual evidence. Interestingly, although the PPQ is generally the preferred form in most of the situations for its versatility, in $B(p)_E(-p)$ scenarios the percentage of scores is lower than in the others. This result leads to the preliminary conclusion that in such situations the adoption of a NPQ better suits the pragmatic needs, increasing the communication efficiency (H2). This result is particularly important when considering application scenarios where common ground inconsistency can occur and lead to understanding problems. This is the case of human-machine interaction, for which the adoption

of the appropriate form of question can better highlight the nature of the conflict in order to recover it. Further investigation will be conducted in this direction. Specifically, we will investigate whether the use of such a form could also bring to better common ground inconsistencies recovery in human-machine interaction. Further investigations will also focus on prosodic analysis which will enable a more detailed comparison of the results collected from the three tasks. In this way, the coherence of the linguistic choices in free, guided and synthesised scenarios will be tested.

Acknowledgments

The authors would like to thank Martina Di Bratto who not only helped during the data collection but who also supported our work with her suggestions and precious remarks. Work funded by the Italian National Project PRIN “Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)” (#B52F15000450001).

References

- AnderBois, S. (2011). *Issues and alternatives*. University of California, Santa Cruz.
- Arnhold, A., Braun, B., & Romero, M. (2021). Aren't prosody and syntax marking bias in questions? *Language and Speech*, 64(1), 141-180. <https://doi.org/10.1177/0023830920914315>
- Asher, N., & Reese, B. (2005). Negative bias in polar questions. In *Proceedings of Sinn und Bedeutung* (Vol. 9, pp. 30-43).
- Bolinger, D. (1978). Yes-no questions are not alternative questions. In *Questions* (pp. 87-105). Springer. https://doi.org/10.1007/978-94-009-9509-3_3
- Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction* (pp. 56-311). Cambridge University Press.
- Buring, D., & Gunlogson, C. (2000). *Aren't positive and negative polar questions the same?*
- Couper-Kuhlen, E. (2012). Some truths and untruths about final intonation in conversational questions. In *Questions: Formal, functional and interactional perspectives*. Cambridge University Press.
- Domaneschi, F., Romero, M., & Braun, B. (2017). Bias in polar questions: Evidence from english and german production experiments. *Glossa: a Journal of General Linguistics*, 2(1). <https://doi.org/10.5334/gjgl.27>
- Dryer, M. S. (2013). Polar questions. In S. D. Matthew, & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology, Leipzig. Retrieved from <https://wals.info/chapter/116>
- Frana, I., & Rawlins, K. (2019). Attitudes in discourse: Italian polar questions and the particle mica. *Semantics and Pragmatics*, 16. <https://doi.org/10.3765/sp.12.16>

- Goffman, E. (1967). On face-work. *Interaction Ritual*, 5-45. <https://doi.org/10.4324/9780203788387-2>
- Gómez, A. B. (1993). Los conectores pragmáticos en español coloquial (i): su papel argumentativo. *Contextos*, (21), 145-188.
- Hayano, K. (2013). Question design in conversation. *The handbook of conversation analysis* (p. 395). John Wiley & Sons. <https://doi.org/10.1002/9781118325001.ch19>
- Heritage, J. (2002). The limits of questioning: Negative interrogatives and hostile question content. *Journal of Pragmatics*, 34(10-11), 1427-1446. [https://doi.org/10.1016/S0378-2166\(02\)00072-3](https://doi.org/10.1016/S0378-2166(02)00072-3)
- Huang, Y. (2017). *The oxford handbook of pragmatics*. Oxford University Press.
- Koshik, I. (2002). A conversation analytic study of yes/no questions which convey reversed polarity assertions. *Journal of Pragmatics*, 34(12), 1851-1877. [https://doi.org/10.1016/S0378-2166\(02\)00057-7](https://doi.org/10.1016/S0378-2166(02)00057-7)
- Koshik, I. (2005). *Beyond rhetorical questions: Assertive questions in everyday interaction* (Vol. 16). John Benjamins Publishing. <https://doi.org/10.1075/sidag.16>
- Krifka, M. (2017). Negated polarity questions as denegations of assertions. In *Contrastiveness in information structure, alternatives and scalar implicatures* (pp. 359-398). Springer. https://doi.org/10.1007/978-3-319-10106-4_18
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. Academic Press.
- Ladd, D. R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In *Papers from the Regional Meeting. Chicago Ling. Soc. Chicago*, 3(17), 164-171.
- Levinson, S. C. (2010). Questions and responses in y é îdnye, the papuan language of rossel island. *Journal of Pragmatics*, 42(10), 2741-2755. <https://doi.org/10.1016/j.pragma.2010.04.009>
- Malá M. (2007). Negative polar questions in English and Czech. *Proceedings of the 4th Corpus Linguistics Conference*.
- Malamud, S., & Stephenson, T. (2015). Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics*, 275-311. <https://doi.org/10.1093/jos/ffu002>
- Metslang, H., Habicht, K., & Pajusalu, K. (2017). Where do polar question markers come from? *STUF-Language Typology and Universals*, 70(3), 489-521. <https://doi.org/10.1515/stuf-2017-0022>

- Orrico, R., & D'Imperio, M. (2020). The interaction of intonation cues and cognitive pragmatic skills in the perception of epistemic bias in Salerno Italian questions. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1). <https://doi.org/10.5334/labphon.238>
- Oshima, D. Y. (2017). Remarks on epistemically biased questions. In *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation* (pp. 169-177).
- Roelofsen, F., Venhuizen, N., & Sassoan, G. W. (2013). Positive and negative questions in discourse. In *Proceedings of sinn und bedeutung* (Vol. 17, pp. 455-472).
- Romero, M., & Chung-Hye, H. (2004). On negative yes/no questions. *Linguistics and Philosophy*, 27(5), 609-658. <https://doi.org/10.1023/B:LING.0000033850.15705.94>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Sidnell, J. (2012). Who knows best? Evidentiality and epistemic asymmetry in conversation. In *Pragmatics and Society* (Vol. 3, pp. 294-320). John Benjamins. <https://doi.org/10.1075/ps.3.2.08sid>
- Sidnell, J., & Stivers, T. (2012). *The handbook of conversation analysis* (Vol. 121). John Wiley & Sons. <https://doi.org/10.1002/9781118325001>
- Stalnaker, R. (1977). Pragmatic presuppositions. In *Proceedings of the Texas conference on per formatives, presuppositions, and implicatures* (pp. 135-148). Arlington, VA: Center for Applied Linguistics.
- Steensig, J., & Drew, P. (2008). *Questioning and affiliation/disaffiliation in interaction: Special Issue of Discourse Processes*. SAGE Publications, 2008. <https://doi.org/10.1177/1461445607085581>
- Stivers, T., & Enfield, N. J. (2010). A coding scheme for question–response sequences in conversation. *Journal of Pragmatics*, 42(10), 2620-2626. <https://doi.org/10.1016/j.pragma.2010.04.002>
- Stowell, S. (2014). *Using R for Statistics*. Apress. Retrieved from <http://www.apress.com/9781484201404>
- Vincent, L. D., & Cassidy, R. S. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, 2(1), 111-128. https://doi.org/10.1162/qss_a_00097
- Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in Statistics* (pp. 196-202). Springer. https://doi.org/10.1007/978-1-4612-4380-9_16

Notes

Note 1. Author Contributions [based on CRediT - <https://casrai.org/credit/>]: Conceptualisation^{abc}; Data curation^{ab}; Formal analysis^{abc}; Funding acquisition^c; Investigation^a; Methodology^{abc}; Resources^{abc}; Supervision^c; Validation^{ab}; Visualisation^a; Writing – original draft^a; Writing – review & editing^{abc}.

Note 2. <https://wals.info/feature/116A2/33.7/153.1> [last consultation on the 4th January 2021]

Note 2. <https://azure.microsoft.com/en-us/services/cognitive-services/text-to-speech/>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)