

**Can you believe that?**  
**The prosody of non-genuine polar questions in English**

Master's thesis

Hauke Lindstädt



Date of submission: 12.03.2024

Examiner: Prof. Dr. Martine Grice

University of Cologne  
Faculty of Arts and Humanities  
Department of Linguistics

# Table of content

Abbreviations .....	III
Chapter 1: Introduction.....	1
Chapter 2: What is a question? .....	3
2.1 Illocutionary force and the theory of speech-acts .....	3
2.2 Statements, mands and questions.....	6
2.3. Questions, interrogatives and inquiry force .....	7
2.4 Sentence type, prototypicality and questionhood .....	12
2.5 What is a non-genuine polar question? .....	15
Chapter 3: The prosody of polar questions.....	16
3.1 Genuine and non-genuine polar questions .....	17
3.1.1 Corpus study I: Banuazizi & Creswell (1999) .....	17
3.1.2 Corpus study II: Hedberg et al. (2017).....	20
3.2 Rhetorical questions as a sub-class of non-genuine questions.....	23
3.3 Contradicting studies.....	27
3.4 Summary .....	29
Chapter 4: The prosody of non-genuine polar questions – an exploratory study.....	30
4.1 Data .....	31
4.1.1 The TV News Archive .....	31
4.1.2 The utterances.....	34
A: Can you repeat that? .....	36
B: Can you believe that?.....	38
C: Do you believe that? .....	39
4.2 Method .....	42
4.2.1 Evaluating, labelling, annotating.....	42

4.2.2 The issue of focus .....	45
4.2.3 Prosodic analysis with periodic energy .....	46
4.2.3.1 Rationale .....	46
4.2.3.2 The measurements .....	48
4.2.3.3 Syllable boundaries.....	50
4.2.3.4 Periograms .....	50
4.3 Results.....	52
4.3.1 Mass.....	52
4.3.2 Synchrony.....	55
4.3.3 Utterance duration .....	60
4.3.4 Summary .....	62
Chapter 5: Conclusion .....	63
References .....	66
Appendix .....	71
Eidesstattliche Erklärung.....	82

## Abbreviations

CoG	Center of gravity
CoM	Center of mass
F0	Fundamental frequency
Hz	Hertz
ms	Millisecond(s)
ToBI	Tones and Break Indices

## Chapter 1: Introduction

In summer 2022 a statement from German chancellor Olaf Scholz caused some indignation. At a press conference he was asked by a journalist: „Mr. chancellor, the G7 have very clearly borne witness to safety guaranties for Ukraine, even after the war. Could you clarify, what kinds of safety guaranties are meant here?”

“Yes, I could”, Olaf Scholz replied. And after a pause he smilingly added: “That’s it” (WELT Nachrichtensender 2022)<sup>1</sup>.

The reason for the medial outrage, that followed, is intuitively comprehensible. But why exactly? It seems like Olaf Scholz answered the question and at the same time he did not. In any case, he didn’t meet the conventionalized expectations of the situation. Many observers reproached him for being impolite or arrogant.

Anticipating, what this study will elaborate on, the mismatch between the journalist’s question and the chancellor’s answer can be explained in the following terms: While the question was non-genuine, i.e. not aiming for an answer in the literal sense, Olaf Scholz – against his better judgement, one can reasonably suspect – treated it as being genuine and acted accordingly. And in doing so he avoided to comply with the question’s actual request, which was to concretely name the mentioned safety guaranties for Ukraine.

This brief example shows, that the matter of questions and its functions in social contexts is far from trivial. Approaching questions from a linguistic perspective starts with a bundle of difficult questions itself: What defines a question? What does it look or sound like? What does it do and what does it mean?

The current study tackles the phenomenon from the field of prosody, a linguistic discipline, that deals with suprasegmental phonetic features of spoken language like melody, duration or intensity. But, as the study’s title suggests, the investigated subject is narrowed down to very specific kinds of questions, namely non-genuine polar questions, like the one from the example above. Without going into detail at this point, polar questions are such questions, that can be answered by either *yes* or *no*. Therefore, they are also referred to frequently as yes-no questions.

The very general question underneath the whole analysis is, if and how functional aspects of human language are reflected in its prosody. In our specific case, this translates

---

<sup>1</sup> The conversation was translated from German to English by the author.

to the following: Do non-genuine polar questions in English exhibit idiosyncratic prosodic characteristics at all? And subsequently: What prosodic characteristics do non-genuine polar questions share and how do different sub-types of non-genuine polar questions differ from each other prosodically.

In order to approach these questions, the study is basically divided into three parts, which are dealt with in the chapters 2, 3 and 4. Chapter 2 starts with asking, what questions are in the first place, and introducing some of the numerous theoretical approaches on that. The two main resources here are the theory of speech-acts with its concept of illocutionary force and the semantic approach to questions by Rodney Huddleston. Step by step the terms *question*, *polar question*, *genuine polar question* and *non-genuine polar question* will be explained to set a profound basis for the further analysis.

Chapter 3 then focuses on previous studies on the prosody of polar questions, with special attention to the distinction between genuine and non-genuine polar questions. The key prosodic feature to be investigated here is utterance-final pitch movement.

Building on that, the conduction and results of an exploratory corpus study will be reported in chapter 4. Here, three kinds of non-genuine polar questions, namely requests for action, rhetorical questions and topic introductions, will be analysed and compared. The speech data is drawn from non-scripted American television programmes and thus can be considered advantageous to experimentally elicited data with regards to spontaneity and naturality. The prosodic analysis is based on the *ProPer* method (Albert et. al 2023), which provides new powerful description tools by taking into account both pitch and periodic energy. The latter can be seen as an acoustic equivalent to the intelligibility of pitch in perception.

As one will see, the endeavour of linking form and function, i.e. in our case the prosodic behaviour and the illocutionary force of polar questions, is a tightrope act between categorizing too strictly and allowing ambiguity to a degree, that fails to explain anything. Some closing remarks on this will be made in the conclusion of chapter 5.

## Chapter 2: What is a question?

Throughout the research history questions have been approached by many linguists from many different perspectives, which leads both to a rich fund of literature and a great degree of terminological confusion. So when speaking about questions, it is important to specify, what is actually meant by *questions*.

As Jan P. de Ruiter (2012) points out, the area of most tension in that matter is that between form and function. While the former refers mostly to syntactic aspects, but especially in phonetic research also implies the prosodic form, the latter refers to meaning and use. Each definition of the term *question* focuses more or less on one of those perspectives or tries to reconcile them in some kind of combination. Importantly, some authors argue, that meaning and use of questions, i.e. their semantics and pragmatics, respectively, should be distinguished just as clearly as form and function (e.g. Huddelston 1994; s. section 2.3).

This chapter provides a broad theoretical overview by introducing and discussing some of the major concepts on questions, and on that basis sets the use of terminology for the following chapters. One important clarification is required right away: Some of the basic ideas on questions claim to be universal, i.e. independent from any specific language. This study, however, is not of typological character, but focusses on English. To avoid undue generalizations, all explanations – unless otherwise indicated – shall thus be understood in reference to the English language. For the sake of simplification, the issue of different varieties of English, will not be dealt with in detail.

### 2.1 Illocutionary force and the theory of speech-acts

Before turning to thoughts on what questions actually are, how they can be distinguished from other linguistic categories or how they can be subdivided into meaningful sub-classes (like non-genuine polar questions), it is necessary to establish the concept of illocutionary force. This term is crucial to many of the following theoretical arguments aswell as for the self-conducted study reported in chapter 4. It is derived from the theory of speech-acts, which was introduced and developed by John. L. Austin (1962) and later continued – among others – by John Searle (1969).

The basic claim of the theory of speech-acts is this: All utterances are speech-acts in the sense, that they have performative character. This gets clearer, when looking at the three kinds of acts, that every speech-act consists of, namely the locutionary, the illocutionary and the perlocutionary act. The locutionary act is the act of saying itself, the act of producing a meaningful utterance, regardless from its specific meaning, intention or effect. The illocutionary act then, is the act, that is performed in the world by saying something. This can be, for example, making a statement, making a promise, asking a question, issuing a command, uttering a threat or christening a ship (examples from Lyons 1977: 730). Finally, the perlocutionary act is the act of causing an effect on others by saying something, like convincing someone of something, persuading someone to do something or consoling someone (examples from Lyons 1977: 730).

While a speech-act can succeed or fail on each of the three levels, the success of the perlocutionary act as opposed to the locutionary and the illocutionary act does not depend on the speaker alone. One simple example can show this. If a speaker asks someone „In which direction is the train station?“, he succeeds in performing the locutionary act just by producing an utterance in a non-defective way. He also succeeds in performing the illocutionary act of asking a question – or more precisely: requesting information, that he needs, from the addressee<sup>2</sup> – by producing the utterance in a way, that conventionally will be understood as exactly that: a request for information. However, whether he succeeds in performing the perlocutionary act, i.e. getting the addressee to actually give away the requested information, is not up to him. There could be several external reasons for failure. Maybe the addressee does not speak English. Maybe the addressee does not know, where the train station is. Or maybe, for whatever reasons, the addressee is not willing to reveal that piece of information. In all those cases the intended perlocutionary effect is not congruent with the actual perlocutionary effect, which consequently leads to a failure of the perlocutionary act.

Having established the term *perlocutionary effect*, it is now time to turn to the concept of illocutionary force, which is related, but must not be confused with it. Basically, the illocutionary force of an utterance describes „its status as a promise, a threat, a request, a

---

<sup>2</sup> This can be seen as a first attempt of defining a question and many researchers in fact stick to some variation of this definition (e.g. Haan 2002). But in accordance with Huddleston (1994) we would call this kind of speech-act an *inquiry*, a term which is not identical to the term *question*. A detailed explanation of this will be given in section 2.3.



statement, an exhortation, etc.“ (Lyons 1977: 731), i.e. its status as an illocutionary act. While the illocutionary force and the perlocutionary effect, in particular the intended perlocutionary effect, are connected in many occasions of everyday language, they should be clearly separated from an analytic point of view. Importantly, the former relates to the utterance itself and the latter relates to the intended effect on the addressee. For example an utterance with the illocutionary force of making a statement typically comes with the intended perlocutionary effect of making the addressee believe the proposition of that statement. In fact, this intention is implied by the illocutionary act itself. But depending on the circumstances, the intended perlocutionary effect might be a different one, e.g. to confuse the addressee. In either case, the conventionalized illocutionary act of making a statement would have been performed, which the speaker wants to be recognized as such by the addressee<sup>3</sup>. Whether the intended perlocutionary effect – be it to convince the addressee of the utterance’s proposition or to confuse him – then matches with the actual perlocutionary effect, i.e. whether the perlocutionary act is successful as a whole, is a different matter.

With regard to the concept of illocutionary force, which is of particular interest here, the following should be noted: The illocutionary force of an utterance describes the kind of thing, that a speaker willingly does in the world by actually uttering it. Is he making a statement? Is he issuing a command? Is he requesting information?

Importantly, this short summary of the theory of speech-acts and the concept of illocutionary force is by no means exhaustive. It is meant only to be as detailed as necessary in order to set the ground for the following analyses of questions. More far-reaching ideas as well as internal controversies of the theory of speech-acts are left out for the sake of brevity. Further explanations will be given, where necessary, namely in section 2.3, when it comes to indirect speech-acts, and in section 2.5, when it comes to sincerity conditions of genuine questions.

---

<sup>3</sup> The recognition by the addressee, that a particular illocutionary act has been performed, is often referred to as *illocutionary uptake* (Lyons 1977: 731).

## 2.2 Statements, mands and questions

In order to approach the complex phenomenon of questions, John Lyons' extensive work on semantics (Lyons 1977) is a good starting point. His classification of questions within the system of human language is strongly based on the just established theory of speech-acts.

Basically, he introduces the paradigm of statements, mands and questions as three major types of utterances, which can be differentiated by their illocutionary force. The coined word *mand* is used here as a generic term for commands, demands, requests, entreaties and the like. While illocutionary acts often depend on "the legal or moral concepts institutionalized in particular societies" (Lyons 1977: 737) – the acts of baptizing a child or swearing an oath in court can exemplify that –, the three "basic illocutionary acts" (Lyons 1977: 737) are considered to be universal. This means concretely, that the acts of making a statement, issuing some kind of mand and asking a question are existent in all human language systems.

It is important to note, that this classification of questions – in a paradigm with and in opposition to statements and mands – is purely functional. It does not say anything per se about formal aspects, i.e. on what questions look like. The formal equivalent to statements, mands and questions would be the paradigm of declaratives, jussives and interrogatives, which refer to types of clauses<sup>4</sup> with particular grammatical structures. The uncommon term *jussive* is used here to avoid the inexact term *imperative*, which specifically refers to the mood of the verb and thus stands in a row with *indicative* and *subjunctive* rather than with *declarative* and *interrogative* (Lyons 1977: 748).

While declaratives, jussives and interrogatives can be described as clauses, that are typically used to produce utterances with the illocutionary force of statements, mands and questions, respectively, there is no obligatory relationship between the two categories. "Can you shut the window?", for example, is an interrogative clause qua its grammatical structure, namely the inversion of the subject *you* and the auxiliary verb *can* (s. section 2.3 for more details). In most contexts, however, it would not function as a question, that is expected to be answered, but as a mand, concretely a request to shut the window. A

---

<sup>4</sup> Lyons (1977) speaks of *sentences*, but I will use the more precise term *clause*. As Huddleston (1994) points out, sentences can hardly be categorized by their type, because they can consist of several clauses of different types. For example, the sentence "Come around six, or is that too early for you?" is a coordination of a jussive and an interrogative clause (Huddleston 1994: 413).

clause like “Ben plays soccer” on the other hand is of declarative type and accordingly can function as a statement. But depending on the context of its utterance and its phonetic realization it might as well function as a question, to which an answer like “Yes” or “No” is both expected and valid. Lyons states: “There is no one-to-one correspondence between grammatical structure, in particular, and illocutionary force; but we cannot employ just any kind of sentence [cmp. footnote 4] in order to perform any kind of illocutionary act” (Lyons 1977: 733).

His explanations of questions and their classification within the theory of speech-acts go much deeper. But for the purposes of this study it is most important to understand the two related, but independent tripartite distinctions – between statement, mand and question on the one hand, and between declarative, jussive and interrogative on the other hand.

### 2.3. Questions, interrogatives and inquiry force

Moving on with the question of questions, a text by Rodney Huddleston (1994) is particularly illuminating. He resumes Lyons’ (1977) ideas in emphasizing and elaborating a sharp distinction between interrogatives and questions as members of different linguistic categories. And while the former refers to the syntactic form, the latter – and here Huddleston contradicts Lyons – refers to properties of meaning rather than properties of pragmatic function. So at this point Lyons’ distinction between form and function is refined into a distinction between form, function and meaning, which stand for the linguistic perspectives of syntax, pragmatics and semantics, respectively.

Just like Lyons (1977) Huddleston (1994) describes interrogatives as members of a clause type, which contrasts with other clause types like declarative and jussive<sup>5</sup>. But before turning to some relevant syntactic details of interrogatives, Huddleston’s striking and rather non-intuitive definition of questions should be introduced. He claims: “[T]he distinctive property of a question is that it defines a set of answers” (Huddleston 1994: 413). This means, that a question should not be looked at isolated, but must always be analysed in the light of its possible answers. And the fact, that there are possible answers

---

<sup>5</sup> Huddleston uses the term *imperative*. But following the suggestion from Lyons (1977) as explained in section 2.2, the less ambiguous term *jussive* will be used throughout this paper, whenever the corresponding clause type is referred to.

to it, makes it a question in the first place. It is not the correct or true answers, that are of relevance here, but all possible answers from a semantic point of view.

Importantly, one has to differentiate answers from responses. While (1b) is both a response and an answer to the question in (1a), (1c) is a valid response, but no answer semantically. (1d), to give one further example, is a response containing an answer (“No”) plus additional information (Huddleston 1994: 413).

- (1) (a) Has she cleaned the carpet?
- (b) Yes, she’s cleaned the carpet
- (c) I don’t know.
- (d) No, she didn’t have the time.

Deriving from Huddleston’s general definition of questions, three major types of questions can be identified, all of which define a different set of answers. These are polar questions, alternative questions and variable questions.

Polar questions, which are also known as yes-no questions (e.g. in the terminology of Sadock & Zwicky 1985 or Lyons 1977), always define a set of two possible answers, *yes* and *no*, respectively. The *yes* answer has the same propositional content as the polar question, the *no* answer has the opposite propositional content. In Huddleston’s example of a typical polar question, “Are you ready?”, affirmative responses like “I am ready”, “Yes, I am” or “Yes” share their propositional content with the question and semantically count as the same answer (Huddleston 1994: 416). Accordingly, versions of the negative answer would be “I am not ready”, “No, I am not” or “No”.

Alternative questions, which might be known by the rather inexact term *either/or questions*, define a finite set of two or more answers, all of which are given as alternatives in the question itself. An example of an alternative question would be the following: “Do you prefer tea, coffee or beer?”

The third type of question, the variable question, is also known as wh-question, x-question (Lyons 1977) or information question (Sadock & Zwicky 1985). Its distinctive feature is, that it contains a variable, for which every answer substitutes a particular value (Huddleston 1994: 416). This leads to an indefinite set of possible answers. A valid answer to the variable question “Who ate the cake?” would be every answer that follows

the pattern *X (ate the cake)*, e.g. “John ate the cake”, “Tina ate the cake”, “The cat ate the cake” and so on.

Having established Huddleston’s semantic definition of questions, some things should be said about interrogative clause type, a “category of grammatical form” (Huddleston 1994: 411), and how it is intertwined with questions and question types. Even though there is no one-to-one correlation between interrogatives and questions, the former are typically used to express the latter. And while questions – as seen – can be categorized into three major types, interrogatives can be subdivided into two classes, namely closed and open interrogatives (Huddleston 1994: 419). Examples for these are shown in (2) and (3), respectively.

- (2) (a) Do you know John?  
(b) Are her shoes black or brown?

- (3) (a) Where is Tina?  
(b) Who plays piano?

In English, closed interrogatives are marked by subject-auxiliary inversion, when occurring as main clauses. This means that – in opposition to the declarative clause type – the inflected auxiliary verb is placed before the subject. This can be seen in the closed interrogatives in (2a) and (2b), where the auxiliaries *do* and *are*<sup>6</sup>, are placed before the subjects *you* and *her shoes*, respectively.

Open interrogatives on the other hand are marked by interrogative words like *who*, *what*, *where* etc. And when occurring as main clauses, open interrogatives are marked by subject-auxiliary inversion, too, as in (3a) – unless the interrogative word is identical to the subject as in (3b)<sup>7</sup>.

The names closed and open interrogatives derive from the fact, that they are typically used to express questions with a closed and open set of answers, respectively. Thus, closed interrogatives typically express polar questions (2a) or alternative questions (2b), while

---

<sup>6</sup> Huddleston contrasts the term *auxiliary verb* with *lexical verb*, not with *main verb*. Thus, the copula verbs in (2b) and (3a) are considered auxiliaries, too. I will take over this terminology.

<sup>7</sup> The subject-auxiliary inversion in open interrogatives is somewhat different to the one in closed interrogatives – Huddleston (1994) distinguishes between triggered and spontaneous inversion –, but a detailed analysis would exceed the scope of this study.

open interrogatives typically express variable questions (3a, 3b) (Huddleston 1994: 419). Again, one must bear in mind, that questions do not necessarily correspond with interrogatives. One classic example is given by so called echo questions.

(4) (a) She's a genius.

(b) She's a genius?

(c) She's a what?

(Examples from Huddleston (1994: 427))

All clauses in (4) are of declarative type. But while (4a) represents a statement, (4b) represents the corresponding polar echo question, which defines a set of two answers, namely *yes* and *no*. (4c) on the other hand represents a corresponding variable echo question, which defines a set of all answers, that substitute a value for the question word *what*. Just as a side note: In echo questions one might find the special case of a variable question in the form of a closed interrogative – for example when the initial question “Did Kim answer the phone?” is responded by the echo question “Did who answer the phone?” (Huddleston 1994: 427). Another example, where the syntactic and semantic categories do not properly match, is the following: “Did you come in and I was still asleep?” (taken from Huddleston 1994: 426). Instead of a single interrogative, this question is expressed by the coordination of a closed interrogative and a declarative.

So while questions are not exclusively expressed by interrogatives, cases of interrogatives, that do not express questions, are rarer. One of them is this: “May the best player win!” (taken from Huddleston 1994: 424). This clause does not define a set of answers and therefore does not express a question. Syntactically, however, it bears the characteristics of a closed interrogative, most importantly the subject-auxiliary inversion.

With the purpose of this study in mind, only unembedded questions and main clause interrogatives have been and will be discussed. Nevertheless, it should be briefly mentioned, that Huddleston (1994) analyses both subordinated interrogatives and embedded questions – an endeavour, which leads to an even richer understanding of syntax, semantics and pragmatics of questions.

Finally, one crucial aspect of Huddleston's terminology should be explained in detail. As already mentioned, for him, questions are defined by the meaning of their

propositional content. This meaning is not to be confused with the function, that questions fulfil, when used in conversation. Thus, for Huddleston, questions do not share a particular illocutionary force – as it is understood from the reading of Lyons (1977).

Regarding the functional aspects of questions, Huddleston claim, that they are “prototypically used with the aim of discovering the answer from the addressee” (Huddleston 1994: 414). He calls this kind of questions *inquiries*, but stresses, that *inquiry force* is by no means a necessary condition for questions.

(5) (a) Can you leave me alone?

(b) Isn't he gorgeous?

Issuing a request for action as in (5a) or making an exclamatory statement as in (5b) are only two among many other functions, that questions can fulfil. Both (5a) and (5b) are closed interrogatives in terms of syntax and polar questions in terms of semantics. But although they define the binary set of answers (*yes* and *no*) each, an actual answer would hardly be expected from the addressee. They are examples for so called indirect speech-acts, i.e. speech-acts, “where the propositional content expressed diverges from the propositional content of the speech[-]act conveyed” (Huddleston 1994: 415).

What Huddleston calls inquiries is labelled as genuine questions by Banuazizi & Creswell (1999). Questions, which do not function as inquiries like the indirect speech-acts of (5a) and (5b) can thus be called non-genuine questions. A more detailed discussion of the distinction between genuine and non-genuine questions will be given in section 2.5.

In conclusion, two major points of Huddleston's paper should be emphasized. First, questions, as a category of meaning, are to be distinguished from interrogatives, as a category of grammatical form. And second, whereas typically used as inquiries, questions can serve a whole range of different functions, and can bear just as many different types of illocutionary force.

## 2.4 Sentence type, prototypicality and questionhood

Huddleston's (1994) rather clear-cut definition of questions as a semantic category, which is both distinct from their syntactic form and their pragmatic use in particular conversational contexts, is very useful to work with. Therefore, this study will refer to questions mostly in his terminology (s. section 2.5).

It is important, however, that there are other analyses of the phenomenon, which are fruitful and enlightening in their own way. In order to understand a complex system like spoken language, it might be helpful to define separate categories and conscientiously disentangle them from each other, like Lyons (1977) and Huddleston (1994) do. But it might be just as reasonable to embrace the interdependencies and the ambiguous character of the subject. To give a broad impression of how differently questions can be looked at, this section briefly introduces three studies, that diverge from the way, in which questions have been discussed so far.

### Sentence type

One example for thinking of form and function as two sides of the same coin is given by Sadock & Zwicky (1985), whose work is also based on the theory of speech-acts. Their concept of sentence type bundles sentences, which share both a certain structure and a conventional conversational use. One specific sentence type would be for instance the combination of subject-auxiliary inversion and final intonational rise, which then coincides with the use of asking a yes-no question (Sadock & Zwicky 1985: 155). Importantly, Sadock & Zwicky speak from a conventionalized use, which does not have to be met in every specific occasion. A sentence of the type yes-no question could also function as a reminder, a demand, an exclamation etc., but normally does not.

Regarding interrogatives, a sentence type that subsumes all kinds of questions under one category, the following is stated: All of the languages, that the authors have surveyed, "have some grammatical signal indicating that the purpose of a sentence that is so marked is to gain information" (Sadock & Zwicky 1985: 178). Again, this observation includes both structural and functional aspects.



Note the deviations from the terminology, that has been introduced by Lyons (1977) and Huddleston (1994): For them, sentences, interrogatives or questions do not refer to the same phenomena as the homonym terms do for Sadock & Zwicky (1985).

### **Prototypicality**

Another way of approaching questions is given by Judith Haan (2002). Focussing on pragmatics, she generally defines questions as “utterance[s] intended to function as a request for information” (Haan 2002: 12). This rules out all kinds of questions (in Huddleston’s terminology), that function as indirect speech-acts, e.g. requests for action. On that basis Haan establishes criteria for prototypicality, by which every possible question type can be located on a scale – from prototypical questions to rather atypical questions.

The first criterion for prototypicality is the amount of information, that is asked for. In that regard, variable questions are considered more prototypical than polar questions. Also, questions, that seek for new information, are more prototypical than questions, that rather seek for the confirmation of available information (Haan 2002: 11).

Next, primary questions, which can be asked independently from the conversational context, are considered more prototypical than secondary questions, which depend on preceding utterances (Haan 2002: 11).

And finally, question types, which combine their interrogative function – a thing, that Huddleston would call inquiry force – with interrogative syntax, are considered more prototypical. This means, for example, that a polar question expressed by an interrogative clause is considered more prototypical than a polar question expressed by a declarative clause (Haan 2002: 12).

Again, one can see, how the perspective on the phenomenon affects the use of terminology.

### **Questionhood**

Related to Haan’s concept of prototypicality is the concept of questionhood introduced by Stephen Levinson (2012), who approaches the subject of questions from a social point

of view. His main idea is, that information exchanged in a conversation is not a free good, but comes at a price.

In his model of the conversational economy of information there are two kinds of currencies: one measures the value of the semantic information, the other measures the social costs. While the amount of information, that a question asks for, can basically be quantified, the social costs are not captured in such a concrete manner. It is stated, however, that the higher the informational costs, the higher are the social costs. As potential social costs of asking a question, Levinson names e.g. face loss due to ignorance, the disclosure of the speaker's interests and concerns, or the demand of the addressee to get a similar information in return (Levinson 2012: 20). These social costs may cause the reluctance of speakers to ask for information, which they are actually interested in.

To classify different types of questions and its social costs, Levinson develops a multidimensional function space, which takes into account, amongst other things, the degree of newness or givenness of the requested information, the degree of sincerity in the speaker's wish for the requested information, and the degree of the hearer's obligation to respond. Within this space the social costs of a question increase as one moves towards greater questionhood.

To lower the social costs, Levinson predicts, speakers will always tend to the smallest degree of questionhood, which is found to be compatible with the speaker's conversational needs. This means, that speakers will avoid variable questions, when polar questions are sufficient. And, when possible, speakers will use the declarative form to express questions rather than the interrogative form (Levinson 2012: 23).

This section very briefly touched the possibilities of analysing questions by the concepts of sentence type (Sadock & Zwicky 1985), prototypicality (Haan 2002) or questionhood (Levinson 2012). Even though these concepts and their inherent perspectives on questions do not come to direct use in the current paper, it is worthwhile to keep them in mind for at least two reasons. First, one should stay aware of the fact, that the term *question* – as used in the following – is by no means uncontroversial. And second, it might be fruitful to look at the studies to be discussed in the next chapter from varying angles, considering for example the social costs of questions or the difference between conventionalized and actual use of certain utterances.

## 2.5 What is a non-genuine polar question?

This chapter gave a broad and non-exhaustive introduction to the theory of questions. But as the current study deals with the prosody of one explicit question category, i.e. non-genuine polar questions, it is now time to define strictly what non-genuine polar questions are.

Following Huddleston's (1994) semantic approach, a question is defined as defining a set of possible answers. A polar question then is a question, which defines a set of two answers, one having the same propositional content as the question, the other having the opposite propositional content. In simple terms: Valid answers to polar questions are "Yes" and "No" or variations of these.

Genuine polar questions are polar questions with inquiry force, i.e. polar questions, that seek to get information from the addressee, which is desired by, but not available to the speaker. Importantly, the genuinity of a question refers to its pragmatic function in conversation. To repeat the quote from section 2.3, genuine questions are "used with the aim of discovering the answer from the addressee" (Huddleston 1994: 414).

Banuazizi & Creswell (1999) have elaborated this definition by putting forward three sincerity conditions for genuine questions. The term sincerity condition again derives from the theory of speech-acts and defines one specific kind of felicity conditions, "which an illocutionary act must fulfil if it is to be successful and non-defective" (Lyons 1977: 733). Generally, three groups of felicity conditions are distinguished, namely preparatory, sincerity and essential conditions. Without going into detail: A speaker performing a speech-act must be in the positions to do so (preparatory conditions), he must be sincere about it (sincerity conditions), and he must behave consistently and coherently with regard to the speech-act (essential conditions)<sup>8</sup>.

It is important to note, that felicity conditions are not fixed, but vary depending on the illocutionary act performed. And while preparatory and essential conditions are ignored at this point, Banuazizi & Creswell (1999) focus on the sincerity conditions of genuine questions. Following Hudson (1975) and Kenworthy (1978), they propose the following conditions, all of which have to be fulfilled by a question to count as genuine.

---

<sup>8</sup> For details on felicity conditions s. Lyons (1977: 733, 734).

1. The speaker does not already know – or believe to know – the answer to the question.
2. The speaker desires the answer from the addressee.
3. The speaker expects, that the addressee can provide the answer<sup>9</sup>.

(Banuazizi & Creswell 1999: 8, 9)

These sincerity conditions are taken over for this paper's definition of genuine polar questions.

Finally, non-genuine polar questions are polar questions without inquiry force, i.e. all kinds of polar questions, that do not sincerely or primarily aim at discovering an answer from the addressee. In terms of the established sincerity conditions, non-genuine polar questions are defined as polar questions with a primary function, that violates at least one of the conditions.

The prosody of non-genuine polar questions and, in comparison to that, the prosody of genuine polar questions – with particular focus on intonation – are dealt with in the next chapters.

### Chapter 3: The prosody of polar questions

It is a widely approved hypothesis, that polar questions generally tend to intonation contours with a final rise. Hedberg et al. (2017) name, among others, Cruttenden (1986, 1997), Pierrehumbert & Hirschberg (1990), Ladd (1996) and Gussenhoven (1983, 2004) to support this claim (Hedberg et al. 2017: 323). Importantly, a distinction between genuine and non-genuine polar questions is hardly drawn in any of these studies. But as the current chapter will show, this distinction seems to be quite relevant for the prosody – and especially for the intonation – of polar questions.

In the first section of this chapter two corpus studies on polar questions will be presented, which take a closer look at general tendencies of their final intonation, functional differences between rising and non-rising intonation contours as well as

---

<sup>9</sup> Instead of *answer* Banuazizi & Creswell speak of *response* in conditions 2 and 3. It is not clear, whether they systematically distinguish between the two terms as suggested in section 2.3. In this study, however, the term *answer* is willfully chosen over *response* for the sincerity conditions.

intonational differences based on the question's status as genuine or non-genuine. A particular sub-class of non-genuine polar questions, namely the class of rhetorical polar questions, and its prosodic characteristics will be examined in more detail in section 3.2. Before summarizing the results of all surveyed studies in the final section of this chapter, two studies, that seem to contradict the initial claim of a tendency towards terminal rises in polar questions, will be discussed in section 3.3. They are especially interesting with regard to their speech data, which is taken from television broadcasts, just as it is done in the self-conducted study to be reported in chapter 4.

### **3.1 Genuine and non-genuine polar questions**

#### **3.1.1 Corpus study I: Banuazizi & Creswell (1999)**

The first study to be presented here was conducted by Banuazizi & Creswell (1999), who analyzed polar questions from a corpus of recorded telephone conversations in American English. Their speech data is “real” in the sense, that it is spontaneous and not designed to elicit specific utterances in a specific manner.

Mainly due to practical reasons, i.e. the pre-assigned annotations of the corpus, it was distinguished between polar questions, rhetorical polar questions and backchannel polar questions. While not much is said about functional aspects of that a priori categorization by the authors, it can be stated, that both rhetorical and backchannel questions are non-genuine questions. The general category of polar questions on the other hand can contain both genuine and non-genuine members.

It should also be noted, that Banuazizi & Creswell (1999) excluded all questions from their analysis, that did not show the syntax of closed interrogatives. Polar questions in declarative form, for example, were not taken into account.

With regards to intonation, the extracted questions were classified as either final falls or final rises – categories, which correlate with low or high boundary tones, respectively, in the ToBI (Tones and Break Indices) annotation system (cmp. e.g. Pierrehumbert & Hirschberg 1990). Importantly, utterances, that ended in a level plateau, were excluded from the study, but they were claimed to make up about 1 % of all tokens (Banuazizi & Creswell 1999: 6).

Discourse Coding	Total Number	Rises (%)	Falls (%)
qy (yes-no question)	2106	89.7	10.3
qh (rhetorical question)	102	55.9	44.1
bh (backchannel question)	130	49.2	50.8

Figure 1: Distribution of final falls and final rises within general polar questions (qy), rhetorical polar questions (qh) and backchannel polar questions (bh). Source: Banuazizi & Creswell 1999: 7.

The table in figure 1 depicts the distribution of falls and rises among the three question categories. From a total of 2106 polar questions – not containing the rhetorical and backchannel questions – 1889 (89.7 %) showed final rising contours, while 217 (10.3 %) showed final falling contours. This is well in line with the hypothesis of a general tendency towards rising contours in polar questions.

More surprising is the fact, that the results for the rhetorical polar and backchannel polar questions don't show this tendency. In the former group 55.9 % of the utterances are categorized as rises and 44.1 % as falls. For the latter group it is 49.2 % rises and 50.8 % falls.

In the light of these numbers, a discourse function analysis was conducted for the 217 general polar questions with falling contours and a comparable number of general polar questions with rising contours. Each of these questions was coded either as genuine or non-genuine by applying the three sincerity conditions explained in section 2.5.

Among the kinds of questions, that violated one or more of the conditions and thus were categorized as non-genuine, were e.g. requests couched in question form ("Can you tell me anything else about the house?"), information couched in question form ("Have you tried the pool at the Spring Creek Fitness Center?") or quoted questions (Banuazizi & Creswell 1999: 8, 9).

	Total	No. of Genuine Questions	Percent Genuine
Rise	237	137	57.8
Fall	217	26	12.0

Figure 2: Shares of genuine questions within the rising and falling polar questions. Source: Banuazizi & Creswell 1999: 9.

The shares of genuine questions in rising and falling contours, respectively, both in absolute numbers and in percent, are shown in the table of figure 2. While ca. 58 % of the rising questions are genuine in the described sense, it is only 12 % of the falling questions. Banuazizi & Creswell conclude: “The patterns in our data indicate that rises are associated with questions that fulfill all our previously stated sincerity conditions. In contrast, speakers use falling intonation for yes-no questions where sincerity conditions do not hold” (Banuazizi & Creswell 1999: 10).

When looking at the considerable amount of non-genuine questions within the group of the rises (42.2 %), this thesis is at least daring. The assumption of Banuazizi & Creswell is, that those cases can be explained by politeness strategies. E.g. informing someone about something, which is possibly already known by the addressee, can be a face-threatening act. Thus, the speaker might wrap up the information in a polar question (“Did you know X?”) with rising intonation, to indicate complete openness regarding the answer and make it less face-threatening and more polite<sup>10</sup>.

For rhetorical and backchannel questions, which showed remarkably higher rates of falling intonation contours than the main group of polar questions (s. figure 1), it is stated, that they violate the first two of the three sincerity conditions and thus are non-genuine by definition.

The authors’ hypothesis, that non-genuine polar questions tend to falling intonation, however, seems to be contradicted by the numbers as well. Unfortunately, an explanation for the relatively high rates of rising intonation contours in rhetorical and backchannel polar questions – 49.2 % and 55.9 %, respectively (s. figure 1) – is not given.

But it is at this point, that Banuazizi & Creswell (1999) give away some thoughts on the functional aspects of rhetorical questions: “The answers to rhetorical questions are, in the speaker’s mind, either perfectly obvious or perfectly obviously unknowable – in neither case would [the speaker] require or expect an answer” (Banuazizi & Creswell 1999: 10). A closer look at rhetorical questions including a more clear-cut definition will be presented in section 3.2, which discusses a study of Dehé & Braun (2020). For now, it should be noted, that rhetorical questions and non-genuine questions are not identical.

---

<sup>10</sup> A discussion of how this hypothesis interferes with Levinson’s (2012) theory of the social costs of questions (s. section 2.4) would be very interesting, but exceeds the scope of this work.

While all rhetorical questions are non-genuine questions, not all non-genuine questions are rhetorical questions.

To summarize the findings of Banuazizi & Creswell (1999): The vast majority of polar questions in their corpus are produced with rising contours, which they associate with the genuinity of these questions. Falling contours on the other hand, which occur less frequently, are associated with non-genuine questions, among which are backchannel and rhetorical questions. Keeping in mind the contradicting cases, they conclude: “Discourse function does correlate – if not categorically, then at least somewhat predictably – with final contour” (Banuazizi & Creswell 1999: 11).

### **3.1.2 Corpus study II: Hedberg et al. (2017)**

Following Banuazizi & Creswell (1999), Hedberg et al. conducted a methodologically similar study in 2017. Again, they took speech data from two corpora of recorded telephone conversations by speakers of American English. Both corpora were different to the one analyzed by Banuazizi & Creswell (1999), but they, too, had the advantage of providing spontaneous utterances from a non-laboratory environment.

Hedberg et al. (2017) determined a total of 410 polar questions, which included only positive polar questions with interrogative syntax (Hedberg et al. 2017: 329). Positive polar questions mean such polar questions, that do not involve grammatical negation. An utterance like “Don’t you see it?”, while being a polar question in form of a closed interrogative clause, would thus have been sorted out. All questions of interest were then categorized phonologically in terms of ToBI with some minor adjustments according to the study’s needs (Hedberg et al. 2017: 329)<sup>11</sup>.

Out of the 410 polar questions 371 (90.5 %) showed final rising contours. These were composed of low-rises, which were most frequent by far (79.8 %), and high-rises (10.7 %). While both low- and high-rises are patterns of final rises, in the former the rise is preceded by a low nuclear accent, and in the latter the rise is preceded by a mid-level or high nuclear accent and typically rises to an even higher end point, relative to the speaker’s pitch range (Hedberg et al. 2017: 333, 336). Hedberg et al. (2017) suggest, that within the rising category, the distinction between low- and high-rises lies in information

---

<sup>11</sup> A table of all nuclear contours and their distribution can be found in the appendix A1.



structure. While low-rises seem to be the unmarked intonation contour for polar questions, high-rises may be used – among other functions – to signal contrastive focus on pronouns or other given words in the discourse, “that would typically be left unaccented” (Hedberg et al. 2017: 337). For the purposes of this very study the distinction between low- and high-rises is not crucial<sup>12</sup>. More interesting is the general distinction between rising and falling contours, from which the former is far more frequent in the corpus of Hedberg et al. (2017).

In opposition to the 90.5 % of pitch contours with a final rise (both low- and high-rises), only 5.6 % of the analyzed questions were found to have falling contours, i.e. a total of 23 cases. The remaining questions showed either level contours (3.4 %) or fall-rises, which were conceptionally separated from the other rising contours, but with only 2 instances (0.5 %) were a minor category anyway.

Just like Banuazizi & Creswell (1999) Hedberg et al. (2017) analyzed the questions with falling intonation in terms of the established sincerity conditions. They found, that 22 out of the 23 (95.65 %) fitted the category of a non-genuine question. Thus, in the whole corpus of 410 positive polar questions with interrogative syntax, there was only one instance of a genuine polar question with falling intonation. This is – within a smaller corpus of course – even less than Banuazizi & Creswell (1999) found (cmp. figure 2).

Out of 96 polar questions with rising contours (here: low-rises), which were randomly chosen for comparison, 84 (87.5 %) were categorized as genuine and 12 (12.5 %) were categorized as non-genuine (Hedberg et al. 2017: 339, 340). As mentioned before and depicted in figure 2, Banuazizi and Creswell (1999) found 57.8 % of their rising polar questions to be genuine, which is quite different to the 87.5 % of Hedberg et al. (2017). This difference gets even more remarkable, when considering the pre-exclusion of the non-genuine sub-categories of backchannel and rhetorical questions by Banuazizi & Creswell (1999). Even though Banuazizi & Creswell (1999) tried to explain their relatively high percentage of non-genuine questions within the rising intonation contours with politeness strategies (s. section 3.1.1), it must be stated, that the two studies provide

---

<sup>12</sup> For a detailed discussion of the meaning of low- and high-rises in polar questions s. Hedberg et al. (2017: 334–337). It should be briefly mentioned, however, that the findings of Hedberg et al. contradict the frequently made assertion, that American English – other than British English – tends to high-rises in polar questions rather than low-rises. In showing, that low-rises outweigh high-rises in speech production, Hedberg et al. (2017) give complementary support to Levis (1999), who presented similar findings from a perception experiment.

contradicting results here. Oddly, the data of Hedberg et al. (2017) provide better proof for the conclusions drawn by Banuazizi & Creswell (1999) than their own data.

One interesting difference between the two studies, which is implied by these numbers, lies in the overall ratio of genuine to non-genuine questions. Even though the data sets are not strictly comparable, it can be said, that Hedberg et al. (2017) found a remarkably smaller share of non-genuine questions. Again, from the large group of polar questions with rising intonation, which makes up about 90 % of all polar questions in their corpus, an estimated 12 % are considered non-genuine<sup>13</sup>. On top of that come the non-genuine questions from the remaining 10 % of the corpus<sup>14</sup>.

In comparison, from the group of rising polar questions in the corpus of Banuazizi and Creswell (1999) an estimated 42.8 % are considered non-genuine. From the considerably smaller category of polar questions with falling intonation, it is even 88 % (cmp. figure 2). And in addition to that, all rhetorical and backchannel questions, regardless from their intonation, are non-genuine by definition<sup>15</sup>.

Both studies claim to apply the same criteria to distinguish between genuine and non-genuine questions, which leads to the suspicion, that the observed differences are based on systematic differences between the data sets, rather than on methodology. This would be yet another example for the difficulty to account for all relevant parameters, when analyzing naturally occurring spoken language.

Regarding meaningful sub-categories of non-genuine questions, Hedberg et al. (2017) write, that such non-genuine questions, that clearly expect a particular response, were most frequent. An example for this is the question “Do you have a cold?”, which a speaker asks after hearing the addressee coughing and talking in a hoarse voice (Hedberg et al. 2017: 341). Other kinds of non-genuine questions were, among others, requests for action (“Yeah, would you please tell them?”), announcements of information (“Did I tell you that I have a new job?”) or backchannel responses (“Did they?”) (Hedberg et al. 2017: 340, 341).

---

<sup>13</sup> The difference between low- and high-rises is ignored here for the sake of simplicity.

<sup>14</sup> A concrete number is not given, but remember that from the 23 utterances with final falls (5.6 % of the whole corpus), 22 were considered non-genuine.

<sup>15</sup> The total amount of rhetorical and backchannel questions within their corpus, remains unclear due to the ambiguous pre-assigned labels. Banuazizi & Creswell (1999) analyse only portions of them to compare the composition of the different groups of questions. A comparison of the group sizes themselves is not readily possible. For the concrete problems, they faced with the corpus, s. Banuazizi & Creswell (1999: 5–7).

Rhetorical questions are not mentioned in particular by Hedberg et al. (2017). But as every categorization of specific speech-acts depends on the underlying definition, this does not necessarily mean, that rhetorical questions do not occur in their corpus. In fact, some of the questions, that clearly expect a particular response, might be considered rhetorical questions.

Interesting is, what Hedberg et al. (2017) write on level contours, which – as mentioned above – make up 3.4 % of all polar questions from their corpus: “Most of our examples of questions ending in a level contour fall into categories of non-genuine questions already established for questions ending in falling contours” (Hedberg et al. 2017: 347, 348). This can be linked to the findings of Dehé & Braun (2020), who distinguish between rising and level contours – rather than rising and falling contours – as indicators for genuine questions and rhetorical questions, respectively. A detailed report of the study by Dehé & Braun follows in the next section. Banuazizi & Creswell (1999), on the other hand, excluded final level contours (ca. 1 % of the polar questions in their corpus) from further analysis.

Generally, one can say that the findings of Banuazizi & Creswell (1999) and Hedberg et al. (2017) support each other in showing, that polar questions are indeed typically uttered with final rising intonation – when functioning as genuine requests for information. But this tendency does not go for polar questions with other illocutionary forces. Whether non-genuine polar questions rather tend to falling intonation – as Banuazizi & Creswell (1999) claim – or don’t show clear tendencies towards one or the other contour, remains somewhat fuzzy. This issue will be addressed in the self-conducted study in chapter 4, when different kinds of non-genuine polar questions and their prosodic characteristics are looked at separately.

### **3.2 Rhetorical questions as a sub-class of non-genuine questions**

In 2020 Dehé & Braun presented a study, which is of particular interest in the context of polar question prosody, because it differs in at least three aspects from the previously discussed studies.

First, Dehé & Braun conducted a production experiment instead of analyzing “natural” speech from a corresponding corpus. This is by no means generally advantageous to other methods, but it helps in controlling the enormous number of intertwined parameters involved in a complex phenomenon like spoken language. The plus in comparability of the speech data, however, comes with a lack of spontaneity and authenticity<sup>16</sup>.

Second, Dehé & Braun (2020) focused on the difference between genuine questions and one particular sub-class of non-genuine questions, namely rhetorical questions. By this specification, they avoid to lump all kinds of non-genuine questions together and thereby come to possibly undue generalizations.

And third, in addition to the intonational analysis, Dehé & Braun (2020) also looked at other prosodic parameters, from which the measurement of duration is particularly relevant for our purposes. The results lead to further insights on polar question prosody and help to interpret the outcomes of the self-conducted study in chapter 4.

To begin with, Dehé & Braun (2020) define the two kinds of questions under investigation, i.e. information-seeking questions and rhetorical questions, respectively. Information-seeking questions are described as “perform[ing] the directive speech[-]act of requesting information” (Dehé & Braun 2020: 607) and as they are congruent with what has been called genuine questions so far, this latter terminology will be used here.

For rhetorical questions, a sub-class of the non-genuine questions, three sincerity conditions are given:

1. Rhetorical questions do not expect an answer.
2. Rhetorical questions have the feel of an assertion.
3. Rhetorical questions do not have to but can optionally be answered.

(Dehé & Braun 2020: 608)

On the basis of the second condition, the feel of an assertion, it can be shown easily, why rhetorical questions are a sub-category of non-genuine questions, but are not identical to them. Questions, that function as requests for action, for example, while clearly being non-genuine, do not have assertive character and thus are no rhetorical questions. An

---

<sup>16</sup> For a discussion on which kind of speech data should be analyzed in prosodic research s. e.g. Couper-Kuhlen & Selting (1996), who argue in favor of naturally occurring speech.

utterance like “Can you repeat that?”, which demands someone to repeat what was just said, can exemplify this. It is not requesting certain information, but it does not claim a certain proposition to be true, either. Again: Non-genuine questions describe a collective category, that contains different kinds of speech-acts, among them rhetorical questions, requests for action and many more.

To investigate the prosodic differences between genuine questions and rhetorical questions, Dehé & Braun (2020) conducted two mirrored production experiments on variable questions and polar questions, respectively. For the sake of brevity and to keep the focus on our purposes, this report will be limited to the relevant aspects of polar questions.

The experiment contained 21 different interrogative clauses, each of which had to be produced by the participants in two different contexts – one eliciting a genuine question, the other eliciting a rhetorical question. One example of such an interrogative and its two contexts can be found in the appendix A2. 21 subjects took part in the experiment and after cleaning the data with regards to defective productions, a total of 433 utterances of polar questions (218 genuine polar questions and 215 rhetorical polar questions) were left for phonological and phonetic analysis (Dehé & Braun 2020: 615).

Regarding the final intonation contours – annotated in terms of MAE\_ToBi, a variation of ToBI for mainstream American English (cmp. e.g. Beckman & Ayers 1997) – the vast majority of the utterances showed one out of two so called edge tones: H-H% and H-L%, which in MAE\_ToBi represent a final rise and a final mid-level plateau, respectively. So, in opposition to the other studies, discussed so far, the main intonational difference is not drawn between final rising and falling contours, but between rising and level contours.

As can be seen in figure 3, the predominant contour for genuine questions was the final rise with 82.1 % (179 out of 218). The final mid-level plateau made up 9.2 % (20 out of 218) and the remaining genuine questions were produced with other final contours, which are not further discussed here (s. Dehé & Braun 2020 for details).

In comparison to that, more than every second rhetorical question (53.5 % or 115 out of 215) was uttered with a final mid-level plateau, while final rises made up 45.6 % (98 out of 215). Again, the minority of rhetorical questions with other final contours is ignored here. Pictures of the oscillogram, the spectrogram and the pitch contour both of a typical

genuine polar question with a final rise and a typical rhetorical polar question with a final mid-level plateau can be found in the appendix A3 and A4, respectively.

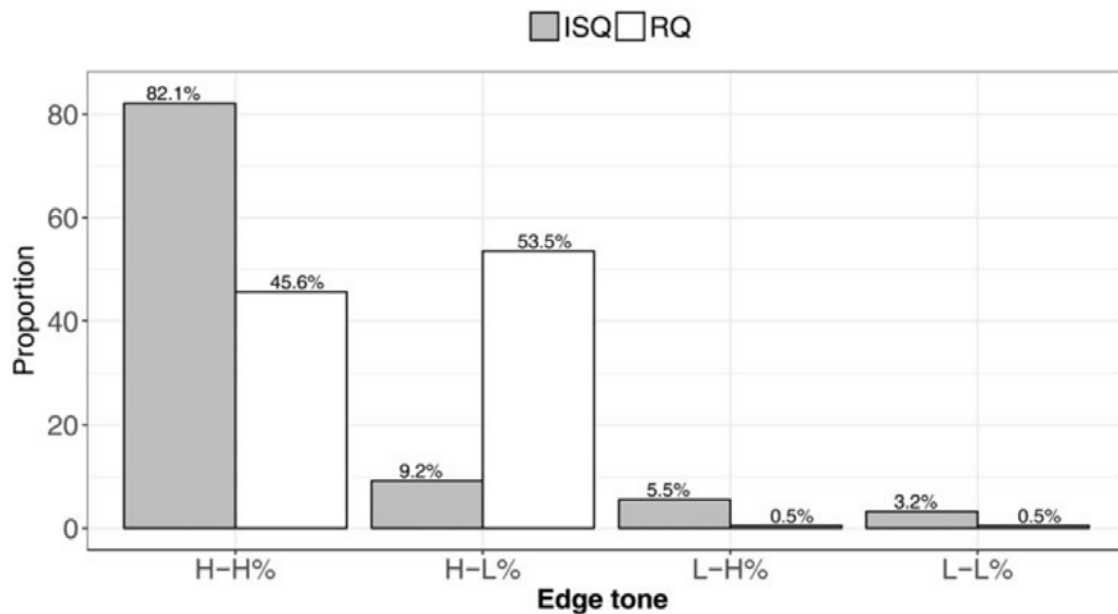


Figure 3: Distribution of edge tones in genuine polar questions and rhetorical polar questions. H-H% means a final rise, H-L% means a final plateau. ISQ stands for information-seeking questions, i.e. genuine questions. RQ stands for rhetorical questions. Source: Dehé & Braun 2020: 623.

These findings can be interpreted to support both Banuazizi & Creswell (1999) and Hedberg et al. (2017) in that genuine polar questions have a strong tendency towards final rises. The rhetorical polar questions generally show greater variety regarding their terminal intonation, with both rising and non-rising contours. And while Hedberg et al. (2017) didn't analyze rhetorical questions in particular, this is in line with the findings of Banuazizi & Creswell (1999), too.

Note, however, that falling contours as described in both corpus studies do not play a role in the production experiment by Dehé & Braun (2020). In the latter it is rather a pattern of level contours around the middle of the speaker's pitch range, that is associated with rhetorical questions. Thus, the term non-rising is used here to reconcile the different findings, which are far from congruent.

Taking into account the relatively high rates of rhetorical questions with rising contours both in Banuazizi & Creswell (1999) (s. figure 1 in section 3.1.1) and in Dehé & Braun (2020), one may conclude: Non-rising intonation is a good indicator for a polar question to be non-genuine, but non-genuine polar questions and specifically rhetorical polar

questions are not necessarily produced with non-rising contours. We will come back to this, when discussing the results of the study in chapter 4.

As mentioned, Dehé & Braun (2020) did not focus on terminal intonation alone. Other parameters like placement and type of pitch accents, voice quality or initial pitch were looked at, too. But the clearest results, and therefore the ones to be reported here, were provided by the measurements of duration, which revealed significant differences based on illocutionary force. Concretely, rhetorical questions showed longer average durations than genuine questions, both in the subject (plus 50 ms) and in the object (plus 57 ms) (Dehé & Braun 2020: 626)<sup>17</sup>. These results should be viewed in the light of former studies, which generally claim faster speaking rates for questions than for statements (s. e.g. Heuven & Zanten 2005 on Manado Malay, Orkney English and Dutch). As rhetorical questions by definition have the feel of an assertion, they seem to share durational properties with statements rather than with genuine questions. Studies on German (Braun et al. 2018) and Icelandic (Dehé et al. 2018) support this hypothesis by reporting longer constituent durations in rhetorical questions than in genuine questions (Dehé & Braun 2020: 612).

In summary, Dehé & Braun presented similar results as Banuazizi & Creswell (1999) and Hedberg et al. (2017) in terms of final intonation – with the discussed difference between mid-level and falling contours. Additionally, they conducted further prosodic analyses, from which especially the duration measurements provided fruitful insights for the self-conducted study to be presented in the next chapter. Generally, when comparing the results of linguistic studies, it is important to keep in mind the procedural differences, the most crucial of which lies in the choice of speech data.

### 3.3 Contradicting studies

Not all studies on the intonation of polar questions have found a clear majority of final rises. Two studies with different results – by Fries (1964) and Hedberg & Sosa (2002), respectively – will be briefly presented in this section. But just as the claim of a tendency towards final rises, which has been shown to be true for genuine polar questions rather

---

<sup>17</sup> Obviously, the lexical identity of genuine and rhetorical questions is crucial for the meaningfulness of such a comparison.

than for all polar questions, the contradicting claim seems to ignore the crucial distinction between genuine questions and other speech-acts.

### **Fries (1964)**

An early empirically supported argument against the final rise hypothesis came from Charles C. Fries in 1964. He obtained speech data from 39 television-radio shows, in which panellists have to figure out the occupation of contestants by asking polar questions (Fries 1964: 247). About three quarters of the analyzed utterances were produced by one of the three regular panellists of the show. The other utterances were produced by a total of 25 guest panellists. From a total of 2561 polar questions, 1589 (61.7 %) showed final falling intonation and 981 (38.3 %) showed final rising intonation. Fries' interpretation of these numbers is as logical as rigorous: "[...] there is no question intonation pattern as such" (Fries 1964: 250).

Banuazizi & Creswell (1999), however, contradict Fries' conclusion by denying the genuine character of the analyzed questions. While it might be true, that "[t]he speed and spontaneity of the language activity of these panellists reduced to practical zero the chance that the intonation forms of that language activity could have been premeditated or deliberately chosen" (Fries 1964: 247), it does not say anything about the conversational function of these spontaneously uttered questions. In this regard Banuazizi & Creswell point out, that the show design encourages the panellists to seek for confirmation of their suspicion and thus ask questions, to which they are confident to know the correct answer (Banuazizi & Creswell 1999: 3). This would mean a violation of the first sincerity condition of genuine questions (s. section 2.5) for an unknown, but potentially big part of the 2561 questions.

Therefore, it is at least questionable, if Fries' study can be taken as a strong indication against the claim of final rising intonation in genuine polar questions. On the contrary, the results seem to be – to different degrees – in line with what Banuazizi & Creswell (1999), Hedberg et al. (2017) and Dehé and Braun (2020) found on non-genuine questions.



### **Hedberg & Sosa (2002)**

Newer, but similarly weak evidence against a general tendency towards final rises in (genuine) polar questions came from Hedberg & Sosa (2002). They analysed 29 positive and 9 negative polar questions from five episodes of two different public affairs discussion programs in American TV (Hedberg & Sosa 2002: 2). From the 29 positive polar questions, only 9 showed final rising intonation contours, while 20 showed either falling or level contours. The group of 9 negative polar questions was composed of 4 rising and 5 falling contours.

Indeed, there is no majority of final rises as observed in the previously discussed studies, but one has to emphasize the relatively small size of the data set, which contains only 38 questions. And interestingly, a qualifying comment on the data was retrospectively given by the authors themselves in the paper of Hedberg et al. (2017). Without naming a concrete number they state, that “most of the questions” (Hedberg et al. 2017: 344) were asked by the moderator of the TV show and had the function to introduce new topics to the discourse. Questions like these are explicitly classified as non-genuine by Banuazizi & Creswell (1999): Instead of aiming at an answer of the type *yes* or *no*, they “attempt[ing] to initiate a more extended discussion” (Banuazizi & Creswell 1999: 8). A closer investigation of such topic introductions will be given in chapter 4.

So comparable to the results of Fries (1964), the non-tendency towards rising intonation in polar questions, indicated by Hedberg & Sosa (2002), seems to be true for a specific kind of non-genuine polar questions, but not for genuine polar questions.

## **3.4 Summary**

This chapter started with the frequently made claim, that polar questions typically show final rising intonation. While this is true for some of the corpora of the presented studies (Banuazizi & Creswell 1999, Hedberg et al. 2017), it is not true for others (Fries 1964, Hedberg & Sosa 2002).

Looking at the sheer numbers of rising and non-rising contours in polar questions, it seems, is not a sufficient way of tackling the issue. For a deeper understanding of the relation between polar questions and intonation one has to take into account functional aspects aswell. And by doing this, it could be shown both by the corpus studies of

Banuazizi & Cresweell (1999) and Hedberg et al. (2017) and by the production experiment of Dehé & Braun (2020), that there is a tendency towards rising intonation for one specific kind of polar questions, namely for genuine polar questions. In contrast, polar questions of different illocutionary forces, summed up in the heterogeneous group of non-genuine polar questions, don't show this tendency.

This insight makes possible a different view at the studies of Fries (1964) and Hedberg & Sosa (2002). Instead of generally showing, that there is no tendency towards final rising intonation in polar questions, their results can be seen as another argument in favour of a thorough distinction between polar questions of varying illocutionary force.

Dehé & Braun (2020) started to disentangle the group of non-genuine polar questions by focussing on the functionally defined sub-class of rhetorical polar questions. While the distribution of intonation patterns here is clearly distinct from genuine polar questions, a comparison to other sub-classes of non-genuine polar questions remains to be done. This is, what the study in the next chapter aims at. Following Dehé & Braun (2020), not only intonation will be analysed, but other prosodic features aswell. In the case of rhetorical questions, especially the measurement of utterance duration seems relevant.

## **Chapter 4: The prosody of non-genuine polar questions – an exploratory study**

In order to shed some light on the prosodic characteristics of specific kinds of non-genuine polar questions and possible differences between them, an exploratory corpus study was conducted. Ca. 100 utterances of each of three different English clauses, all closed interrogatives syntactically and polar questions semantically, were taken from television broadcasts, provided by the TV News Archive (2023). These clauses were:

- (A) Can you repeat that?
- (B) Can you believe that?
- (C) Do you believe that?

The idea was, that utterances of A, B and C typically are indirect speech-acts, i.e non-genuine questions, which function as requests for action, rhetorical questions and topic

introductions, respectively. After inspecting, whether this assumption held true, the utterances were analysed in terms of the *ProPer* method (Albert et al. 2023) to find and compare prosodic patterns of the different speech-acts. This chapter thoroughly reports all steps of the study and puts in the previously discussed thoughts on question theory and polar question prosody.

In the first section, the kind of speech data used for this study is introduced. On the one hand, this includes more information on the TV News Archive aswell as general arguments against and in favour of speech data from TV broadcasts. On the other hand, it includes a discussion on the three specific clauses and the so called *form and function problem* (Cameron et al. 1989), which raises the issue of deducing certain speech-acts from lexical strings. Especially clause C and its concrete utterances reveal themselves as a borderline case and need closer examination.

The second section focusses on the methodological procedure. It reports in detail the coding and annotation of the data and introduces *ProPer*, a rather new method of prosodic analysis, which does not focus on F0 trajectories alone, but, importantly, takes into account periodic energy.

The results and a comprehensive discussion of these follow in the third section.

## **4.1 Data**

### **4.1.1 The TV News Archive**

In their paper on interactional aspects of prosodic research Couper-Kuhlen & Selting (1996) argue for the analysis of real conversations rather than constructed examples. While this, in principle, is a reasonable approach, the choice of speech data almost always remains a compromise between the poles of speaker's authenticity and scientific usability. In this regard, the TV News Archive is no exception.

It is a sub-section of the Internet Archive (2023), a non-profit endeavour to build an extensive digitalized “library of [i]nternet sites and other cultural artifacts” (Internet Archive About 2023). The Internet Archive provides free access to millions of historic web pages, books, audio recordings, videos, images or software programs – and in form of the TV News Archive it provides clips from more than 2.5 million US television news

shows from 2009 until today<sup>18</sup> (TV News Archive 2023). Automatically generated transcripts make it possible to search for specific lexical strings within the pool of those clips.

Thus, the first and undeniably important reason for using the TV News Archive as a source of speech data is a pragmatic one: its size and accessibility. It just offers an easy way to get a huge number of English utterances, legally and free of charge. The fact, that the clips are publicly accessible also prevents a problem, that often arises for studies with experimentally collected speech data. Due to legal restrictions, their raw data cannot be published easily, which makes it harder for third parties to replicate and verify the study's results. Such obstacles do not occur with data from the TV News Archive.

Another point in favour of the TV News Archive is, that it provides data of non-scripted and spontaneous speech. By taking utterances from hearings, interviews or news broadcasts, one avoids the artificial character of experimental set-ups, which – as criticised by Couper-Kuhlen & Selting (1996) – aim at eliciting specific utterances in a specific manner and by design ignore what might be called real conversations or everyday speech.

But as pointed out before, this advantage comes at the price of fuzziness. Identifying certain utterances as members of the same group of speech-acts is not as straightforward in natural speech data as it might be in data from the laboratory. Aside from the lexical string there are few parameters, that can be controlled. This leads to a somewhat heterogeneous corpus, which needs more effort to interpret and categorize in meaningful ways.

At the same time, one has to acknowledge, that the different kinds of TV formats bear different kinds of artificial traits themselves. While the utterances looked at are not scripted (as in movies, series or commercials), they nevertheless happen to occur in staged contexts. The people producing the utterances, for example in news studios or on a lectern, are aware, that they are being filmed for public. And from witnesses in public hearings to politicians or anchorpersons they are to different degrees trained and experienced in performing for an audience. This kind of TV broadcast idiosyncrasy can hardly be quantified, but must be kept in mind, when working with such data. As section

---

<sup>18</sup> As it is in the nature of television to work with archive material, like quotes from former interviews or replays of whole programmes, this does not mean, that all utterances to be collected from the TV News Archive have necessarily been made since 2009.

3.3 and its examples of studies on speech data from television broadcasts (Fries 1964, Hedberg & Sosa 2002) have shown: Considering the specific situational contexts of utterances is crucial for the interpretation of their pragmatic functions. Using the terminology of Couper-Kuhlen & Selting (1996), one can say that speech data from non-scripted television broadcasts as chosen for this very study might not be “everyday conversational data” or “naturally occurring talk” (Couper-Kuhlen & Selting 1996: 25). It can be described, however, as “genuine real-time communication” (Couper-Kuhlen & Selting 1996: 24).

Finally, one methodologically favourable aspect of collecting data from the TV News Archive lies in the independency of speakers and utterances. In production experiments, the test subjects, whose number is limited for trivial pragmatic reasons, are typically asked to produce the same utterance repeatedly in order to gather enough data points – a fact, that has to be accounted for carefully in the statistical analysis of such experimental studies. In our corpus, by contrast, there is a wider and fully independent range of speakers, which differs for each of the three questions. This means, that a certain speaker of question A will not – or only by chance – appear as a speaker of question B and C aswell. Occasionally there are utterances of the same question by the same speaker in our corpus. For example, the former US president Donald Trump appears more than once as the speaker of question B, and the Fox News moderator Bill O’Reilly appears several times as the speaker of question C. These repetitions, however, are not a necessary consequence of the study’s design. It rather is an artefact based on two things: First, the person has produced the utterance of interest more than once in TV, and second, the TV News Archive has provided the corresponding clips. And since these repetitions are an exception in our data rather than a regularity, all utterances can be treated as being independent.

Generally and despite some existing caveats, speech data from the TV News Archive, when processed with care, is a reasonable and in many aspects advantageous choice for studies on prosody.

### 4.1.2 The utterances

The basis of this study lies in the choice of the three clauses A, B and C, listed above. Before looking at the actual utterances of these clauses, some things can be already stated.

From a semantic point of view, all three clauses represent polar questions in that they define a set of two answers, *yes* and *no*, respectively. Again, this does not necessarily mean, that such an answer would be expected or even adequate in the context of the question's utterance. Anticipating some of the discussion on the functional aspects ahead, it is reasonable to assume, that a simple "Yes" in response to question A ("Can you repeat that?") would feel inappropriate in most contexts. The act of repeating the before said, on the other hand, would most probably be an appropriate response, even though it would leave the question unanswered.

Syntactically speaking, the polar questions come in the form of closed interrogatives, with the subject-auxiliary inversion being the main indicator here. The clauses can be said to consist of four constituents – auxiliary verb, subject, lexical verb and object, in that order –, each expressed through a single word. While there is variation regarding the auxiliary verb (either *can* or *do*<sup>19</sup>) and the lexical verb (either *repeat* or *believe*), the subject *you* and the object *that* are the same for all three clauses.

This lexical resemblance, which goes hand in hand with a phonological resemblance, is intended and serves the purpose of better comparability. All clauses consist of five phonological syllables, distributed to three monosyllabic words and the disyllabic lexical verb. The latter receives lexical stress on its second syllable, which has the long vowel /i:/ as its nucleus, both in *repeat* and in *believe*.

Everything said so far on the clauses A, B and C seems uncontroversial – at least when accepting the framework on questions and interrogatives of Huddleston (1994), which, as has been pointed out in chapter 2, is not the only reasonable way of approaching the phenomenon. But when it comes to the function, that these questions serve, if uttered in discourse, it becomes far more difficult. The general assumption is, that all three – at least

---

<sup>19</sup> There is a categorical difference between the two auxiliaries *can* and *do*. While the former is a modal auxiliary verb, which carries modality information, the latter is an auxiliary verb in its pure form, needed only to build a correct interrogative clause. This can be seen clearly, when forming the corresponding declarative clauses of B and C. The modal auxiliary *can* in B must be conserved to keep the meaning of B intact: *You can believe that*. In the declarative clause of C on the other hand, the *do* vanishes: *You believe that*.

in the context of TV broadcasts – typically are non-genuine. Non-genuine means, that they are not primarily uttered to get information from the addressee, i.e. to narrow down the set of the two possible answers, *yes* and *no*, to the presumably correct answer. Remember, for a question to be non-genuine, at least one of the three sincerity conditions introduced in section 2.5 must be violated.

The polar questions A, B and C, however, not only are expected mostly to be non-genuine, they also are expected to perform different indirect speech-acts each.

The illocutionary force associated with question A is that of a request for action, concretely the indirectly conveyed request to repeat a previously made utterance, which is referred to by the demonstrative pronoun *that*.

The illocutionary force associated with question B is that of a rhetorical question. It indirectly conveys what Huddleston (1994) would call an exclamatory statement, which here predicates a certain circumstance, referred to by *that*, as unbelievable, remarkable, extraordinary, outrageous or the like.

Finally, the illocutionary force associated with question C is that of a topic introduction. It indirectly conveys an invitation to the addressee to set out his thoughts on an issue, which exceed the simple answer of whether he believes a certain statement, referred to by *that*.

At this point an interjection is necessary. As explained before, there should be reasonable doubt about the assumption, that the illocutionary force of an utterance can be predicted by its lexical string. Both Lyons (1977) and Huddleston (1994) have raised this issue (s. chapter 2). Cameron et al. (1989) call it the *form and function problem*. They claim, that one cannot simply link a specific form, be it syntactic or intonational, to one specific discourse function. The latter is both variable and complex and one single utterance can – and normally does – perform multiple functions at the same time. This does not mean, that there is no relation between form and function whatsoever<sup>20</sup>. But it leads to the “absolute necessity of considering forms in their linguistic and social context” (Cameron et al. 1989: 77).

For this study, assumptions based on the author’s intuition can serve as a starting point for exploring the functional aspects of A, B and C, but not as evidence for concrete

---

<sup>20</sup> Sadock & Zwicky (1985) for example, whose concept of sentence type has been briefly introduced in section 2.4, speak of conventionalized use, which implies a common connection between form and function, but at the same time denies a necessary connection between the two.

functions of concrete utterances. So in order to respond to the *form and function problem*, two things must be done. First, the three speech-acts presumably conveyed by the questions A, B and C must be defined in more detail by setting up a number of sincerity conditions. And second, the actual utterances of A, B and C must be checked with regards to these conditions.

For polar questions A and B, this turned out to go rather smoothly, as the initial assumptions on their pragmatic function seem right – at least for enough utterances to build a corpus of well comparable speech-acts. In the case of question C things were less clear and the corpus of utterances appeared to be more heterogeneous.

### **A: Can you repeat that?**

The indirect request for action – here to repeat a specific utterance –, as which the polar question of A is assumed to function mostly, can be described by the following sincerity conditions.

1. The speaker desires the addressee to carry out the requested action, i.e. to actually repeat *that*.
2. The speaker expects the addressee to be able to do so; and thus, the speaker knows – or believes to know – the answer to the question, whether the addressee *can* repeat it.
3. An answer by the addressee, i.e. *yes* or *no*, is not expected, but optional.

Indeed, most of the utterances provided by the TV News Archive, that formally matched clause A, fulfilled these conditions. A concrete ratio cannot be given out of two reasons. First, because keeping track of all utterances of A with other primary functions would have meant an expenditure, that did not seem worthwhile. And second, because not all of the several thousand clips, that came up as results of the search string “can you repeat that” were actually looked at (cmp. section 4.2.1).



Important for this study is, that after all necessary exclusions a corpus of 91 utterances of question A could be built, that were considered requests for action<sup>21</sup>. These utterances took place mostly in interview situations, either face-to-face or via video transmission. Due to the inherent technical obstacles, the latter case is especially prone to acoustic issues, which often lead to requests for repetition. Other common situational contexts of the utterance were public hearings or press conferences.

A representative chunk of dialogue occurred on September 15, 2018, in the MSNBC rolling news program. With journalist Jonathan Allen, who is connected via video call, the anchorwoman Alex Witt discusses some allegations against Brett Kavanaugh, at that time designated Justice of the Supreme Court.

“Does this change the confirmation hearing outcome in the process?”, Alex Witt asks Jonathan Allen. And after a pause of more than 2 seconds, which might be due to transmission delay, he responds: “I’m sorry, Alex, can you repeat that?”<sup>22</sup>

Alex Witt then replies: “Yeah, in general, do you think, this will affect the ability for Brett Kavanaugh to be confirmed?”

In this case, both the request is complied with – if Alex Witt’s second utterance is interpreted as a paraphrased repetition of her first –, and the question is answered (“Yeah”). But neither is required for the utterance of Jonathan Allen to be considered as an indirect request for action, as the relevant criterion here is the illocutionary force, not the actual perlocutionary effect (cmp. Section 2.1). Crucial is, that all of the above stated conditions – by a conventional understanding – are fulfilled.

Interestingly, the dialogue exemplifies another phenomenon, that could be observed. When used as an indirect request, the phrase “Can you repeat that?” is often accompanied by some kind of excuse from the speaker. 65 out of 91 eventually considered utterances of A were preceded or followed by an overt excuse like “sorry” or “excuse me”, an additional 4 cases included at least an explanation, that might be interpreted as an excuse<sup>23</sup>. While question’s social costs like potential face loss are not the primary subject of this study, it can be stated that certain kinds of politeness strategies are definitely at play here. A thorough analysis of the data with regard to Levinson’s “social economics of

---

<sup>21</sup> It was not filtered for different varieties of English. Nevertheless, it can be stated, that all considered utterances of A, B and C were produced by native speakers, most of which were speakers of American English.

<sup>22</sup> This utterance of A can be found in the table of the digital appendix A12 under the ID a004.

<sup>23</sup> For this and other meta information compare section 4.2.1 and the table in appendix A5.

interrogatives” (Levinson 2012; cmp. Section 2.4) could be one promising continuation of this work.

### **B: Can you believe that?**

The rhetorical question, as which the polar question of B is assumed to function in most cases, can be described by the following sincerity conditions, which are loosely based on the ones from Dehé & Braun (2020) introduced in section 3.2.

1. The utterance of B has the feel of an assertion, which here can be paraphrased roughly as: *That is remarkable*.
2. It can, but does not have to be directed at a single addressee.
3. An answer, i.e. *yes* or *no*, is not expected, but optional.

Again, a huge portion of the inspected utterances of B fulfilled these conditions and thus were regarded as rhetorical questions.

In opposition to both A and C, a single direct addressee was not necessary here. So, in the case of B, the double meaning of *you*, which in English refers both to the personal pronoun in the second person singular and plural, has to be pointed out. It was decided, however, to ignore this homonymy as it is not expected to affect the utterance’s illocutionary force or prosody – upcoming studies are welcomed to prove this assumption wrong.

In fact, 40 out of the 86 eventually analyzed utterances of B were not directed at one concrete interlocutor. These cases occurred for example in political speeches in front of an audience or in news broadcasts, where the anonymous TV audience was addressed. Utterances of B with direct addressees took place mostly in dialogues between anchorpersons or in interviews.

One example of a rhetorical question performed by uttering clause B could be found in the *News at 8* on television station KRON 4, broadcasted in the evening of march 31, 2021. In the weather section of the program the moderator presents the current temperatures in California, which are unseasonably warm: “Still it is 70 degrees outside

right now in San José, 75 in Fremont, 75 still in downtown San Francisco. Can you believe that?”<sup>24</sup>

There is no answer to that question, not even a response. And as no one is there to respond – at least no other person can be seen in the frame – a response is clearly not expected. Instead, the weatherman continues without a break: “74 in Berkeley, 77 in St. Helena, 73 in Vallejo.” It takes roughly half a minute before another moderator takes the floor, after the weather report is over by schedule.

Notice, that an answer to B in the strict semantic sense, i.e. *yes* or *no*, wouldn’t inform the speaker on whether a potential addressee believes a certain statement, but whether he has the ability to believe it. This seems “either perfectly obvious or perfectly obviously unknowable” (Banuazizi & Creswell 1999: 10).

### **C: Do you believe that?**

Less clear was the case for utterances of C. To be fully transparent: Originally, they were assumed to function mostly as inquires, i.e. as genuine questions, sincerely desiring an answer on whether the addressee believes a certain statement. But after looking at the first chunk of instances of C as provided by the TV News Archive, this assumption was revised. It turned out, that a lot of the questions were uttered by interviewers not primarily to get a *yes* or *no* answer, but to give the addressee the floor to more extensively explain their view on the brought up issue.

Sometimes, these utterances have challenging character, putting the addressee into a situation, in which he has to defend a controversial stance. Sometimes they are more inviting, putting the addressee in the rewarding position to elaborate on his seemingly good ideas. A lot of cases, however, lie somewhere in between, and often a classification into one category or the other is not possible offhand.

What the utterances have in common is, that they introduce new topics or new aspects of certain topics to a conversation. This resembles what Hedberg et. al (2017) write about the data from Hedberg & Sosa (2002), who analysed polar questions from a comparable source, i.e. from public affairs discussion programs (cmp. Section 3.3). Banuazizi & Creswell (1999) include polar questions, that function as introductions to assigned or new

---

<sup>24</sup> Utterance ID: b063

topics, in their list of non-genuine questions. They explain: “[The speaker] may wish for [the hearer] to provide an extended answer, or often [the speaker] will use her question as a springboard for making further statements of her own” (Banuazizi & Creswell 1999: 8). The latter, however, was not the case for many of the utterances in our corpus – one indicator of that is, that 61 out of 72 relevant utterances of C led to an immediate floor change.

An example of such a topic introduction can be found in the FOX News show *The O'Reilly Factor*, aired on December 19, 2014, when a clip from an older interview with Russel Simmons is played. The co-founder of the hip hop label Def Jam here discusses racism and police violence against Black people with show host Bill O'Reilly. The latter provocatively says: “I believe the strife between poor Blacks and the police is driven by Black crime. That’s the nexus point. Do you believe that?”<sup>25</sup>

Russel Simmons replies: “No. I believe, that we can create a good relationship between police and community with sensitivity training both for the community and for the police.” He then goes on explaining his understanding of the issue and his ideas to solve it for about half a minute before being interrupted by O'Reilly.

In this case an answer (“No”) is given and might even be expected by Bill O'Reilly. But importantly, this answer does not seem like the primary goal of the utterance of C. Instead, the question seeks for further explanations from Russell Simmons. Its function here might rather be that of a (genuine) variable question, which could be paraphrased as “Why do or don’t you believe that?” or “What do you think of that?”.

Additionally, it can be suspected, that the moderator Bill O'Reilly asks the question, especially because he believes to know, that his interlocutor will disagree with its proposition. This would mean a violation of the first sincerity condition for genuine questions (s. section 2.5) – but of course, one can never know for sure, what people know or believe to know. And it must be emphasized, that not all utterances of C suggest such an interpretation. In a lot of cases there are no indications on whether the speaker already knows – or believes to know – the answer to his question. Thus, one might argue, that some of the utterances of C, while expecting more than a *yes* or *no*, still fulfil all sincerity conditions of a genuine question. These cases highlight, that a single utterance can – and normally does – fulfil several pragmatic functions at the same time.

---

<sup>25</sup> Utterance ID: c006

For our purposes, the following approach was applied: If the primary function of a question is interpreted to be different from that of an inquiry, it is considered non-genuine. After all, one should keep in mind, that the theory of speech-acts is a made-up concept consisting of made-up categories like illocutionary force. It was developed to gain better understanding of human language and communication, and it should not be confused with what it is built to understand, i.e language and communication itself.

If inquiry force of questions would be conceptualized as a continuum between genuine and non-genuine rather than a strict dichotomy – as the works on prototypicality (Haan 2002) or questionhood (Levinson 2012), introduced in section 2.4, imply –, the topic introductions of C could be considered more genuine than the requests for action of A and the rhetorical questions of B. This assumption is supported by the fact, that 47 % (34 out of 72) of the eventually analysed utterances of C actually received an answer, i.e *yes* or *no*, while it was only 23 % (21 out of 91) for A and 8 % (7 out of 86) for B<sup>26</sup>.

Interestingly, the distribution of affirmative and negative answers also varies from speech-act to speech-act. While the request for action shows a *yes* bias (20 affirmative answers, 1 negative answer)<sup>27</sup> and the rhetorical question shows a *no* bias (all 7 answers are negative), the answers in response to the topic introduction are equally distributed (17 affirmative, 17 negative). This does not necessarily mean, that the utterances of C are sincere in aiming for an answer, which is unknown to the speaker, as the example of Bill O'Reilly has shown. But it might be interpreted as supporting the idea of C being more genuine – or: less non-genuine – than A and B.

We keep that thought in mind, but for now proceed with the categorical distinction, considering only utterances interpreted to be non-genuine by the criteria stated above. Consequently, the utterances of C make up a rather heterogenous corpus of utterances, which are grouped together under the somewhat vague label of topic introduction. This kind of speech-act includes all kinds of biases in favour or in disfavour of the addressee

---

<sup>26</sup> The case of B, however, is not properly comparable to A and C, as 40 out of 86 utterances were not directed at a single addressee – a circumstance, that would have been an exclusion criterion for utterances of A and C.

<sup>27</sup> Note, that this count is based on a purely formal analysis. It is reasonable to assume, that an answer to question A means an approval or denial of the indirect request, rather than an affirmation or negation of the question's semantic proposition. Thus, a *no* answer does not imply, that the addressee is not able to repeat the before said. Rather, the addressee denies to carry out the indirectly requested action to repeat it. One might argue, that such responses shouldn't even be regarded as answers. Similar considerations can be made regarding answers to the rhetorical question B (s. above).

and all kinds of implications towards an affirmative or negative answer. The following is an attempt to draw up sincerity conditions as done for A and B before.

1. The utterance of C introduces a new topic or a new aspect of an already established topic to a conversation.
2. A response is expected, which can but doesn't have to include an answer, i.e. *yes* or *no*.
3. It has the character of a variable question, which can roughly be paraphrased as: *What do you think of that?*

## 4.2 Method

### 4.2.1 Evaluating, labelling, annotating

The work on the data started with watching clips, that the TV News Archive provided as results for the search strings of A, B and C, and evaluating the corresponding utterances regarding their relevance for the study. The clips included a lot of repetitions and formally irrelevant utterances, which were ignored right away. An example of the latter would be an utterance, which includes the relevant string, but doesn't syntactically end on the word *that* as in "Do you believe, that it is summer already?"<sup>28</sup>.

Only a fraction of the many clips – over 30.000 for search string A, over 300.000 for search string B and over 500.000 for search string C – were watched. The process of watching and evaluating was stopped after collecting ca. 100 formally relevant utterances of each question, from which some were excluded afterwards, e.g. due to accidental repetitions of the same utterance, special focus conditions (s. section 4.2.2) or phenomena, that disturb phonetic analysis like speech overlap or background noise.

As the concrete structure of the relevant speech-acts and therefore their sincerity conditions evolved in the process of looking through the data – with the biggest changes of course in the case of question C (s. section 4.1.2) – some formally flawless utterances were excluded afterwards due to functional aspects. One example of such a belated exclusion is an utterance from the ABC show *Eyewitness News*, aired on February 11,

---

<sup>28</sup> Syntactically, *that* is not a demonstrative pronoun here, but a conjunction.

2016. Here a sports journalist gives a long explanation of a new and seemingly complicated rule in car racing sports. After finishing, he is addressed by his co-host, who jokes: “I heard everything you said, but just in case: Can you repeat that?”<sup>29</sup>

This is not a genuine question and might seem like an indirect request for action at first. But this request for action is not genuine itself as it does not seek for the requested action to be actually carried out. This means a violation of the first sincerity condition for an utterance of A to count as a request for action (s. section 4.1.2), which consequently leads to exclusion. While such jokes and ironic utterances, which knowingly play with their different layers of non-genuine character, are very interesting from a pragmatic perspective, they are not the subject of this very study.

Every utterance, that was initially chosen to be relevant, was given a unique identification code. And next to information on their illocutionary force (A, B or C) all utterances were also labelled for some characteristics, which were assumed to be of possible value later on. These labels included information on:

- The sex of the speaker (superficially categorized into *m* or *f* by the author)
- The broad situational context of the utterance (e.g. interview, news, speech, hearing)
- Whether or not a floor change occurs immediately after the utterance
- Whether or not there is a verbal response at all, whether or not such a response includes an answer, and whether such an answer is affirmative (*yes*) or negative (*no*)

Utterances of A received additional labels on:

- Whether or not the utterance is preceded or followed by an excuse or explanation from the speaker
- Whether or not the requested action (to repeat *that*) is carried out by the addressee

Utterances of B received an additional label on:

- The type of addressee (i.e. direct or indirect)

This last information was not of interest for utterances of A and C as they are directed at one concrete interlocutor by definition (s. section 4.1.2).

---

<sup>29</sup> Utterance ID: a006

Ultimately, not all of the collected meta information came to actual use. The speaker's sex for example did not seem to be relevant for our purposes. And while it is an interesting fact, that the speakers of the 249 utterances, that finally made it to prosodic analysis, were male in 172 cases and female in only 77 cases, conclusions on representation inequalities in American television should be left for more profound social studies.<sup>30</sup>

In a next step, all utterances of A, B and C were manually segmented into syllables by using the phonetics program *praat* (Boersma & Weeninks 2023). Figure 4 exemplary shows the oscillogram, the F0 trajectory and the syllable segmentation of an utterance of A.

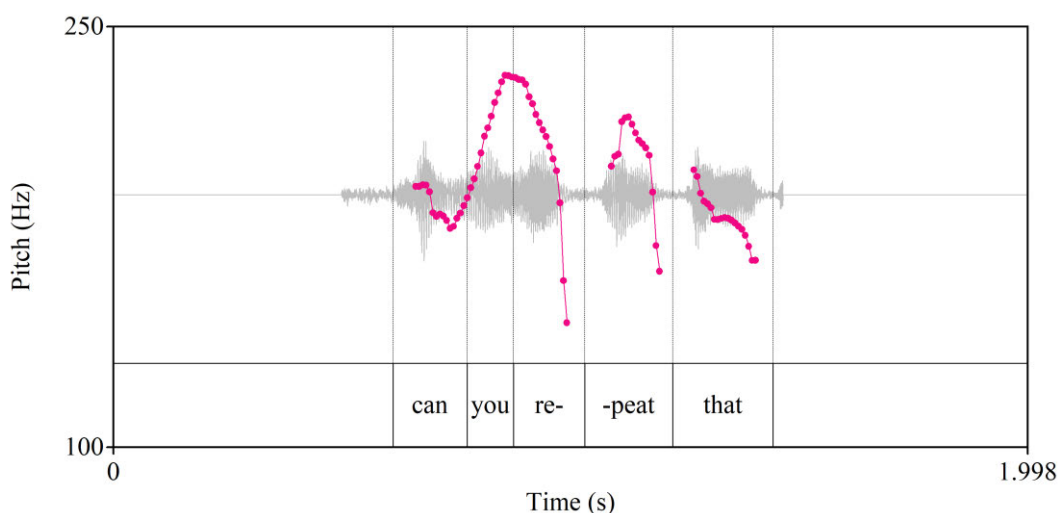


Figure 4: Oscillogram (grey), F0 contour (pink) and syllable annotation of an utterance of A. Utterance ID: a020.

After the corpus was cleared of utterances, that did not bear the acoustic quality for further phonetic analysis, a provisional group of 278 utterances was left (95 utterances of A, 98 utterances of B and 85 utterances of C). While these utterances were both technically flawless and relevant from a pragmatic perspective – i.e. they were considered requests for action, rhetorical questions or topic introductions in terms of the presented sincerity conditions (s. section 4.1.2) –, one issue remained: their focus condition.

<sup>30</sup> Concrete numbers for all of the above mentioned categories can be drawn from the table in the appendix A5.



#### 4.2.2 The issue of focus

Focus, a linguistic category established to describe differences in information structure, marks new or important information within an utterance as opposed to given or expected information. It has been shown to affect prosody in English, e.g. in terms of intonation, duration or intensity (Breen et al. 2010). Thus, the issue of focus needs to be addressed, when comparing other linguistic categories like different speech-acts with regards to their prosody. Typically, one distinguishes between narrow focus, where the scope of the focus is limited to one single constituent, and broad focus, where the scope includes a whole phrase.

As narrow focus is reported to be especially salient prosodically, i.e. realized with higher F0, greater intensity or longer duration (Breen et al. 2010), it was decided to concentrate on narrow focus and ignore broad focus for the current study. On that basis, the author went through all utterances again and decided, whether one of the four syntactic constituents (auxiliary, subject, lexical verb, object) receives narrow focus from the speaker and – if so – whether this focus is perceived as either mild or strong.

From the 278 utterances, 111 were perceived as having narrow focus on one constituent, mostly on the lexical verb (33 with mild narrow focus, 17 with strong narrow focus), but also on the subject (25 mild, 8 strong) and object (22 mild, 3 strong). The auxiliary received narrow focus in only 3 cases (2 mild, 1 strong). 26 cases of narrow focus (4 of which were strong) occurred in utterances of A, 38 (12 strong) in B, and 47 (13 strong) in C.

The inclusion or exclusion of the different focus conditions did not turn out to radically change the overall picture of the results of the prosodic analysis. But for the sake of clarity and to avoid possible statistic artefacts, it was decided to exclude all utterances with strong narrow focus from the study anyway. This left the final corpus with a total of 249 utterances (91 of A, 86 of B, 72 of C)<sup>31</sup>.

---

<sup>31</sup> The results for different data sets – a) including all utterances regardless from their focus condition, b) excluding all utterances with strong or mild narrow focus, c) including only utterances with strong or mild narrow focus – can be found in the appendix A6–A8.

### 4.2.3 Prosodic analysis with periodic energy

To investigate general prosodic characteristics of the non-genuine polar questions as well as differences between their three concrete types, an analytical method called *ProPer* was used. *ProPer*, which is short for *prosodic analysis with periodic energy*, was developed and made available as an open source project on OSF by Albert et al. (2023).

From a practical perspective, it provides a workflow for speech data, based on *praat* and *R* (R Core Team 2023) scripts. And as the name suggests, its main and distinctive analytical feature is, that – next to fundamental frequency (F0) – it takes into account periodic energy as a crucial source of linguistic information.

#### 4.2.3.1 Rationale

For decades most studies on speech prosody have focused in one way or another on one specific description tool: fundamental frequency over time. In fact, the presented studies from section 3 are no exception to that. And while a lot of insights on human language could be gained by that method – the described differences between genuine and non-genuine polar questions in terms of terminal pitch movement is only one of many examples – it leaves unexposed at least one important aspect: the degree of pitch intelligibility.

The F0 trajectory of an utterance provides information on the presence or absence of periodicity, i.e. voiced or vocalic sounds, and its frequency (as measured in Hz) at any point of time. Compare for example the F0 trajectory of “Can you repeat that?” in figure 4 from section 4.2.1. It starts after the initial voiceless stop of the auxiliary verb *can* and then goes on without interruption until the end of the first syllable of the lexical verb *repeat*, while showing a rise over the subject’s syllable *you* and a fall after that. The two breaks in the trajectory, before and after the vowel of the second syllable of *repeat* can easily be explained by the voiceless stops in the onset and coda of that syllable. The contour then ends in a fall over the object’s single syllable *that*.

This depiction, however, cannot account for the fact, that not all of the periodic parts of a signal, i.e. those parts where we see the F0 contour, are perceived with equal ease and thus are equally important for the processing of speech (Albert et al. 2018).

The approach of *ProPer* tries to solve this issue by adding another dimension to the analysis: periodic energy. “Periodic energy reflects the degree to which pitch is intelligible, a higher value representing a stronger F0 signal that is consequently more easily perceived” (Aviad et al. 2018: 804). Simplifying from the technical details, one can say, that periodic energy describes the intensity of pitch.

By taking into account periodic energy, those parts of speech, in which F0 is more relevant for the recipient can be identified and therefore distinguished from those parts, where F0 is less relevant. Importantly, periodic energy does not provide a dichotomy between relevant and irrelevant pitch, or between presence and absence of pitch, as shown by a simple F0 trajectory. It rather describes pitch in terms of its degree of strength on a continuous scale.

Trajectories of periodic energy, by their nature, “show peaks in correspondence with vocalic sounds, and troughs in correspondence with non-vocalic sounds. As such, they can be used as a proxy for syllabic cycles, thus reducing the necessity for segmenting the text” (Cangemi et al. 2019: 1). Concretely, the local minima of a periodic energy curve, which can be located by its derivatives, reflect the boundaries of periodic fluctuations and are supposed to correspond with syllabic boundaries. So in addition to exposing pitch intelligibility, periodic energy offers a formal and non theory driven way of segmenting speech data into syllables, which is less time consuming and more definite than manual segmentation. Working practice has shown, however, that these periodic energy fluctuations do not always correspond to phonologically defined syllables. Being a purely phonetic measure, periodic energy reflects the continuous acoustic signal of spoken language, rather than meaningful linguistic categories offhand.

While F0 remains a powerful description tool in speech analysis, periodic energy can enrich and adjust its findings. Especially when zooming in on a very specific linguistic topic – like the prosodic comparison between different kinds of non-genuine polar questions – fine-grained tools of description might be necessary to discern subtle differences. By combining F0 with periodic energy, *ProPer* provides a range of such tools.

#### 4.2.3.2 The measurements

The final *ProPer* analysis comprises different kinds of measurements, all of which rely at least partly on periodic energy. Two of these measurements are especially interesting for our study: mass and synchrony<sup>32</sup>.

Mass – or: mass of periodic energy – describes the area under the periodic energy curve, i.e. its integral, for a certain syllable in relation to all syllables of the relevant utterance. The mass value is interpreted “as reflecting the relative strength of each periodic cycle” (Albert et al. 2018: 806) and thus the relative periodic energy strength of each syllable<sup>33</sup>. A mass value of 1.0 reflects the mean mass of an utterance’s syllable. Therefore, syllables with mass values higher than 1.0 have an above-average periodic energy strength, syllables with mass values lower than 1.0 have a below-average periodic energy strength. Note, that both periodic energy and syllable duration – more precisely: the duration of the periodic parts of a syllable – affect the measurement of mass.

The second measurement of *ProPer* is the synchrony between the center of gravity (CoG) and the center of mass (CoM). The CoG, as proposed by Barnes et al. (2012a, 2012b), marks a time value, which separates the area under a given F0 contour (subtracted by a set baseline) into two halves of the same size. By that procedure, one can account both for general direction of the F0 contour, i.e. whether it is rising or falling, and for its concrete shape, i.e. whether it is concave or convex. For example, a rising contour would lead to a relatively later CoG, while a falling contour would lead to a relatively earlier CoG. By the same logic, a concave rising contour would lead to later CoG than an otherwise comparable convex rising contour. Of course, F0 contours of speech signals comprise all kinds of changes of shape and direction within certain time intervals. Importantly, the CoG accounts for all of that as it does not primarily look at the contour itself, but at the integral below.

While the CoG applies to F0, the CoM is the equivalent for periodic energy, i.e. a time value, that divides a certain interval into two equally sized halves of periodic energy mass. The regions of interest, for which both CoG and CoM are calculated, are the same

---

<sup>32</sup> Another common measurement of the *ProPer* method is delta F0. As its analysis did not prove very fruitful for our data, it is ignored here. A depiction of the results of the delta F0 analysis can be found in the appendix A9.

<sup>33</sup> As described in section 4.2.3.1, a periodic cycle can be seen as a phonetic correlate of the phonological category syllable.

syllables, which are taken for the calculation of periodic energy mass. An explanation of how the syllable boundaries are determined concretely will be given in the following section 4.2.3.3. As syllables roughly correspond with periodic energy cycles, with troughs at their onset and coda and a peak in their nucleus, the CoG is – with natural variations – placed around the middle of a syllable. Of course, it is these variations, that make the CoM relevant in the first place, as they reflect the periodic energy characteristics of specific syllables.

Finally, the synchrony between a syllable's CoG and CoM describes the distance between the two points and can be interpreted as the relation between text and tune in perception (Cangemi et al. 2019: 2). A value of 0.0 reflects perfect synchrony, which is perceived neither as rising nor as falling pitch. If the synchrony value is positive, i.e. the CoG occurs after (or: on a time axis right from) the CoM, the syllable is considered to have a perceived rising pitch. On the other hand, if the synchrony value is negative, i.e. the CoG occurs before (or: on a time axis left from) the CoM, the syllable is considered to have a perceived falling pitch. As synchrony describes a distance on a time scale, it can be expressed as an absolute duration in ms. But for better comparability, it is typically expressed as a relative value, i.e. in percent of the duration of the corresponding interval (Albert 2023: 156).

The functions of the two described measurements can be summed up as reflecting the relative prosodic strength of a certain syllable (mass) and the pitch change within a single syllable (synchrony). It should be emphasized once again, that adding periodic energy to the analysis of pitch movement brings the advantage of taking into account not only F0, but also its intelligibility.

A third measurement, that was looked at, is utterance duration. While this is not an idiosyncratic *ProPer* measurement as it is not directly related to periodic energy<sup>34</sup>, it can be analysed easily within the *ProPer* workflow.

---

<sup>34</sup> There is an indirect relation between utterance duration and periodic energy, though, as the location of the syllable boundaries are based on the troughs in the periodic energy trajectory (s. section 4.2.3.3). And obviously, the start point of the first syllable and the end point of the final syllable of an utterance at the same time mark the start and end point of the whole utterance.

#### 4.2.3.3 Syllable boundaries

The syllable, as the place of the measurements of mass and synchrony, is crucial for the *ProPer* analysis. Therefore, the locations of the syllable boundaries directly affect the results and should be determined with great care.

The conducted manual segmentation, which relies on the perception of phonological syllables rather than acoustic features of the signal, was taken as a basis for that endeavour. But as already mentioned, *ProPer* also provides a way of automatically detecting syllable boundaries at the troughs of the periodic energy curve. And to benefit from these formal computations, too, it was decided to take a reasonable hybrid of the manual and the automatic segmentation. Concretely, the following procedure was conducted.

Automatically detected boundaries were selected, if they did not differ more than a fixed time span from a manually detected boundary. Trial and error revealed 25 ms to be an adequate value for this time span. Manually detected boundaries were selected, if they had no automatically detected boundaries within the preceding or following 25 ms. Automatically detected boundaries, that did not fall within this 25 ms region of any manually segmented boundary, were ignored.

#### 4.2.3.4 Periograms

An exemplary outcome of the semi-automatic syllable detection can be seen in figure 5, which is an abstract depiction of the same utterance shown in figure 4. The dotted vertical black lines show the original manual segmentation. The solid vertical red lines show the adjusted boundaries, determined in the above mentioned way. In this case, the difference from the manual segmentation concerns mostly the start and end point of the subject *you* (which of course correlate with the end point of the preceding auxiliary verb *can* and the start point of the following first syllable of the lexical verb *repeat*, respectively).

Fairly new is the way, in which pitch is depicted here. Instead of just showing F0 over time, as in figure 4, the blue trajectory shows the interaction between F0 and periodic energy over time. This is done by changes of colour transparency and width of the F0 curve, with less transparency and a thicker line indicating greater periodic energy.

Periograms, as these multidimensional depictions of F0 and periodic energy are called, are another handy innovation of the *ProPer* method.

Additionally, the other discussed *ProPer* parameters are also shown in figure 5. Mass is depicted as the integral under the red periodic energy curve for each syllable. The relative mass values can be read at the bottom (in red). The dashed vertical red lines show the locations of the corresponding CoM. The CoG are marked by the solid vertical blue lines. Finally, the distance between corresponding CoM and CoG, i.e. their synchrony, can be read from the purple labels, both in absolute (ms) and in relative terms (%).

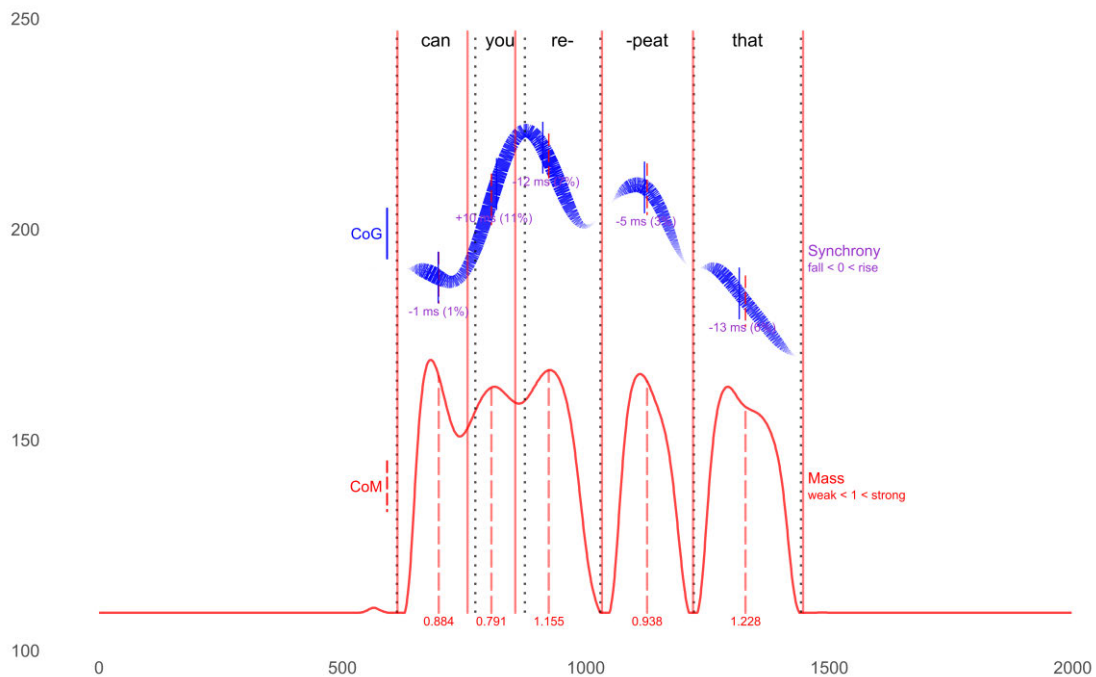


Figure 5: ProPer depiction of an utterance of A. The x-axis shows time values in ms. The y-axis shows pitch values in Hz, which relate to the periogram (blue), but not to the periodic energy curve (red). Utterance ID: a020.

A good reading example of the figure can be exercised for the final syllable *that*. It has a relative mass value of 1.228, which means that it is 22.8 percent above the average syllable mass and thereby the syllable with the greatest prosodic strength of that utterance. As can be seen quite clearly, the CoG (solid vertical blue line) occurs before the CoM (dashed vertical red line). The distance between the two is -13 ms – or 6 % of the syllable duration (purple label). The negativity of synchrony is associated with received falling pitch, which one can see in the periogram (blue trajectory) as well.

## 4.3 Results

### 4.3.1 Mass

Figure 6 shows the relative mass values for each syllable of every utterance of A, B and C. Within each syllable column on the x-axis each dot represents one specific utterance and by its placement on the y-axis shows its relative mass value. Remember, that a relative mass value of 1.0, here marked by the solid horizontal black line, means the average mass of all syllables of an utterance. As the legend explains, dots representing an utterance of A (request for action) are red-coloured, the colours for B (rhetorical question) and C (topic introduction) are yellow and blue, respectively. The violins depict the distribution of all data points of one utterance type within a syllable. They have larger width in highly frequented areas of the y-axis and smaller width in less frequented areas. The crosses mark the mean values for each utterance type in the corresponding syllable. By definition, all means of a given utterance type add up to a sum of 5.0 – because every utterance consists of five syllables with an average mass of 1.0. Differently put: The mean of the means necessarily is 1.0, too. So when looking at the means, it is interesting, whether they deviate from the average value 1.0, and if so, to which degree and in which direction.

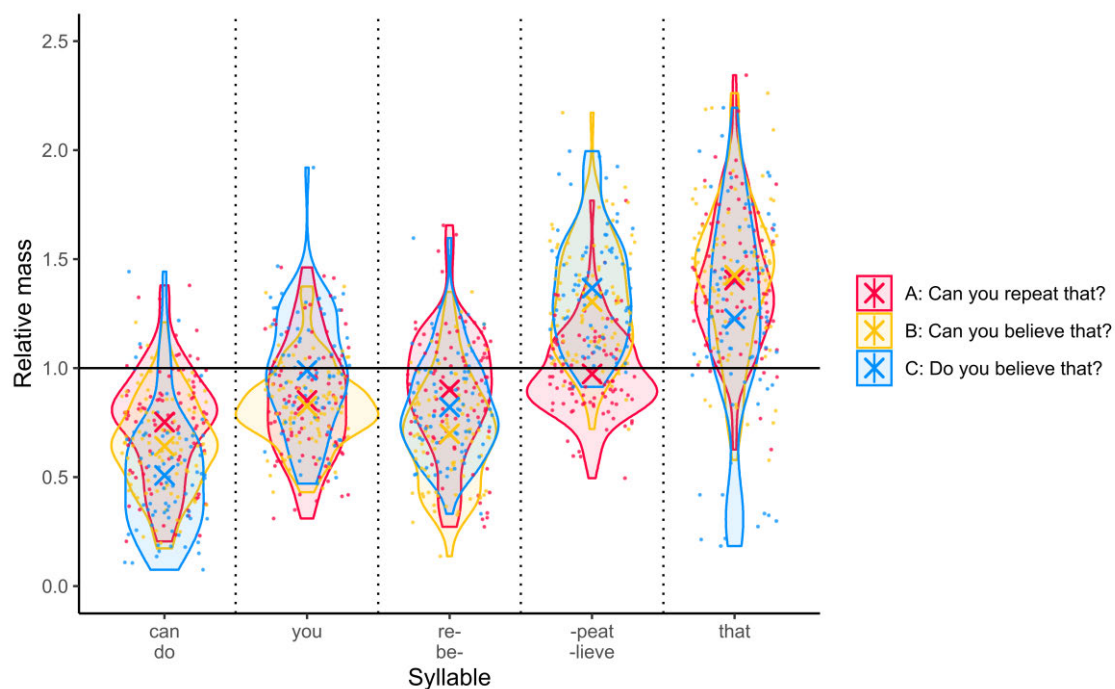


Figure 6: Relative mass values, sorted by utterance type (A, B, C).



Regardless from the utterance type, the utterance final syllable, i.e. the object *that*, receives the highest mass values on average. It can thus be understood as being prosodically prominent and relevant for pitch perception. This syllable seems suitable to investigate final pitch movement, the crucial feature of the studies presented in chapter 3, which will be done in the following section in terms of synchrony.

One should notice, however, that the type of utterance is not fully irrelevant in the object syllable. While the request for action A and the rhetorical question B show quite similar distributions with means of 1.40 and 1.43, respectively, and few mass values below 1.0, the topic introduction C behaves a little different. Its mean value is dragged down to 1.23, mainly by a group of utterances with mass values of 0.5 or lower. So in opposition to the other utterance types, the topic introduction does not receive its highest mean value in the syllable *that*, but in the preceding syllable, the ultimate syllable of the lexical verb *believe*.

To test, whether these differences between utterance types in the object are statistically significant, a linear model of mass as a function of utterance type was constructed. And indeed, significant is both the difference between A and C ( $F(2,246)=6.642$ ,  $p<0.01$ ) and the difference between B and C ( $p<0.001$ ). But as the corresponding adjusted  $R^2$  value is quite low (0.04), which means that very few of the data variance can be accounted for by the utterance type, these differences should not draw too much attention in our case<sup>35</sup>.

More interesting are the differences within the ultimate syllable of the lexical verb (*repeat/believe*). As this syllable is lexically stressed, it is prone to show relatively high mass values and thus be prosodically relevant. As one can see, this is the case for the rhetorical questions and the topic introductions with mean values of 1.30 and 1.37, respectively, but not for the requests for action, the mass values of which are distributed around a mean of 0.97. The corresponding linear model is significant both for the difference between A and B ( $F(2,246)=67.39$ ,  $p<0.001$ ) and the difference between A and C ( $p<0.001$ ). And with an adjusted  $R^2$  value of 0.35, the fixed effect, i.e. utterance type, accounts for the variance in the data quite well.

It has to be pointed out, however, that in our data the functional difference between utterance type A on the one hand and utterance types B or C on the other hand, correlates with a difference in the phonemic material of the lexical verb. Concretely, the syllable

---

<sup>35</sup> It should be kept in mind, however, that the categorization into utterance types as done here, is not set in stone. One might speculatively link the low outliers of C to undiscovered pragmatic differences within the group of topic introductions, which was found to be rather heterogenous in some respects anyway.

/pi:t/ in *repeat* bears less voiced phonemes than the syllable /li:v/ in *believe*, which might lead to a naturally shorter duration of the periodic part and thereby affect the mass of periodic energy. Nevertheless, a natural durational difference of that kind seems unlikely to be big enough to account for the observed difference in mass. The fact, that vocalic parts of a syllable can be quite flexibly stretched in time and amplitude according to the speaker's needs, is expected to outweigh the effects of phonemic differences, especially so in otherwise comparable utterances like the ones from our data.

An actual look at the durations of the periodic part of the verb-final syllable, again in relation to the other syllables, supports this hypothesis<sup>36</sup>. The distributions of the three utterance types are closer to each other here, with mean values of 1.16 for A, 1.24 for B and 1.31 for C. Most strikingly, the mean value of A is much higher – and by the way clearly above 1.0 – than in the measurement of relative mass.

Thus, even though there are differences in the relative duration of the periodic parts of the syllable – which can, but do not necessarily need to be based on the lexical material –, these differences do not account for the differences in periodic energy mass alone. The amplitude of periodic energy, which should not be affected significantly by the concrete phonemes, evidently plays an important role here, too. It therefore can be suspected, that in our case the functional differences of the utterances are to some degree reflected in differences of the relative mass in the position of the lexically stressed syllable of the lexical verb, with the request for action receiving significantly lower values on average. This observation, however, should not be generalized offhand for all kinds of non-genuine polar questions with the illocutionary force of a request for action. Larger studies would be needed to verify or adjust this exploratory developed hypothesis.

While the subject *you* and the first and unstressed syllable of the lexical verb (*re-/be-*) don't show unexpected tendencies for any of the three utterance types – with mass values around and below the average of 1.0 – the utterance-initial auxiliary verb is worth a closer inspection. With the lowest overall mass values, it might seem rather negligible at first. But in fact it shows some interesting differences between the utterance types. All three of them show similarly, while not identically shaped violins, which differ slightly, but clearly in their mean mass values: 0.75 for A, 0.64 for B and 0.51 for C. It is especially the

---

<sup>36</sup> A figure of the relative durations of the periodic parts – in the same style as figure 6 – can be found in the appendix A10.

difference between the requests for action of A and the topic introductions of C, that catches one's eye and which indeed is statistically significant ( $F(2,244)=19.82, p<0.001$ ). An adjusted  $R^2$  value of 0.13 suggests, that utterance type as a fixed effect is not without impact here.

But again, this observation comes with a caveat, as the functional difference between utterance type A and C correlates with a difference in the segmental makeup of the auxiliary verb. Not only do *can* and *do* differ in their phonemic material, they also differ in their syllabic structure with the latter lacking a consonant in coda position. In combination with *you*, *do* often gets reduced in natural speech, so that the two form a single syllable, orthographically expressed as “d’you”. For better comparability, in the annotation process such cases were segmented into two separate syllables anyway, leaving the auxiliary with few more than its initial stop sound<sup>37</sup>. This contraction of the auxiliary and the subject, however, is a phenomenon of segmental phonology and rather independent from aspects of illocutionary force.

And looking at the relative durations of the periodic part of the syllable cannot alleviate the concerns here, as the distributions resemble the ones from the mass measurement<sup>38</sup>. So while the observable mass differences between utterance types in the auxiliary verb are worth mentioning, their meaning for this study remains unclear.

### 4.3.2 Synchrony

Figure 7 shows the results for synchrony in the same way figure 6 does for mass. Here, the y-axis represents relative synchrony, i.e the distance between the CoG and the CoM in percent of the syllable duration. The solid horizontal black line marks the value of 0.0, which represents perfect synchrony between the CoG and the CoM, reflecting a pitch pattern, that is perceived neither as rising nor as falling. Positive synchrony values, placed above that line, represent rising pitch. Negative synchrony values, placed below that line, represent falling pitch.

Importantly, the means, which again are marked by crosses, do not need to add up to 0.0 or any other determined value. Other than relative mass, which relates to all syllables

---

<sup>37</sup> This circumstance by the way might explain to some part the slightly bigger relative mass of utterance C in the subject syllable *you*.

<sup>38</sup> Cmp. the figure in the appendix A10.

of an utterance, the relative synchrony of a given syllable is independent from other syllables of the utterance. Thus, it might be the case, for example, that every syllable of an utterance shows a positive synchrony value, which then would indicate, that pitch rises throughout the whole utterance. The same, of course, goes vice versa for negative synchrony and falling pitch.

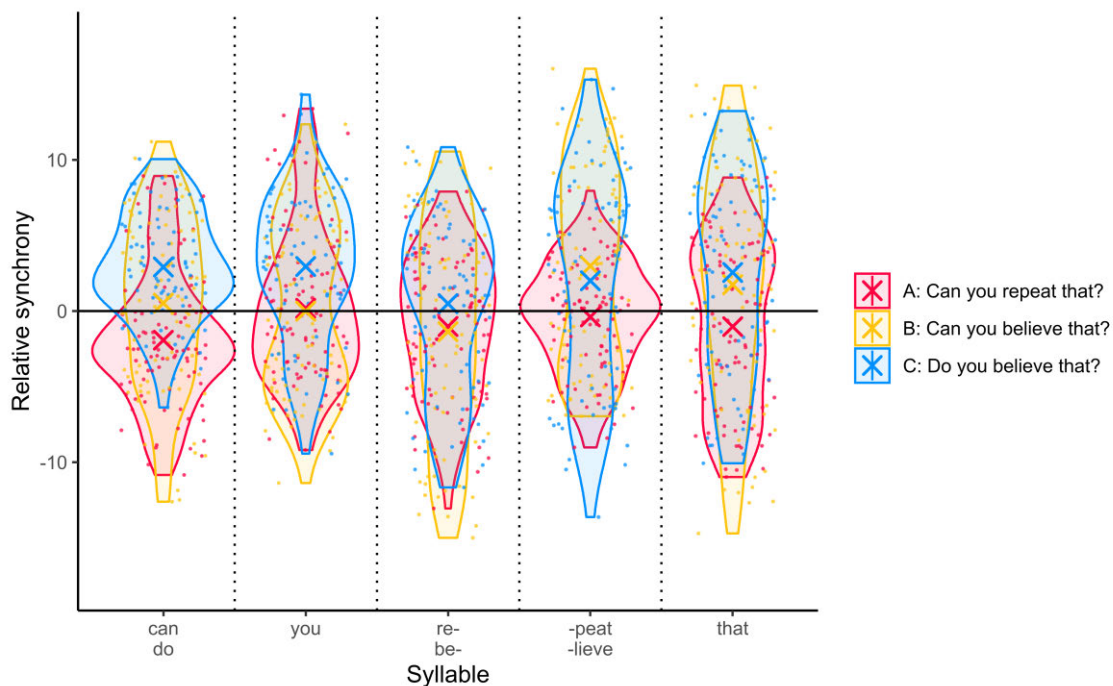


Figure 7: Relative synchrony values, sorted by utterance type (A, B, C).

By a first broad look at figure 7 the results appear rather fuzzy. There are hardly syllables, that stand out within one utterance type by showing clear tendencies towards rising or falling pitch. On the contrary, all syllables of all utterance types exhibit both positive and negative synchrony values, with only gradual differences in their ratios. Nevertheless, a closer look at certain syllables lead to some insights.

The first syllable of interest again is the utterance-final syllable, i.e. the object *that*. As the relatively high mass values suggested (cmp. figure 6), this syllable's synchrony values can be taken as a reasonable indicator of final pitch movement. And the fact, that none of the three utterance types shows exclusively – or even nearly exclusively – final rising pitch, is well in line with the previously discussed studies on polar questions. As explained in chapter 3, Banuazizi & Creswell (1999), Hedberg et al. (2017) aswell as Dehé & Braun

(2020) unanimously claim, that such a strong tendency towards final rises is a characteristic of genuine polar questions. So when abstracting from the different utterance types, our results can already be seen as supporting the former studies in showing, that non-genuine polar questions don't show this tendency.

An issue, on which the studies presented unclear outcomes, is, whether non-genuine polar questions tend to final falls or don't show clear tendencies at all. As can be seen easily in figure 7, our results support the latter hypothesis. In that, they resemble the results from Banuazizi & Creswell (1999), who still drew different conclusions (cmp. section 3.1.1).

The eyeballed non-tendency can be exemplified, when categorizing each utterance of our data set into either finally rising or finally falling by the plus or minus sign of its synchrony value in the object syllable – a procedure, which simplifies from the differences between smaller and bigger synchrony values, and thus from the degree of steepness of the F0 slope. The utterances of A consist of 48 falls (52.7 %) and 43 rises (47.3 %), the utterances of B consist of 32 falls (37.2 %) and 54 rises (62.8 %), and the utterances of C consist of 27 falls (37.5 %) and 45 rises (62.5 %). These numbers, of course, should be taken as a very broad estimation, rather than absolute results. The established dichotomy between falls and rises, with a synchrony value of 0.0 as the ultimate threshold, just serves the purpose of illustrating a more complicated pattern. Notice for example, that final level contours are not even considered here. Nevertheless, the ratios underline, that there is no overall tendency towards falling pitch in the group of non-genuine polar questions.

At the same time, a fine difference between the requests for action (A) on the one hand, and the rhetorical questions (B) and the topic introductions (C) on the other hand can be seen, which is also reflected in their mean values. While the utterances of B and C have positive mean synchrony values of 1.75 and 2.55, respectively, in the object syllable, the utterances of A have a negative mean value of -1.03. The linear model of synchrony as a function of utterance type shows, that the difference between A and B is significant ( $F(2,246)=7.149$ ,  $p<0.01$ ), and even more so is the difference between A and C ( $p<0.001$ ). But with an adjusted  $R^2$  value of 0.05, the role of utterance type should not be overestimated here.

And yet, one interesting thought can be derived from that. In section 4.1.2 the possibility of conceptualizing genuinity of questions as a continuum rather than a categorical dichotomy has been discussed. And by the positions of the mean synchrony values in the object syllable, it could be speculated, that requests for action are less genuine than rhetorical questions and topic introductions.

The topic introductions of C already were suspected to be more genuine by the insights of their pragmatic analysis in section 4.1.2. And indeed, they show the highest mean synchrony value of all three utterance types. But how rhetorical questions, which by definition have assertive force are more genuine than requests for action is not evident offhand. On the contrary, there have been frequent discussions on how requests and genuine questions are pragmatically related in that they are both subtypes of *mands*<sup>39</sup>. This analysis would rather lead to the assumption, that a request for action in interrogative form like A is closer to a genuine question than an assertive exclamation like B.

If both genuinity as a continuum and final pitch movement as an indicator of the degree of genuinity would be accepted – which by no means is self-evident –, the relation between requests for action and rhetorical questions requires further investigation. But while figure 7 gives cause for such interesting thought games, they remain in the purely speculative mode. Still, the data set is limited to only one lexical string per utterance type, which bears the great risk of undue generalization.

Before looking at other syllables, one thing should be said particularly about terminal pitch in the rhetorical questions. Dehé & Braun (2020) found in their experimentally elicited speech data, that the distinction between genuine polar questions and rhetorical polar questions is reflected in different ratios of final rising and final level contours. Final falling contours did not play a role in their study (cmp. Figure 3 in section 3.2). If this was true for the rhetorical questions in our data, the utterances of B would receive few if any clearly negative synchrony values. But as can be seen in figure 7, this is not the case. In fact, the rhetorical questions show the widest range of synchrony values in the object syllable *that*, from extremely rising to extremely falling pitch. The particular observation from Dehé & Braun (2020), that rhetorical polar questions are produced either with final rising or with final level contours, therefore cannot be verified by our study.

---

<sup>39</sup> A discussion on that matter can be found in Lyons (1977: 753 ff.) – even though Lyons argues for a categorical separation of questions and *mands* as explained in section 2.2.

As the mass values of the ultimate syllable of the lexical verb (*repeat/believe*) revealed a difference between utterance type A and the other two utterance types, the synchrony values at this position are of particular interest. The requests for action of A are centred around the perfect synchrony of 0.0, with few extreme values in any direction. Thus, neither clearly rising nor clearly falling pitch patterns do frequently occur here. A mean of -0.38 reflects that circumstance. While the means of B and C are rather moderate, too, with values of 2.98 and 2.01, respectively, the overall distributions of B and C show broader ranges, with more extreme values, mainly in the positive area of the spectrum.

The differences between A and B ( $F(2,246)=7.906$ ,  $p<0.001$ ) and between A and C ( $p<0.01$ ) are significant. But the inferential statistic blurs, what is overt in the descriptive statistic of figure 7: The violins of B and C, in clear opposition to A, depict bimodal distributions with narrow sections around the black flatline. This indicates, that these two utterance types tend to either rising or falling pitch in the lexically stressed syllable of the lexical verb, whereas level pitch contours are rare. Together with the overall high mass values of B and C in that position, their more extreme behaviour with regards to synchrony points to an exposed status of the lexical verb. The same cannot be observed for the requests for action of A, which show lower mass values and more moderate synchrony values. Of course, further studies with variation in the lexical material for all three utterance types would be needed to verify these findings.

Finally, we take a look at the initial syllable of the utterances, i.e. the auxiliaries *can* and *do*, respectively. Other than in the contemplation of mass, the segmental differences of the two are negligible, when inspecting synchrony. And as figure 7 shows, the three utterance types behave quite differently here. No tendency towards any direction can be seen in the rhetorical questions of B, which are equally distributed around a mean value of 0.53. A and C on the other hand mirror each other in moderately tending to falling and rising pitch, respectively. This difference is reflected in their mean values, -1.92 for A and 2.92 for C, and it is found to be statistically significant by the corresponding linear model ( $F(2,244)=22.01$ ,  $p<0.001$ ). The adjusted  $R^2$  value of 0.15 indicates, that utterance type indeed can account for some of the variance of the data.

However, the moderate tendencies towards rising initial pitch in topic introductions and towards falling initial pitch in requests for action, should be viewed in the light of the generally low mass values of the auxiliary verb. Especially *do* in the topic introductions

of C is prosodically weak, receiving only half of the average mass of all syllables (cmp. Figure 6). Therefore, the relevance of the auxiliary verb for tonal events is rather low. Taking this caveat into account, one can state the following: A single utterance of a non-genuine polar question might not be affected considerably in its pragmatic function by a particular change of pitch in the initial auxiliary, but a fine difference in prosodic behaviour between utterances of different functions can be observed in that position over the course of many utterances.

### 4.3.3 Utterance duration

The last measurement to be looked at is the utterance duration. As outlined in section 3.2, several studies have found rhetorical questions to bear slower speaking rates, i.e. overall longer durations, than genuine questions. This is explained by their assertive force and their resemblance to statements, which also show slower speaking rates than genuine questions.

In our study, rhetorical polar questions are not compared to genuine polar questions – as done by Dehé & Braun (2020) – but to other kinds of non-genuine polar questions, namely requests for actions and topic introductions. Figure 8 shows the durations of the utterances, sorted by their type (A, B or C), as density functions, i.e. as smoothed and normalized distribution curves. As density is a purely relative measure, the concrete numbers on the y-axis are not meaningful in themselves. They only derive from the fact, that the integral of each curve equals 1.0<sup>40</sup>, which makes the three curves comparable.

And the distributions clearly show, that the rhetorical questions of B tend to longer utterance durations than the requests for action of A and the topic introductions of C. There is a distance of ca. 75 ms between their mean values (765 ms for B, 686 ms and 690 ms for A and C, respectively), which are marked by the solid vertical lines. A linear model of utterance duration as a function of utterance type statistically confirms this difference, both in the comparison of A and B ( $F(2,246)=8.914$ ,  $p<0.001$ ) and B and C ( $p<0.001$ ).

---

<sup>40</sup> So if utterance duration on the x-axis would be measured in seconds rather than in milliseconds, the density values on the y-axis would change accordingly, i.e. they would multiply by the factor 100.



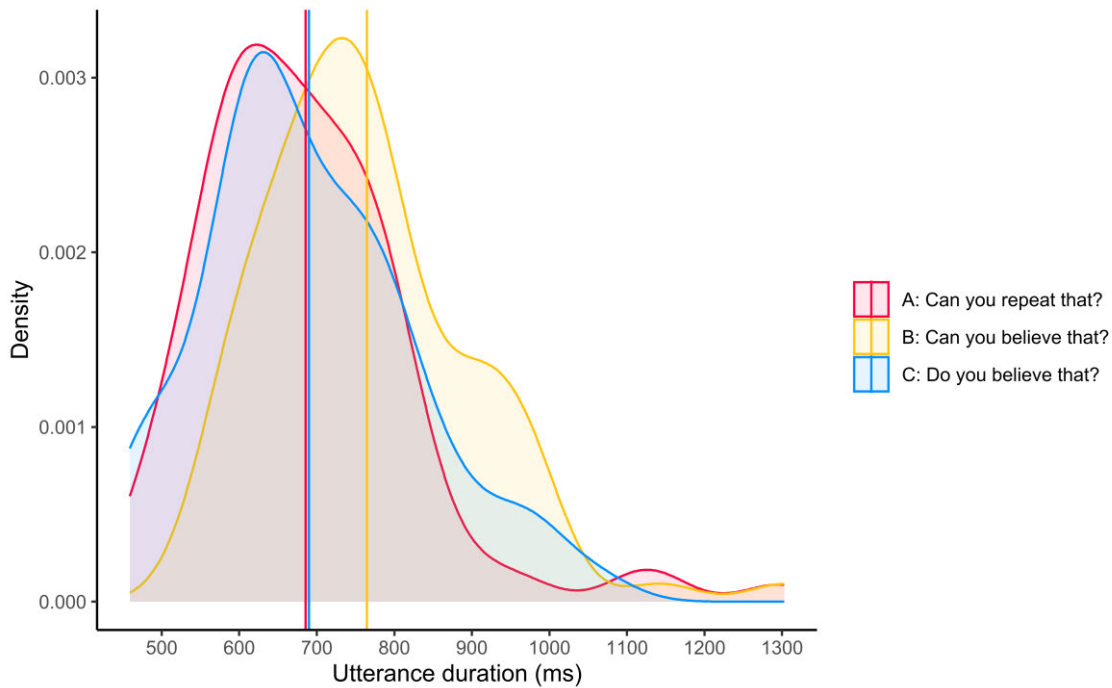


Figure 8: Density curves of utterance durations, sorted by utterance type (A, B, C).

From that it can be concluded, that longer duration seems to be a characteristic of rhetorical questions – not only in opposition to genuine questions (as reported by Dehé & Braun 2020), but also in opposition to other kinds of non-genuine questions, concretely requests for action and topic introductions.

One specific reservation about that claim needs to be made, though. Interestingly, more than the half of the durational difference between B and C results from the durational difference within the syllable of the auxiliary verb, where the mean values are 117 ms for B and 78 ms for C<sup>41</sup>. As already discussed in the analysis of mass, in our study the difference between utterance types correlates with a difference in lexical material at this position, namely *can* for B and *do* for C. The latter, in combination with the following *you* often gets reduced to its initial stop sound, which neatly explains its shorter average duration. Leaving out the auxiliary verb, the difference in mean duration of the remaining four and phonemically identical syllables between B and C is reduced to 36 ms, which is still relevant, but less clear.

The comparison of A and B on the other hand is not similarly affected by the lexical material. The main durational differences between these utterance types come from the

<sup>41</sup> A figure of absolute syllable durations for each utterance type can be found in the appendix A11.

subject syllable *you* and the object syllable *that*, which are lexically identical for A and B.

The possibility of an idiosyncratic bias in the rhetorical question B is also worth mentioning. Utterances of B are often produced in a theatrical manner, possibly leading to more emphasis and longer duration. While rhetorical questions generally seem to be prone to that kind of theatricality, it is by no means a necessary condition. So again, larger studies with non-genuine polar questions of the same kind, but with different lexical material, could shed more light on the fine differences observed here.

But despite the mentioned caveats, it should be repeated, that our results are generally in line with the previous literature in showing a tendency towards longer utterance duration for rhetorical polar questions.

#### **4.3.4 Summary**

The results, especially for mass and synchrony, don't show very clear tendencies, which suggest easy interpretations. Nevertheless, differences between the three utterance types could be found in some positions.

With regards to the discussed literature, the most interesting findings concern the utterance-final syllable *that*, which was found to be prosodically relevant with mostly above-average mass values. Here, none of the three analysed utterance types shows a clear tendency towards final rises, as it was found for genuine polar questions. Rather, there is a gradual difference between the distinct speech-acts: The requests for action show an almost equal distribution of falls and rises, whereas the rhetorical questions as well as the topic introductions only slightly tend to rising patterns. This led to speculations on whether synchrony can be taken as an indicator of the degree of genuinity in polar questions.

Worth highlighting is also, that the polar rhetorical questions show instances of both final rising and final falling patterns. In this, the results contradict Dehé & Braun (2020), who found hardly any final falls in rhetorical polar questions, but level contours instead.

Another interesting phenomenon could be observed in the lexically stressed syllable of the lexical verb, where both rhetorical questions and topic introductions show overall high mass values, paired with rather extreme synchrony values, both positive and negative.

The requests for action, on the other hand, reveal a more moderate behaviour, with average mass and synchrony values centred around the *neutral* zero mark.

While the position of the initial auxiliary with its overall low mass values does not seem to be prosodically relevant at first sight, fine differences between the investigated indirect speech-acts, could be found at this unlikely spot. This regards both differences in relative mass – with the reservations concerning the lexical material as discussed in section 4.3.1 – and differences in synchrony. Especially the requests for action and the topic introductions reveal opposing tendencies here, with the former showing higher mass and negative synchrony values, and the latter showing lower mass, but positive synchrony values.

Finally, the measurement of utterance duration confirms the previous findings on rhetorical questions. Former studies have shown, that rhetorical questions have slower speech rates than genuine questions. The current results indicate, that they have longer durations than other kinds of non-genuine questions, too. Importantly, this finding has to be qualified by the fact, that our study design contains a difference in the lexical material between the three utterance types, which directly appears to affect the absolute duration of the auxiliary verb.

Coarsely summarizing, it remains difficult to establish straight one-to-one connections between the illocutionary force of an utterance and its concrete prosodic features. Nevertheless, connections between the two could be observed in terms of mass, synchrony and duration. But the prosodic differences between the utterance types, where present, are differences in degree, rather than in category.

## Chapter 5: Conclusion

This study has investigated semantic, pragmatic and prosodic characteristics of questions in general and non-genuine polar questions in specific. The definitional issues, both theoretically and practically, i.e. when evaluating concrete utterances, have shown the complexity of the phenomenon. This complexity was reflected in the ambiguous results of the phonetic analysis, too.

Questions as a linguistic category and thereby their subtype of polar questions were, in terms of Huddleston (1994), defined semantically by their set of potential answers. To

analyse the various pragmatic functions of questions, the theory of speech-acts and its concept of illocutionary force were applied, which led to the crucial distinction between genuine and non-genuine polar questions.

As former studies on question intonation could reveal, genuine polar questions strongly tend to final rises, whereas non-genuine polar questions don't show such a tendency. And to investigate, whether similar prosodic distinctions could be found between different speech-acts from the heterogeneous group of non-genuine polar questions, an exploratory study was conducted. At first, a corpus of utterances from American television programs was created, which consisted of three utterance types, differing in their illocutionary force. Concretely, requests for action, rhetorical questions and topic introductions, represented by utterances of one specific polar question each, were analysed.

A thorough pragmatic analysis revealed the more or less ambiguous character of the functional categories. Especially the utterances, which were subsumed under the term topic introduction, showed variations with regards to speaker expectation or question bias. Once again, it became clear, that speaking is not the application of a formal grammatical system. It is acting in a complex system of social relations and constantly shifting conventions. By saying something, one inevitably performs several and multilayered speech-acts at the same time.

With this premise in mind, but nevertheless relying on the established primary speech-act categories, the utterances were analysed prosodically. Three concrete features were looked at in detail, namely mass and synchrony on syllable level and duration on utterance level. Mass and synchrony are idiosyncratic measurements of the *ProPer* method, which in form of periodic energy takes into account not only pitch, but the intelligibility of pitch, too.

The analysis' results could not show any one-to-one correlations between specific types of non-genuine polar questions and their prosodic behaviour. Small differences between the utterance types, however, could be observed for all measurements. And these differences, as subtle they might seem, indicate, that illocutionary force is conveyed prosodically to some degree. But as fluid and ambiguous as pragmatic functions are in the interplay of human language, as fluid and ambiguous are their prosodic realizations.

While linguistic categories like non-genuine polar questions – or the separate indirect speech-acts, that non-genuine polar questions can be sub-divided into – are very useful

and indispensable from a heuristic perspective, they hardly depict, what is *really* going on. And the same goes for the fine-grained prosodic measurements, that were used in our study. Coming from a different angle, they help finding patterns in the use of human language, and thereby gaining a richer understanding of it. But human language, nevertheless, remains a dynamic social phenomenon, that can be approximated in a formalized way, but never deciphered entirely in all its intertwined dimensions.

## References

- Albert, Aviad. 2023. A model of sonority based on pitch intelligibility. *Studies in Laboratory Phonology* 13. Berlin: Language Science Press. DOI: 10.5281/zenodo.7837176
- Albert, Aviad & Cangemi, Francesco & Ellison, T. Mark & Grice, Martine. 2023, July 27. ProPer: PROsodic analysis with PERiodic energy. DOI: 10.17605/OSF.IO/28EA5.
- Albert, Aviad & Cangemi, Francesco & Grice, Martine. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of the 9th International Conference on Speech Prosody*. 804–808.
- Austin, John Langshaw. 1962. How to do things with words. Oxford University Press.
- Banuazizi, Atissa & Creswell, Cassandre. 1999. Is that a real question?: Final rises, final falls and discourse function in yes-no question intonation. In Sabrina J. Billings, John P. Boyle & Aaron M. Griffith (eds.), *CLS 35-1: Papers from the main session*. Chicago: Chicago Linguistic Society. 1–13.
- Barnes, Jonathan & Brugos, Alejna & Shattuck-Hufnagel, Stefanie & Veilleux, Nanette. 2012a. On the nature of perceptual differences between accentual peaks and plateaux. In: Oliver Niebuhr & Hartmut Pfitzinger (eds.), *Understanding prosody – the role of context, function and communication*. 93–118.
- Barnes, Jonathan & Veilleux, Nanette & Brugos, Alejna. & ShattuckHufnagel, Stefanie. 2012b. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, 3(2). 337–383.
- Beckman, Mary E. & Ayers, Elam. 1997. *Guidelines for ToBI labelling*, version 3. Columbus, OH: The Ohio State University Research Foundation.

- Boersma, Paul & Weenink, David. 2023. Praat: doing phonetics by computer. Version 6.3.12. URL: <http://www.praat.org>. Download: July 2023.
- Braun, Bettina & Dehé, Nicole & Neitsch, Jana & Wochner, Daniela, & Zahner, Katharina. 2018. The Prosody of Rhetorical and Information-Seeking Questions in German. *Language and Speech*, 62(4). 779–807.
- Breen, Mara & Fedorenko, Evelina & Wagner, Michael & Gibson, Edward. 2010. Acoustic correlates of information structure. *Language and Cognitive Process*, 25 (7-9). 1044–1098. DOI: 10.1080/01690965.2010.504378
- Cameron, Deborah & McAlinden, Fiona & O’Leary, Kathy. 1989. Lakoff in context: The social and linguistic functions of tag questions. In Jennifer Coates & Deborah Cameron (eds.), *Women in their speech communities*. London, New York: Longman. 74–93.
- Cangemi, Francesco & Albert, Aviad & Grice, Martine. 2019. Modelling intonation: Beyond segments and tonal targets. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*. 572–576.
- Couper-Kuhlen, Elizabetz & Selting, Margret. 1996. Towards an interactional perspective on prosody and a prosodic perspective on. *Prosody in conversation: Interactional studies*, 11.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.
- Cruttenden, Alan. 1997. *Intonation*. Second edition. New York: Cambridge University Press.
- De Ruiter, Jan P. 2012. Introduction: Questions are what they do. In Jan P. De Ruiter (ed.), *Questions: Formal, Functional and Interactional Perspectives* (Language, Culture & Cognition). Cambridge: Cambridge University Press. 1–8.

- Dehé, Nicole & Braun, Bettina. 2020. The prosody of rhetorical questions in English. *English Language & Linguistics*, 24(4). 607–635.
- Dehé, Nicole & Braun, Bettina & Wochner, Daniela. 2018. The prosody of rhetorical vs information-seeking questions in Icelandic. In *Proceedings of Speech Prosody, 9th International Conference, 13–16 June 2018, Poznan, Poland*. 403–407.
- Fries, Charles Carpenter. 1964. On the Intonation of ‘Yes-No’ Questions in English. In David Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott & J. L. M. Trim (eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*. London: Longmans, Green & Company, Ltd. 242–255.
- Gussenhoven, Carlos. 1983. *A semantic analysis of the nuclear tones of English*. Bloomington: Indiana University Press.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation* (Research surveys in linguistics). Cambridge: Cambridge University Press.
- Haan, Judith. 2002. *Speaking of Questions. An Exploration of Dutch Question Intonation*. Dissertation. Utrecht: Lot.
- Hedberg, Nancy & Juan M. Sosa. 2002. The prosody of questions in natural discourse. *Proceedings of Speech Prosody 2002, Aix-en-Provence, France*. 375–378.
- Hedberg, Nancy & Sosa, Juan M. & Görgülü, Emrah. 2017. The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory*, 13(2). 321–368.
- Heuven, Vincent J. van & Zanten, E. van. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, 47(1–2). 87–99.



- Huddleston, Rodney. 1994. The contrast between interrogatives and questions. *Journal of Linguistics*, 30(2). 411–439.
- Hudson, Richard. 1975. The meaning of questions. *Language*, 51. 1–31.
- Internet Archive. Accessed November 2023. <https://archive.org>.
- Internet Archive About. Accessed November 2023. <https://archive.org/about>.
- Kenworthy, Joanne. 1978. The intonation of questions in one variety of Scottish English. *Lingua*, 44(2–3). 267–282.
- Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Levinson, Stephen C. 2012. Interrogative intimations: On a possible social economics of interrogatives. In Jan P. De Ruiter (ed.), *Questions: Formal, Functional and Interactional Perspectives* (Language, Culture & Cognition). Cambridge: Cambridge University Press. 11–32.
- Levis, John M. 1999. The intonation and meaning of normal yes/no questions. *World Englishes*, 18(3). 373–380.
- Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet & Hirschberg, Julia. 1990. The meaning of intonation contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Maratha E. Pollack (eds.), *Intentions in communications*. Cambridge: MIT Press. 271–311.
- R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org>. Download: July 2023.

Sadock, Jerrold M. & Zwicky, Arnold M. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language typology and syntactic description*. Cambridge: Cambridge University Press. 155–196.

Searle, John Rogers. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge University Press.

TV News Archive. Accessed March 2023. <https://archive.org/details/tvarchive>.

WELT Nachrichtensender. 2022. G7-Pressekonferenz: “Könnte ich!” Nach dieser Frage kanzelt Olaf Scholz eine Journalistin ab. YouTube video, accessed February 2024. URL: <https://www.youtube.com/watch?v=1NMM2tGbxZo>

## Appendix

### A1: Nuclear contours and their distribution (Hedberg et al. 2017)

Nuclear contour	Percentage	ToBI category	Total N
Low-rise	79.8%	L*H-H%	323
		L*H-↑H%	2
		L*L-H%	2
High-rise	10.7%	H*H-H%	39
		!H*H-H%	5
High-fall	2.4%	H*L-L%	9
		!H*L-L%	1
Rise-fall	1.7%	L+H*L-L%	7
Low-fall	1.5%	L*L-L%	6
Fall-rise	0.5%	H*L-H%	1
		L+H*L-H%	1
Level	3.4%	H*H-L%	8
		!H*H-L%	3
		ǀH*H-L%	1
		L*+HH-L%	2
<b>Total</b>	<b>100%</b>		<b>410</b>

Source: Hedberg et al. 2017: 332.

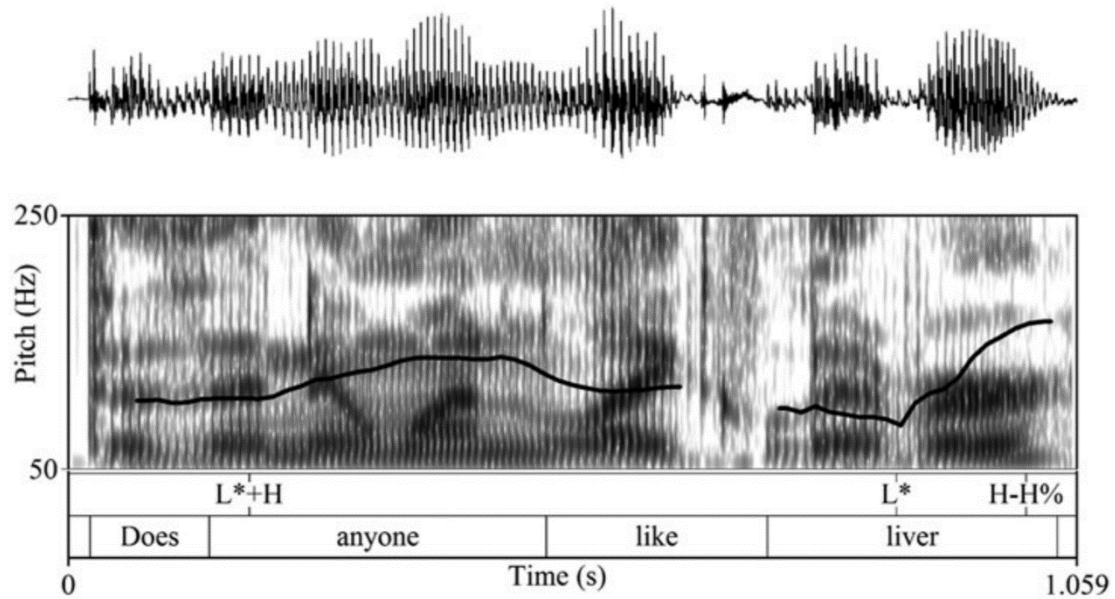
### A2: Example of a polar question and its two contexts (Dehé & Braun 2020)

Context for ISQ	Context for RQ
At a garden party, you offer canapés with Limburger cheese. You would like to know which of the guests eat this and whether they want some of it or not. You say to the guests:	Your friend offers his guests a cheese tray, including Limburger. However, it is well known that none of your friends like stinky cheese and therefore nobody will touch it. You say to your friend:

Target Q: Does anyone eat Limburger?

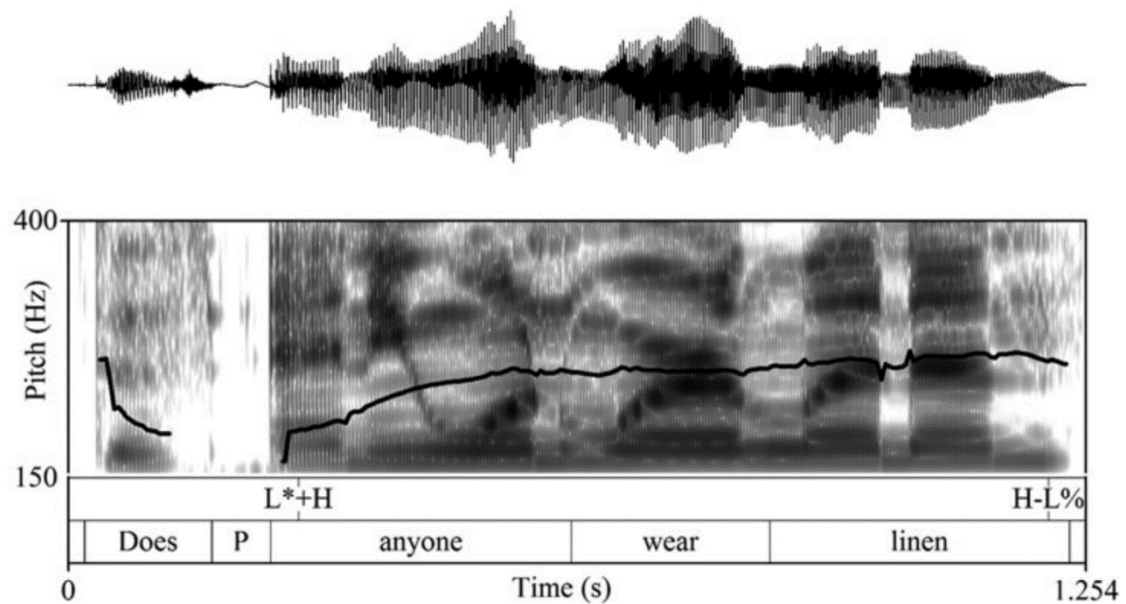
Target Q stands for target question. ISQ is short for information-seeking question, which is equivalent to genuine question. RQ is short for rhetorical question. Source: Dehé & Braun 2020: 614.

### A3: Genuine polar question with a final rise (Dehé & Braun 2020)



Oscillogram, spectrogram, pitch contour and MAE\_TOBI annotation of the genuine polar question "Does anyone like liver?". Source: Dehé & Braun 2020: 626.

### A4: Rhetorical polar question with a final mid-level plateau (Dehé & Braun 2020)



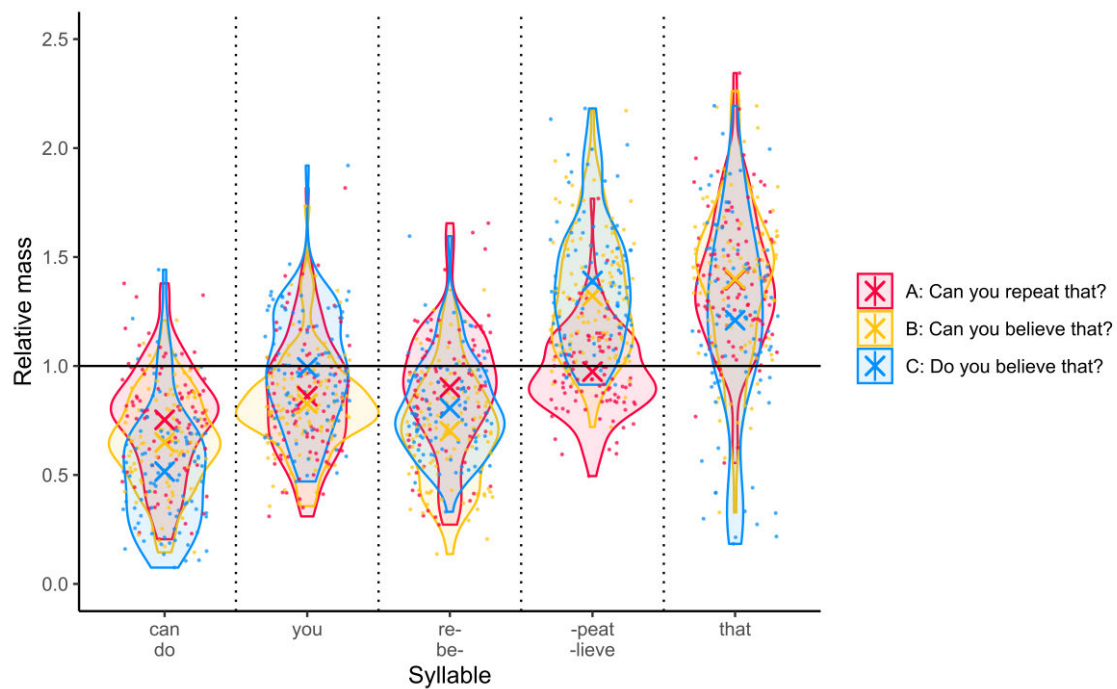
Oscillogram, spectrogram, pitch contour and MAE\_TOBI annotation of the rhetorical polar question "Does anyone wear linen?". Source: Dehé & Braun 2020: 627.

## A5: Meta information on utterances A, B and C

Utterance type	A				B					C		
Total	91				86					72		
Situational context	Hearing	Interview	Press conference	Other	Interview	News	Speech	Weather report	Other	Interview		
	25	56	5	5	11	39	13	17	6	72		
Sex of the speaker	m		f		m		f		m		f	
	60		31		55		31		57		15	
Immediate floor change	Change	No change	Overlap		Change	No change	Overlap		Change	No change	Overlap	
	62	21	8		15	67	4		61	8	3	
Verbal response	Response	No response	Unknown		Response	No response	Unknown		Response	No response	Unknown	
	88	0	3		28	56	2		72	0	0	
Answer	Affirmative	Negative	No answer/unknown		Affirmative	Negative	No answer/unknown		Affirmative	Negative	No answer/unknown	
	20	1	70		0	7	79		17	17	38	
Excuse by the speaker	Excuse	Explanation	Neither		/					/		
	65	4	22									
Request complied with by the addressee	Complied	Not complied	Unknown		/					/		
	87	3	1									
Type of addressee	Direct				Direct	Indirect	Unclear		Direct			
	91				44	40	2		72			

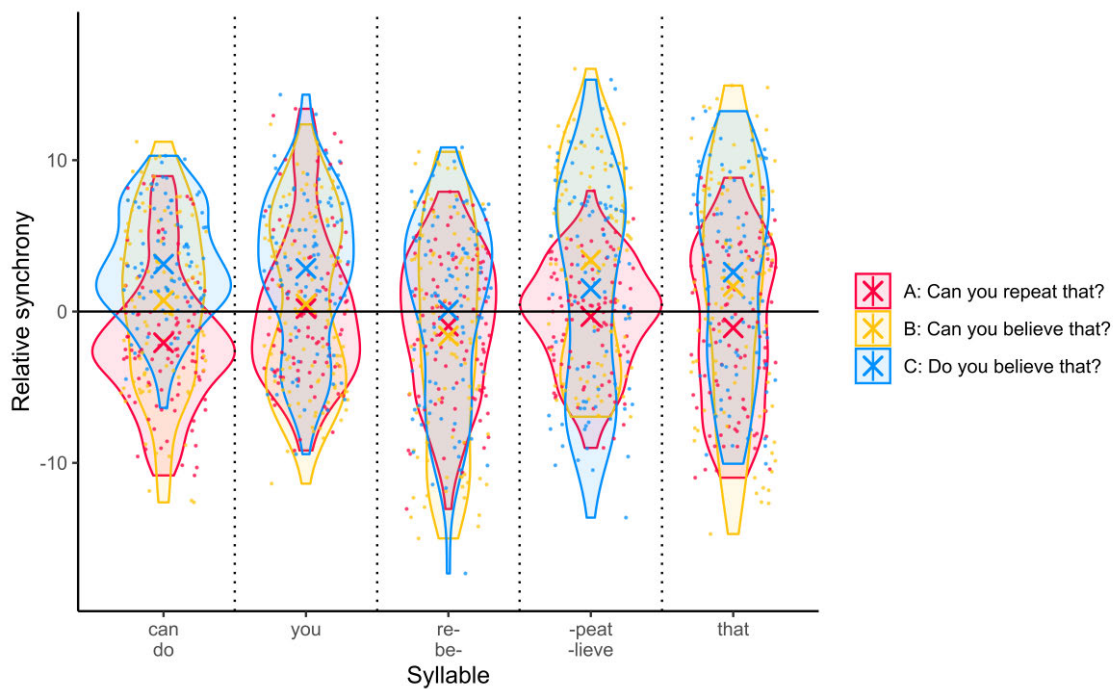
## A6: Results N=278 (all utterances, regardless from focus conditions)

### A6.1: Relative mass



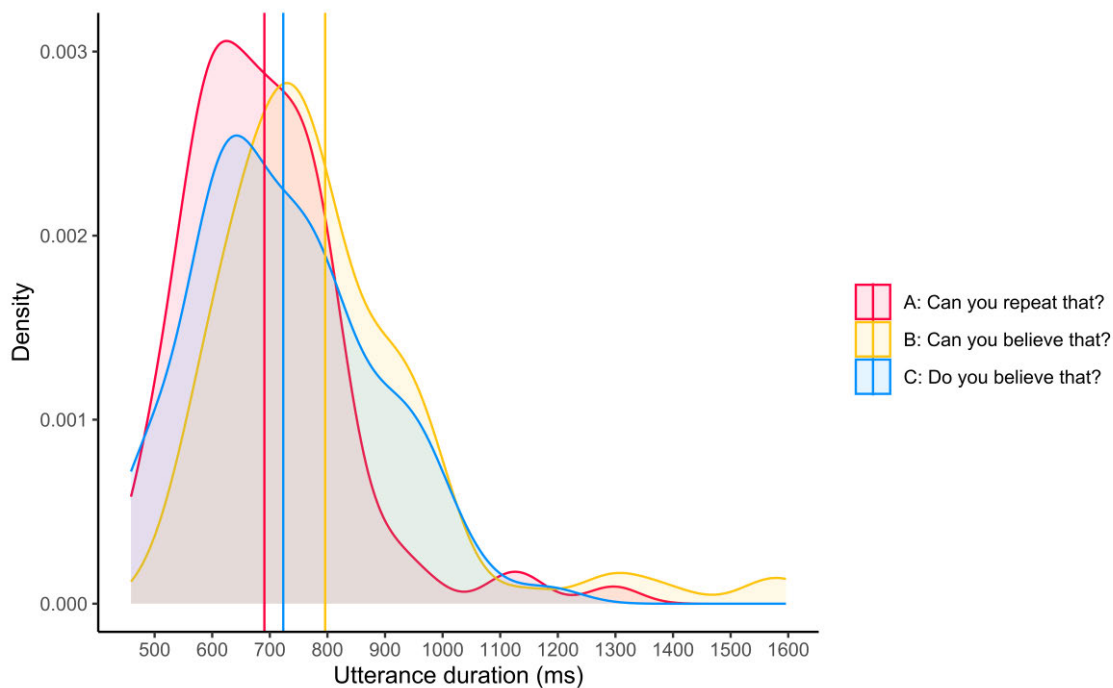
Data set: all utterances, regardless from their focus condition, N=278.

**A6.2: Relative synchrony**



Data set: all utterances, regardless from their focus condition, N=278.

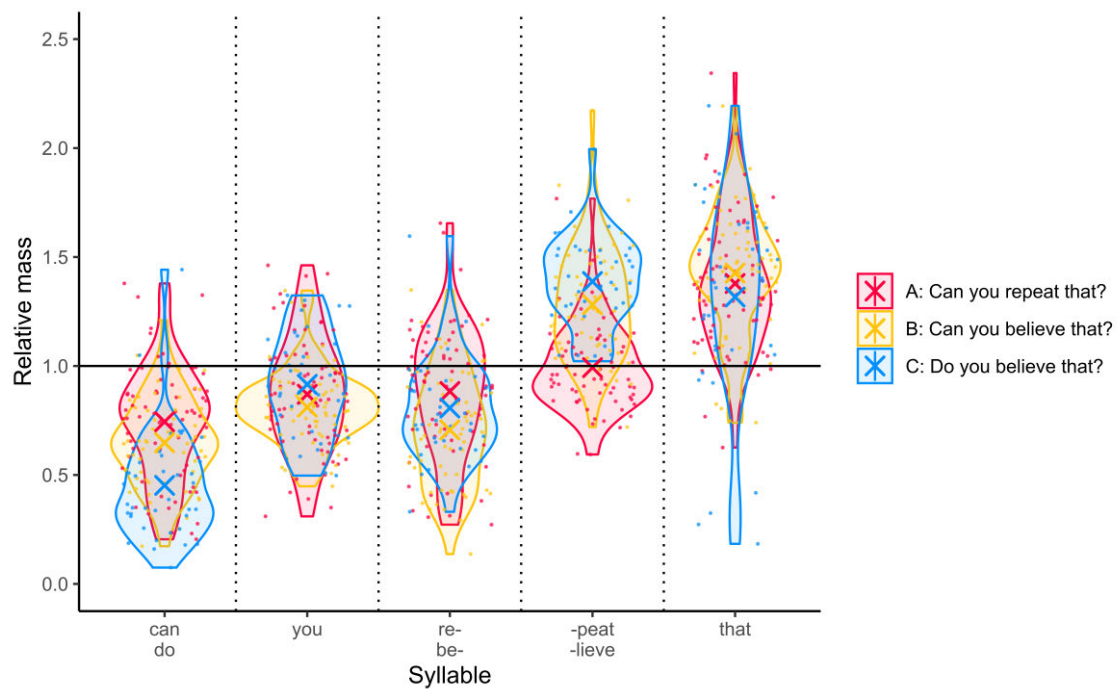
**A6.3: Utterance duration**



Data set: all utterances, regardless from their focus condition, N=278.

## A7: Results N=167 (utterances without any narrow focus)

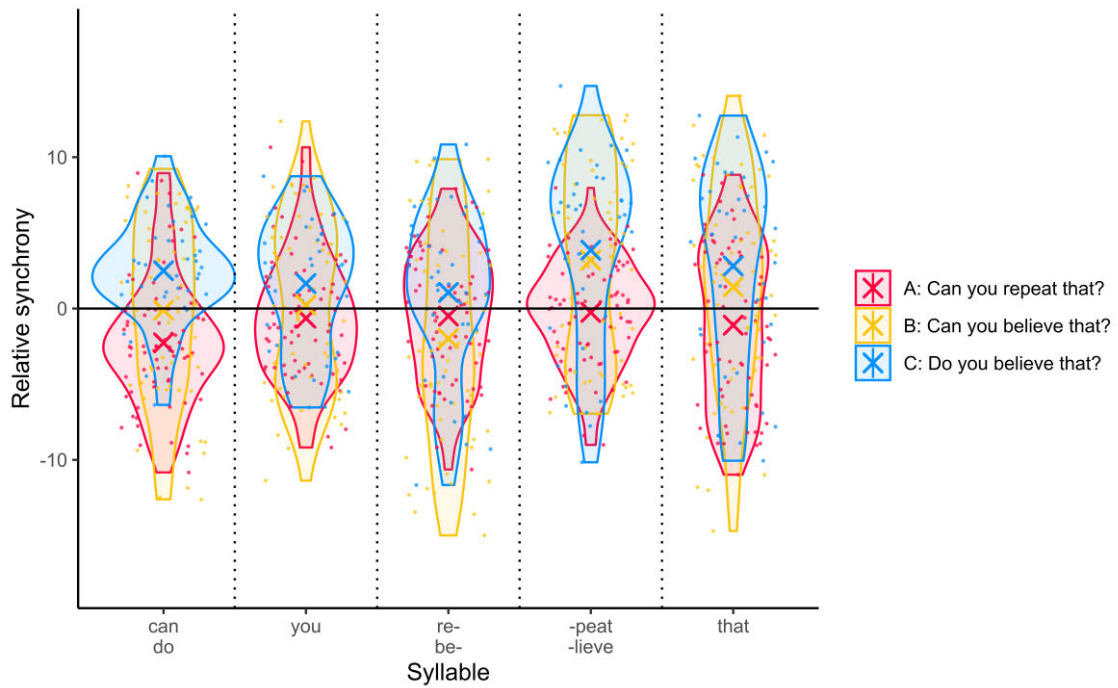
### A7.1: Relative mass



Data set: utterances without narrow focus, N=167.

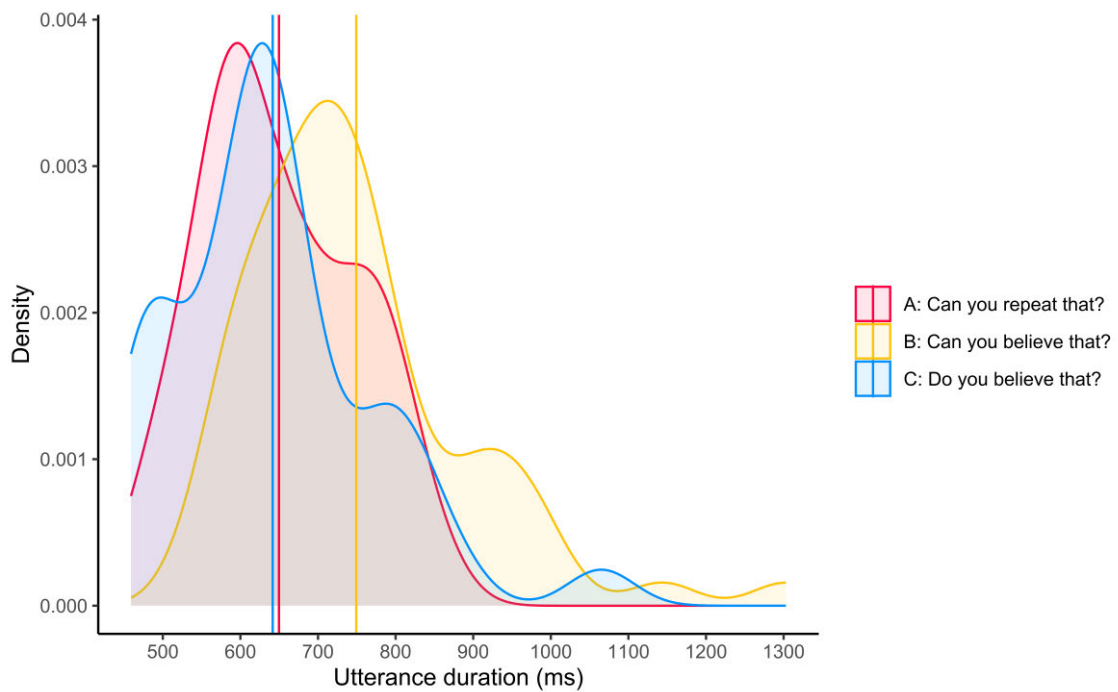


## A7.2: Relative synchrony



Data set: utterances without narrow focus, N=167.

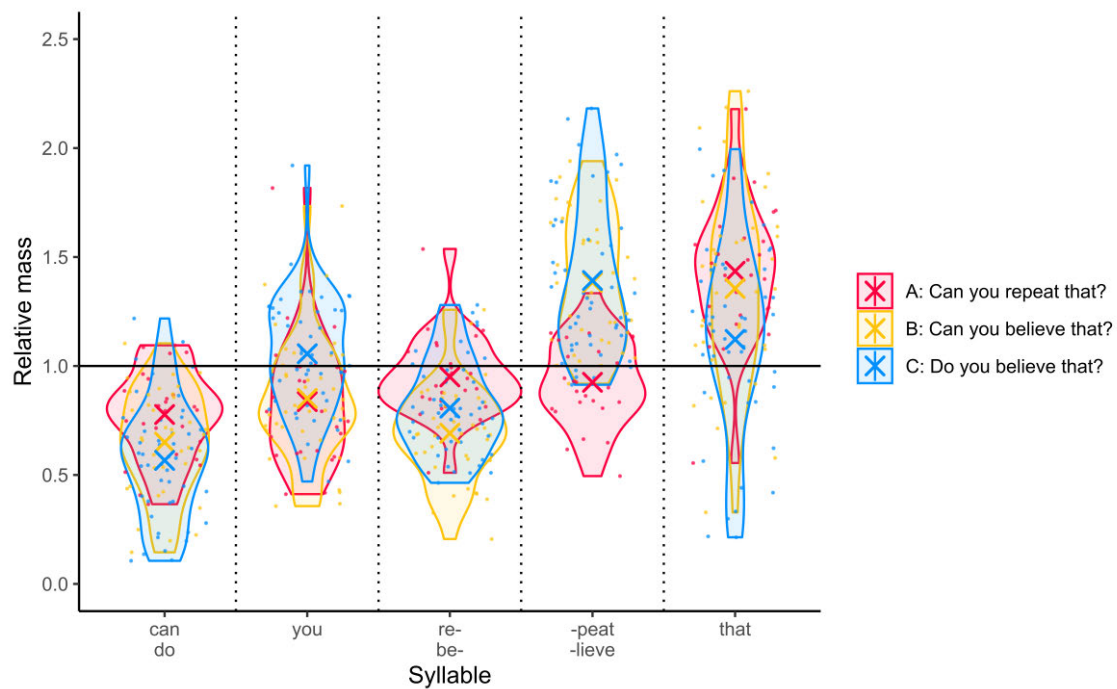
## A7.3: Utterance duration



Data set: utterances without narrow focus, N=167.

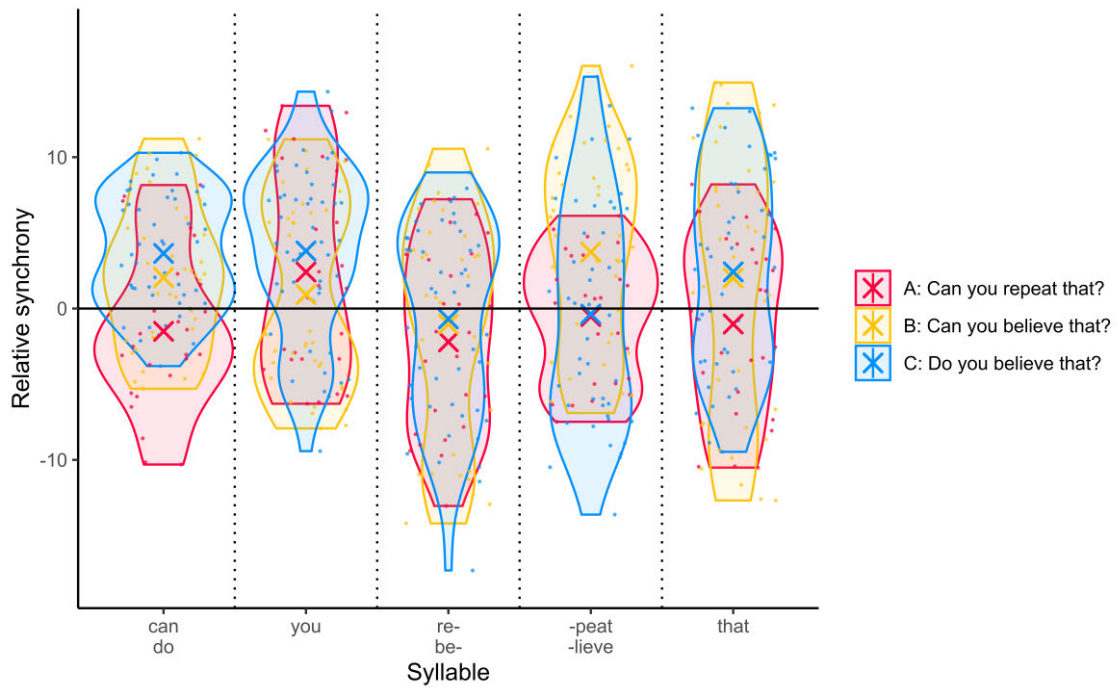
## A8: Results N=111 (only utterances with narrow focus, mild or strong)

### A8.1: Relative mass



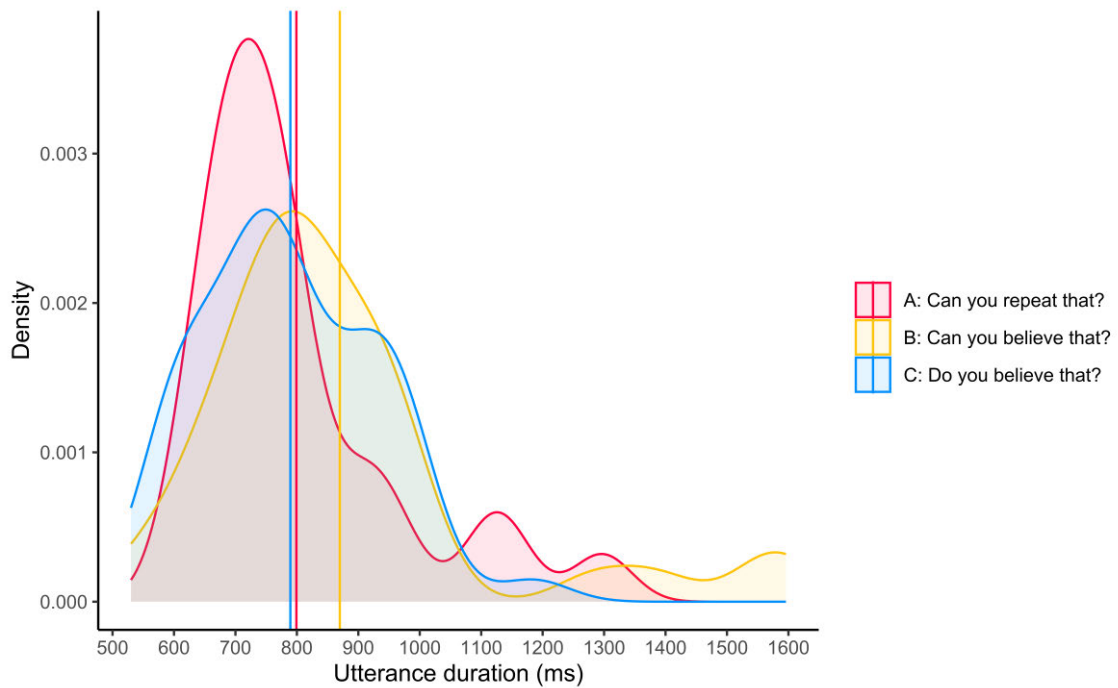
Data set: only utterances with narrow focus, N=111.

### A8.2: Relative synchrony



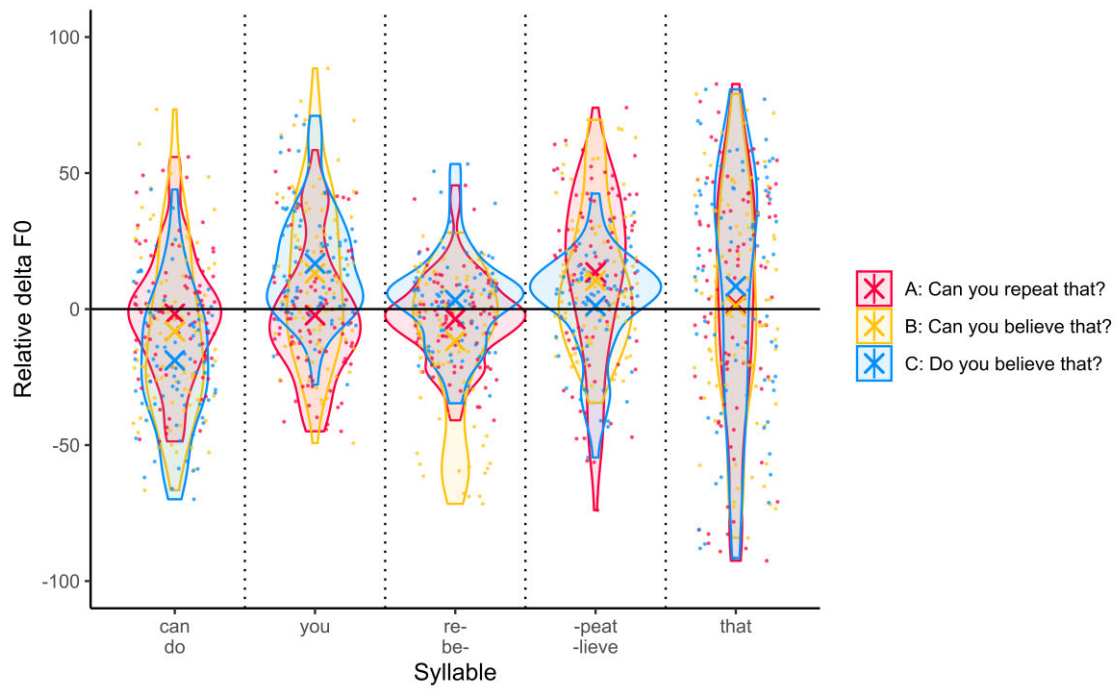
Data set: only utterances with narrow focus, N=111.

### A8.3: Utterance duration



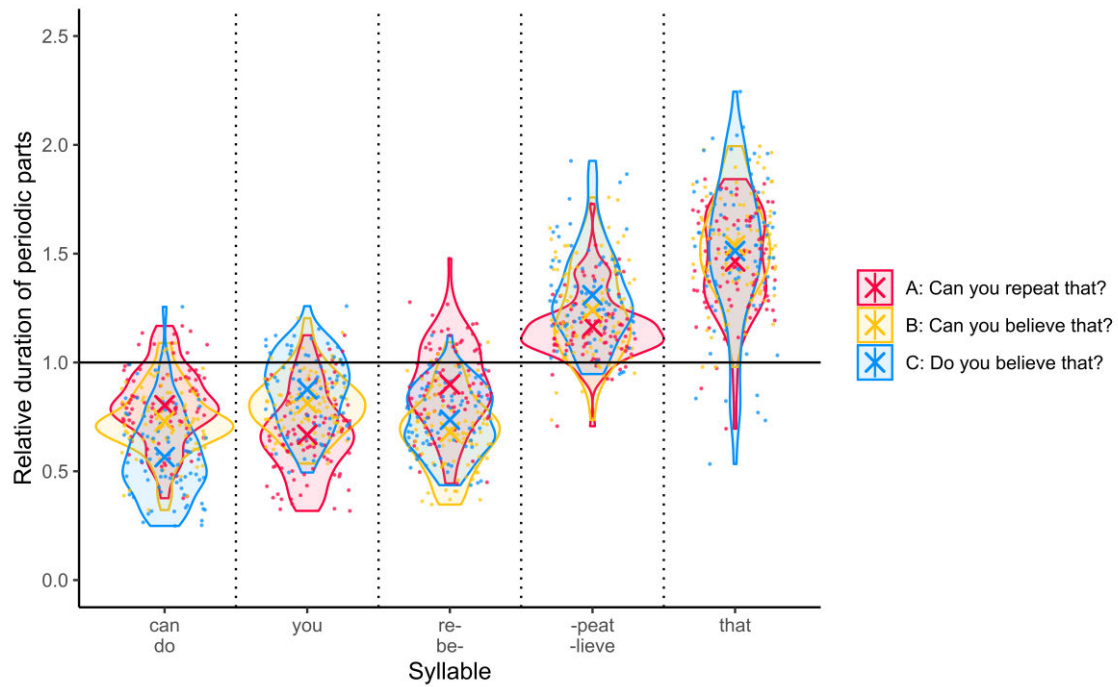
Data set: only utterances with narrow focus, N=111.

### A9: Delta F0 (standard data set)



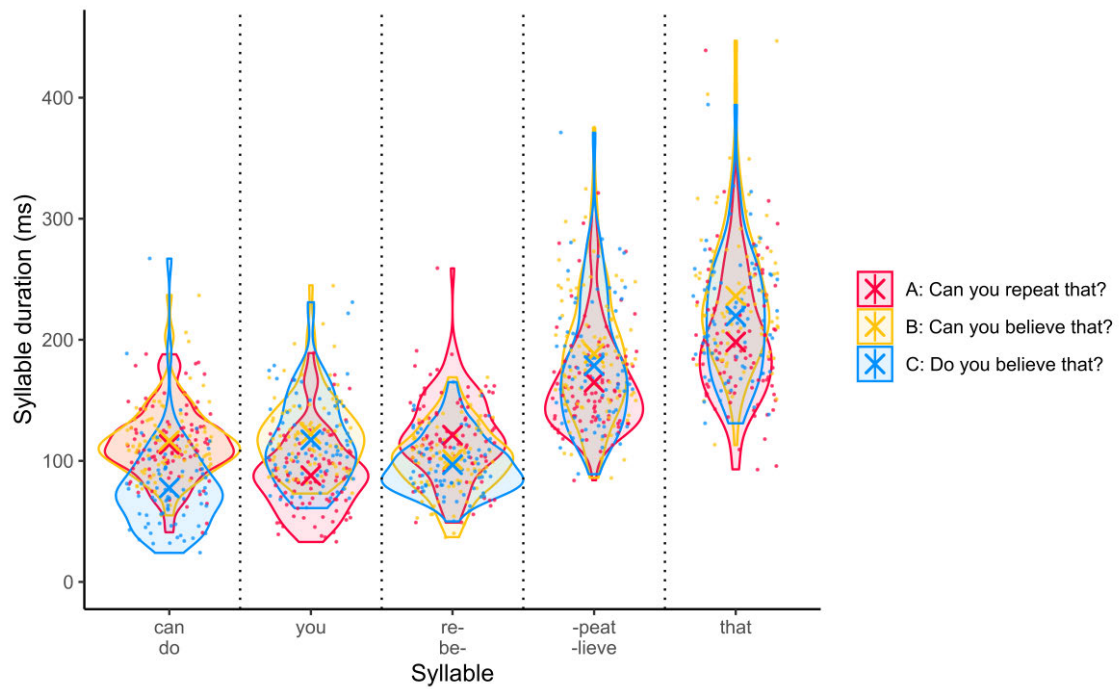
Data set: all utterances without strong narrow focus, N=249.

### A10: Relative duration of periodic parts (standard data set)



Data set: all utterances without strong narrow focus, N=249.

### A11: Absolute syllable duration (standard data set)



Data set: all utterances without strong narrow focus, N=249.

### A12: Table of IDs

S. "Appendix\_A12\_Table\_of\_IDS.xlsx" in the digital appendix.

## Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche unter Angabe der Quelle kenntlich gemacht.

Diese Arbeit ist in gleicher oder ähnlicher Form im Rahmen einer anderen Prüfung noch nicht vorgelegt.

Köln, 12.03.2024 Unterschrift: 